

Networks with communities and clustering

Citation for published version (APA):

Stegehuis, C. (2019). *Networks with communities and clustering*. [Phd Thesis 1 (Research TU/e / Graduation TU/e), Mathematics and Computer Science]. Technische Universiteit Eindhoven.

Document status and date:

Published: 31/01/2019

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

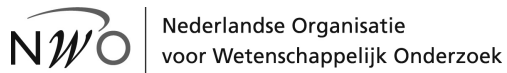
If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Networks with communities and clustering

This work was financially supported by The Netherlands Organization for Scientific Research (NWO) through the TOP grant 613.001.451.



© Clara Stegehuis, 2018

Networks with communities and clustering

A catalogue record is available from the Eindhoven University of Technology Library
ISBN: 978-90-386-4688-6

Cover created by Freepik

Printed by Gildeprint Drukkerijen, Enschede

Networks with communities and clustering

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Technische Universiteit Eindhoven, op gezag van de rector magnificus, prof.dr.ir. F.P.T. Baaijens, voor een commissie aangewezen door het College voor Promoties, in het openbaar te verdedigen op 31 januari 2019 om 16.00 uur

door

Clara Stegehuis

geboren te Amersfoort

Dit proefschrift is goedgekeurd door de promotoren en de samenstelling van de promotiecommissie is als volgt:

voorzitter: prof. dr. J.J. Lukkien
1e promotor: prof. dr. R. van der Hofstad
2e promotor: prof. dr. J.S.H. van Leeuwen
leden: prof. dr. N. Litvak
prof. dr. L. Massoulié (MSR-INRIA joint centre)
prof. dr. T. Müller (Rijksuniversiteit Groningen)
prof. dr. P.F.A. Van Mieghem (TU Delft)
prof. dr. A.P. Zwart

Het onderzoek dat in dit proefschrift wordt beschreven is uitgevoerd in overeenstemming met de TU/e Gedragscode Wetenschapsbeoefening.

Acknowledgments

The completion of this thesis would not have been possible without the help and support of many others.

First of all, I would like to thank my supervisors Remco and Johan for their guidance during the past four years. Without you, none of this work would have been possible. Remco, thank you for your enthusiasm about the research during our meetings and for all your patient explanations. Johan, thank you for giving me the opportunity to share my research with a wide audience at many different places, this really has made my PhD a lot more fun. Thanks to both of you for all your feedback on my writing, and paying attention to all the mathematical details in my work, I learned a lot from both of you.

I would also like to thank all members of my defense committee: Nelly Litvak, Laurent Massoulié, Tobias Müller, Piet Van Mieghem and Bert Zwart. Thank you for taking the time to read my thesis and for providing me with valuable feedback.

While studying applied mathematics, I did not think I would become a PhD student after graduating. Nelly, thank you for giving me the idea me that a PhD could be something that would be interesting for me. You were an excellent master student supervisor, and I am grateful for your advice and for our collaborations after that.

During my PhD, Laurent provided me with the opportunity to spend four months at the INRIA research institute in Paris. It was a very nice experience to work on a topic which was completely different from the projects I had worked on before. Laurent, thank you for providing me with this opportunity and for continuing our discussions about the work after my return to Eindhoven.

Several parts of this thesis are the result of collaborations with other researchers. Guido, I learned many new mathematical techniques during our chats about the collaborations that resulted in Chapters 3 and 4, and I appreciated your preciseness when reading draft versions of a paper. I would also like to thank Ellen for our collaboration that resulted in Chapter 10, Jane and Angus for collaborating on parts of Chapter 5 and Alessandro for collaborating with me on Chapter 9. Furthermore, I would like to thank Tom, Pim and Nelly for their collaboration on other papers that do not appear this thesis. I learned a lot from all of you while writing these papers together.

The STO section has been a great environment for working, and I would like to thank all other PhDs of the STO group for creating such a positive environment with many social activities. Thanks in particular to Souvik, Alberto and Joost for sharing

an office with me where we could discuss the good and the more frustrating parts of our PhDs. Also thanks to Alessandro, for accompanying me on travels to many conferences. Thanks to Enrico as well, for providing me with a step by step guide on how to complete a PhD, and to Murtaza and Alessandro for going through these final stages of our PhDs together. I would also like to thank Petra, Chantal and Pascale for helping out with all my non-mathematical questions.

I would also like to thank the members of the NETWORKS program, in particular for the training weeks. The courses provided in these weeks gave an interesting overview of areas of mathematics I did not know yet. I would also like to thank all NETWORKS PhD students for all the fun we had in the evenings of the training weeks while talking and playing games.

I truly enjoyed blogging about my research and presenting it to a wider audience as one of the Faces of Science. For this, I would like to thank Martine in particular. Thank you for proofreading all my blogs and putting them on the website. Furthermore, I would like to thank Suzanne and René for giving me the opportunity to present my work at many different places and helping me to improve my presentation skills.

I would also like to thank my parents, Bas and my friends for their continuous friendship and support. Finally, I would like to thank Jorn. Thank you for always being there for me. I am happy I got to share this experience with you and I am looking forward to our future together.

Contents

Contents	vii
1 Introduction	1
I Clustering and correlations	15
2 Degree-degree correlations	17
3 Global clustering in inhomogeneous random graphs	43
4 Local clustering in inhomogeneous random graphs	67
5 Local clustering in erased configuration models and uniform random graphs	91
6 Local clustering in dynamic and spatial models	127
II Subgraph structure	145
7 Subgraphs in erased configuration models	147
8 Subgraph fluctuations in inhomogeneous random graphs	181
9 Subgraphs in preferential attachment models	201
10 Finding induced subgraphs in inhomogeneous random graphs	229
III Networks with community structure	243
11 Hierarchical configuration models	245
12 Power-law relations in networks with community structures	273
13 Epidemics on networks with community structures	285

14 Mesoscopic scales in hierarchical configuration models	299
Bibliography	331
Summary	345

1 Introduction

A network is a collection of vertices, connected in pairs by edges. Many objects of interest can be thought of as networks. Examples include social networks, the Internet, biological networks, the brain or communication networks. Most of these networks are large, containing millions or even billions of vertices. For example, the network of English Wikipedia pages and their hyperlinks contains several millions of vertices, whereas some social networks contain more than a billion vertices. Even though these large networks are very different in application, their connectivity patterns often share several universal properties. For example, most networks contain communities: groups of densely connected vertices. In social networks these communities may correspond to groups of friends or groups of people with similar interests, but many other types of networks also contain community structures. Another frequently observed network property is that two neighbors of a vertex are more likely to be connected as well. In a social network for example, this means that two of your friends are likely to know each other.

Networks are typically modeled with random graphs: mathematical models generating large networks that may serve as null models for real-world networks. Whereas a real-world network often consists of just one network observation, random graphs are able to generate many network samples, allowing for statistical analyses. Furthermore, properties of these random graph models can often be analyzed mathematically. For this reason, random graph models are used to study network properties. In this thesis, we study several observed properties of real-world networks using random graph models, aiming to understand the similarities and the differences between random graph models and real-world networks.

1.1 Scale-free network models

Many real-world networks contain *hubs*, vertices that are substantially more connected than most other vertices in the network. Figure 1.1 illustrates this for the Gowalla social network [137] and plots the fraction of vertices of degree at least k against k on a log-log scale. Vertices in this network are members of the Gowalla social network, and edges indicate the friendship between two members of the network. The figure shows that indeed the network contains hubs: while most members only have a few friends, some members have several thousands of friends. This implies that the degree distribution of the network has a heavy right tail. One way of modeling such heavy-tailed degree distributions is through a power-law distribution.

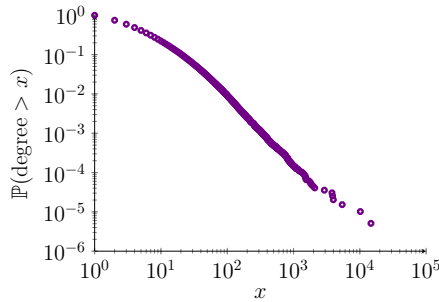


Figure 1.1: A loglog plot of the degree distribution of the Gowalla social network [137].

That is, if p_k denotes the fraction of vertices of degree k , then

$$p_k \approx Ck^{-\tau}, \quad (1.1.1)$$

for some $C > 0$, where τ is also called the degree exponent. The probability that the degree of a vertex is at least k then scales as $k^{1-\tau}$. Thus, if the fraction of vertices of degree at least k is plotted against k for a power-law degree distribution, this plot should follow a straight line on a log-log scale, as can indeed be observed in Figure 1.1.

Many networks were found to have degree distributions that can be approximated by a power law with degree exponent $\tau \in (2, 3)$ [6, 79, 122, 212]. These networks are called *scale-free networks*, because the degree of a random vertex in the network does not have a typical scale: the difference between the smallest and the largest degree in a network may be several orders of magnitude. Among the real-world networks that were found to be scale-free are the Internet [79], citation networks [188], e-mail networks [212] and online social networks [59]. However, several other networks were found to have different types of degree distributions [52, 66].

Over the course of several decades, various models for generating and studying scale-free networks have been developed. Most of these models are probabilistic, and fall in the class of random graph models. We now describe several popular scale-free random graph models that will reappear later in this thesis.

1.1.1 Configuration model

Given a degree sequence of n positive integers $\mathbf{d} = (d_1, d_2, \dots, d_n)$, the *configuration model* (CM) [41] is a random graph model that generates a graph on n vertices where vertex i has degree d_i . Given a degree sequence such that $\sum_{i=1}^n d_i$ is even, the configuration model equips each vertex j with d_j free half-edges. The configuration model is then constructed by successively pairing free half-edges into edges, uniformly at random, until no free half-edges remain. When all half-edges have been paired, the resulting graph has degree sequence \mathbf{d} , as illustrated in Figure 1.2a and 1.2b. The generated network may be a multigraph, so that it may contain self-loops and multiple edges between a pair of vertices. Most real-world networks on the other hand are *simple*, i.e., they do not contain self-loops and multiple edges. For a simple graph

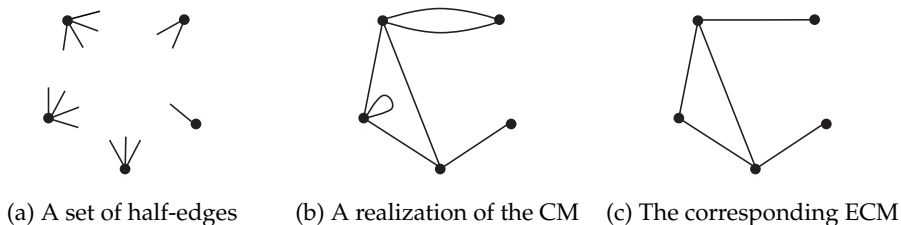


Figure 1.2: Illustration of the CM and ECM.

with degree sequence \mathbf{d} , to exist, \mathbf{d} needs to be graphical, see [104]. Conditionally on the configuration model resulting in a simple graph, it is uniform over all simple graphs with the prescribed degree sequence. Thus, every simple graph on degree sequence \mathbf{d} is then generated with equal probability. This is why the configuration model is often used as a standard model for real-world networks with given degrees.

In this thesis, we usually take the degree sequence as an i.i.d. sample of a random variable D . To model scale-free networks, we sample from the degree distribution

$$\mathbb{P}(D = k) = Ck^{-\tau}, \quad (1.1.2)$$

for $\tau \in (2, 3)$ and some constant $C > 0$, so that the configuration model indeed constructs a scale-free network. For $\tau > 2$, a degree sequence constructed in this manner is graphical with probability tending to one, conditionally on the sum of the degrees being even [10]. When the sum of the sampled degrees is odd, we add an extra half-edge to the last vertex.

An important property of the configuration model is that it is *locally tree-like*, which means that in the large network limit, any small neighborhood of a vertex will look like a tree.

1.1.2 Uniform random graph

Given a graphical degree sequence, the uniform random graph samples each simple graph with that degree sequence with equal probability. As stated before, the configuration model creates a uniform random graph conditionally on the configuration model resulting in a simple graph. When $\tau > 3$, the configuration model constructs a simple graph with positive probability [116]. One can then obtain a uniform random graph with prescribed degree sequence by repeating the configuration model construction until it results in a simple graph. However, when $2 < \tau < 3$, the probability that the configuration model creates multiple-edges and self-loops tends to one [116], so that the configuration model can no longer be used to sample simple uniform random graphs. In this regime, sampling uniform random graphs is difficult. Several algorithms start with an initial graph, and rewire some of the edges of the graph at each time step, finally resulting in a uniform random graph sample [11, 97, 151]. However, the number of necessary time steps for these algorithms to result in a uniform sample is unknown for $2 < \tau < 3$.

1.1.3 Erased configuration model

Since the configuration model often creates multigraphs, and uniform random graphs are difficult to construct for $2 < \tau < 3$, we often study the *erased configuration model* (ECM) instead, which first constructs a configuration model, and then removes all self-loops and multiple edges, as illustrated in Figure 1.2c. This erasing procedure does not affect the degree distribution in the large-network limit [50], so that the erased configuration model generates networks with a scale-free degree distribution as long as the original degree sequence is scale free. Like the configuration model, the erased configuration model is locally tree-like.

1.1.4 Rank-1 inhomogeneous random graph

Whereas the configuration model constructs a network with an exact given degree sequence, the *rank-1 inhomogeneous random graph* (also called *hidden-variable model*) constructs networks with soft constraints on the vertex degrees. The model starts with n vertices, where each vertex i is equipped with a weight h_i . Then, each pair of vertices is connected independently with probability $p(h_i, h_j)$. Several choices for the function $p(h_i, h_j)$ exist. For example, the Chung-Lu version of the rank-1 inhomogeneous random graph uses the connection probability [61]

$$p(h_i, h_j) = \min\left(\frac{h_i h_j}{\mu n}, 1\right), \quad (1.1.3)$$

where μ denotes the average weight. The connection probability $p(h_i, h_j) = 1 - \exp(-h_i h_j / (\mu n))$ has been introduced as the Norros-Reittu model [166] and the connection probability $p(h_i, h_j) = h_i h_j / (\mu n + h_i h_j)$ as the maximally random graph or generalized random graph [174], but many other connection probabilities are possible. It can be shown that the degree of a vertex is close to its weight for these three connection probabilities [36, 50], which makes the original weight sequence a soft constraint on the resulting degree sequence. Therefore, choosing the weights to have a power-law distribution results in a random graph with a power-law degree distribution with the same exponent.

The rank-1 inhomogeneous random graph is a locally tree-like model, as the configuration model and the erased configuration model.

1.1.5 Preferential attachment model

Another important network model is the *preferential attachment model*, a dynamic network model that can generate scale-free networks for appropriate parameter choices [4, 54]. In contrast to the random graph models described before, the preferential attachment model is a *dynamic* random graph model, adding vertices step by step. Many different versions of the preferential attachment model exist. In this thesis, we consider a modification of [20, Model 3]. This model has parameters m and δ and starts with two vertices, vertex 1 and 2, with m edges between them. Then, at each step $t > 2$, vertex t is added with m new edges attached to it. These m edges are paired to existing vertices one by one. The probability that the j th edge of vertex t

attaches to vertex $i < t$ is given by

$$\mathbb{P}(\text{jth edge of } t \text{ connects to } i) = \begin{cases} \frac{D_i(t-1)+\delta}{2m(t-2)+(t-1)\delta} & j = 1, \\ \frac{D_i(t-1,j-1)+\delta}{2m(t-2)+(j-1)+(t-1)\delta} & j = 2, \dots, m, \end{cases} \quad (1.1.4)$$

where $D_i(t-1)$ denotes the degree of i before adding vertex t , while $D_i(t-1, j-1)$ denotes the degree of vertex i after the first $j-1$ edges of vertex t have been attached.

The first term in the denominator of (1.1.4) ensures that a new vertex has higher probability of attaching to a vertex of high degree than to a vertex of low degree, which creates power-law degrees. The parameter δ controls the degree exponent τ of the power-law degree distribution as [107, Lemma 4.7]

$$\tau = 3 + \delta/m. \quad (1.1.5)$$

Choosing $\delta \in (-m, 0)$ then results in a scale-free network.

The preferential attachment model may, like the configuration model, result in a multigraph. However, whereas in the configuration model the number of edges between two vertices may grow in n , the maximum number of edges between any pair of vertices in the preferential attachment model is bounded by the parameter m . A significant difference between the preferential attachment model and the previously mentioned models is that the edges of the preferential attachment model can be interpreted as *directed* edges, pointing from the younger vertex towards the older vertex.

1.1.6 Hyperbolic random graph

The last scale-free random graph model that we describe is the *hyperbolic random graph*, where vertices are connected based on their embedding in a geometric space [130]. The model samples n vertices in a disk of radius R . Each vertex i can then be described by its angular coordinate $\phi_i \in [0, 2\pi]$ and its radial coordinate $r_i \in [0, R]$. The density of the radial coordinate of a vertex is given by

$$\rho(r) = \alpha \frac{\sinh(\alpha r)}{\cosh(\alpha R) - 1} \quad (1.1.6)$$

with $\alpha = (\tau - 1)/2 > 1/2$, so that $\tau > 2$. The angular coordinate ϕ is sampled uniformly from $[0, 2\pi]$. Then, two vertices are connected if their hyperbolic distance is at most R . The hyperbolic distance between points $u = (r_u, \phi_u)$ and $v = (r_v, \phi_v)$ is defined by

$$\cosh(d(u, v)) = \cosh(r_u) \cosh(r_v) - \sinh(r_u) \sinh(r_v) \cos(\theta_{uv}), \quad (1.1.7)$$

where θ_{uv} denotes the relative angle between ϕ_u and ϕ_v . Asymptotically, this creates a simple random graph with a power-law degree distribution with exponent τ [100]. The radius is chosen as $R = 2 \log(n/\nu)$, where the parameter ν fixes the average degree of the graph. Figure 1.3 shows an illustration of a hyperbolic random graph.

The hyperbolic random graph is able to generate simple, scale-free networks, but in contrast to the erased configuration model and the rank-1 inhomogeneous

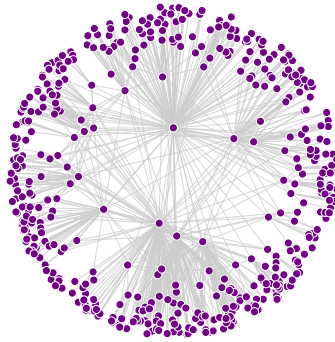


Figure 1.3: An example of a hyperbolic random graph with $n = 500$ and $\tau = 2.5$. Vertices are embedded based on their radial and angular coordinates.

random graph, it is not locally tree-like. Instead, the geometric nature of the model causes many triangles to be formed [56, 130]. Indeed, if two vertices are connected, they must be geometrically close to one another. Therefore, two neighbors of a vertex are also likely to be geometrically close to one another, and are likely to be connected, creating many triangles. Furthermore, using hyperbolic geometry rather than Euclidean geometry causes the network to be scale free. For this reason, the hyperbolic random graph is widely used as a model for real-world networks [8, 38, 47, 89].

1.2 Network structures

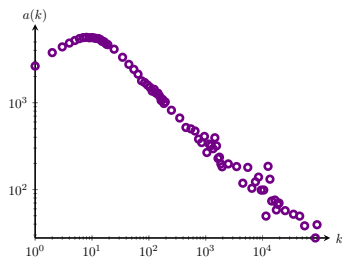
In this thesis, we aim to understand several structural network properties of random graph models described above as well as in real-world networks. We now describe these properties in more detail.

1.2.1 Degree correlations

The most simple structure in a network is given by its edges and describes which pairs of vertices connect to one another. Do high-degree vertices typically connect to other high-degree vertices, or are they more likely to connect to lower-degree vertices? One popular measure for such degree correlations is *assortativity*, Pearson's correlation coefficient of the degrees at the two ends of an edge [161]. To obtain more detailed information about the degree correlations in a network, one can investigate the average degree of neighbors of a vertex of degree k , denoted by $a(k)$, and defined as

$$a(k) = \frac{1}{kn_k} \sum_{i:d_i=k} \sum_{j \in \mathcal{N}_i} d_j, \quad (1.2.1)$$

where n_k denotes the number of vertices of degree k , and \mathcal{N}_i denotes the set of neighbors of vertex i . In networks without degree correlations, $a(k)$ is independent of k and $k \mapsto a(k)$ forms a flat curve. When $a(k)$ decays in k , high-degree vertices are typically connected to lower-degree vertices. On the other hand, when $a(k)$

Figure 1.4: $a(k)$ for the Youtube social network [137]

increases in k , this implies that high-degree vertices are likely to connect to other high-degree vertices. Several real-world networks display a decaying $a(k)$ [146, 176], as is illustrated for the Youtube social network in Figure 1.4. This decaying curve is often ascribed to the constraint that the network must be simple. In a simple scale-free network, a high-degree vertex can only connect to a few other high-degree vertices, because only few high-degree vertices exist. The remaining connections therefore have to be to lower-degree vertices, reducing the average degree of a neighbor of a high-degree vertex.

Degree correlations have been found to influence the behavior of various important processes acting upon networks. For example, degree correlations influence the behavior of an epidemic process on a network [36, 39, 222] or the robustness of a network when a fraction of edges is removed [211].

1.2.2 Clustering

The second network property describes the relation between the neighbors of a vertex. In many networks, two neighbors of a vertex are likely to be connected to one another as well, so that a large number of triangles is present. The tendency for a network to create triangles is often captured in terms of the clustering coefficient. The clustering coefficient of a network is defined as

$$C = \frac{6\Delta}{\sum_{i=1}^n d_i(d_i - 1)}, \quad (1.2.2)$$

where Δ denotes the number of triangles in the network. This clustering coefficient can be interpreted as the number of pairs of neighbors that are connected themselves divided by the total number of pairs of neighbors. However, when the degree exponent satisfies $\tau < 3$, this clustering coefficient tends to zero in the large graph limit for any sequence of graphs, even when a large number of triangles is present [172].

Since this is an undesirable property, an alternative clustering coefficient is defined as

$$C_{\text{avg}} = \frac{1}{n} \sum_{i=1}^n \frac{2\Delta_i}{d_i(d_i - 1)}, \quad (1.2.3)$$

also called the average clustering coefficient. Here Δ_i denotes the number of triangles attached to vertex i . This clustering coefficient measures the average fraction of pairs of neighbors of a vertex that connect to one another. Many random graphs with

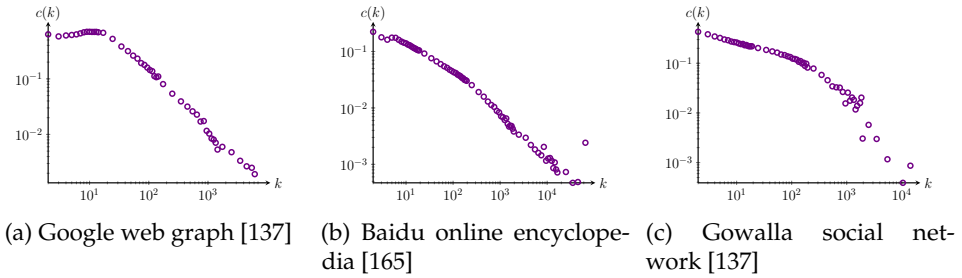


Figure 1.5: Clustering spectrum $c(k)$ of three real-world networks.

$\tau \in (2, 3)$ exist that have a non-vanishing average clustering coefficient [49, 70, 101]. For the rank-1 inhomogeneous random graph and the erased configuration model on the other hand, C_{avg} vanishes as n grows large due to the locally tree-like nature of these models.

The average clustering coefficient gives a global picture of the presence of triangles in a network. A more detailed picture of the triangle structures in a network can be obtained in terms of the *local clustering coefficient* $c(k)$: the probability that two neighbors of a vertex of degree k are connected to one another. More precisely, $c(k)$ is defined as

$$c(k) = \frac{1}{n_k} \sum_{i:d_i=k} \frac{2\Delta_i}{k(k-1)}, \quad (1.2.4)$$

where n_k again denotes the number of vertices of degree k . In many real-world networks $c(k)$ has been observed to decay in k [131, 146, 212], as illustrated in Fig. 1.5. This indicates for example that two random friends of a popular person are less likely to know each other than two random friends of a less popular person. The clustering spectrum of a network influences the spread of epidemic processes on a network, and it contains information about its community structure, making it an essential network property [198, 202].

1.2.3 Subgraphs

The clustering spectrum of a network describes the presence of triangles in the networks. However, other subgraphs, such as squares or larger complete graphs can also provide relevant structural information. Subgraphs that occur frequently in a network are also called *motifs*, and are viewed as important building blocks [9, 152]. Whereas most real-world networks display high clustering, their larger building blocks may be significantly different. For example, frequently occurring subgraphs in the World Wide Web were found to be very different from those in food webs or gene regulation networks [152]. Network motifs are believed to contain information about the function of the network [152, 200]. In biological networks for example, several motifs have been assigned a specific biological function they perform [200].

It is also possible to count induced subgraphs, where edges not present in the subgraph are required not to be present in the network. For example, when we are interested in counting all square induced subgraphs, we count all sets of four vertices

that contain a square in the edges between them, but no more edges than that. On the other hand, when we count all square subgraphs, we count all squares, but also all complete graphs of size four, since they also contain a square. Frequently occurring induced subgraphs in networks are sometimes called *graphlets*.

1.2.4 Community structures

The final network property we discuss describes larger network subgraphs. Many real-world networks contain communities: groups of densely connected vertices [92]. These community structures have been observed in many different types of networks. In social networks for example, communities can be thought of as groups of friends, but in other types of networks, such as collaboration networks, biological networks and technological networks, community structures have also been observed [80, 92]. The presence of communities is related to the presence of clustering and motifs. For example, we may expect a group of friends to consist of many triangles or larger complete graphs. Thus, many dense motifs are expected to be present in networks with a community structure. The major difference between motifs and community structures lies in the scale of the subgraphs that are considered. Motifs are usually assumed to be of fixed size, while the graph is large. Thus, motifs describe the presence of microscopic structures in the network. Communities on the other hand, often occur on a mesoscopic scale: a network of n vertices may contain communities of size n^α for some $\alpha \in (0, 1)$ [80, 181] so that community sizes grow in the network size n .

1.3 Bond percolation

One of the goals of studying structural network properties is to find out how they influence processes acting upon these networks. For example, we would like to know how the presence of non-trivial clustering influences the spread of a message on a social network, or how the presence of hubs influences the spreading of computer viruses on computer networks. Many models for processes taking place on networks have been defined, for example aiming to model epidemic spreading, synchronization phenomena, traffic congestion or searching on networks (see [18, 163] for an overview).

In this thesis, we analyze the process of *bond percolation*, a simple model for a process acting on a network, where every edge of the network is deleted independently with probability $1 - p$. This models a situation where each edge of the network is subject to a random failure with probability $1 - p$. By investigating the effect of this edge-removal process on the structure of the network, we measure the robustness of a network to random failures.

One of the simplest measures of robustness is the fraction of vertices in the largest connected component of the network after the edge deletions. The configuration model and the rank-1 inhomogeneous random graph undergo a phase transition under bond percolation when $\tau > 3$. When p is larger than some value $p_c > 0$, the largest component of the network remains a positive fraction of the total number of vertices in the large-network limit. Thus, a large part of the network remains

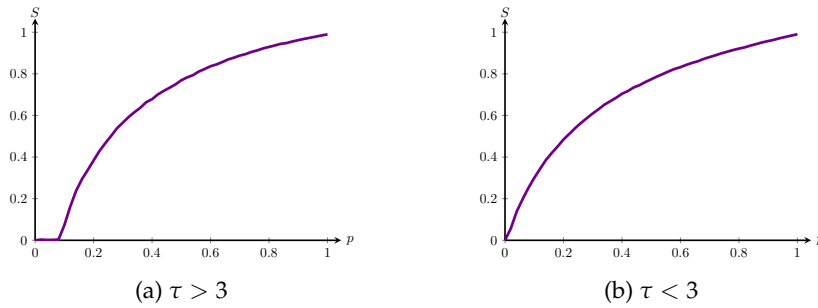


Figure 1.6: Illustration of the behavior of the fraction of vertices in the largest connected component S under bond percolation with probability p in the configuration model.

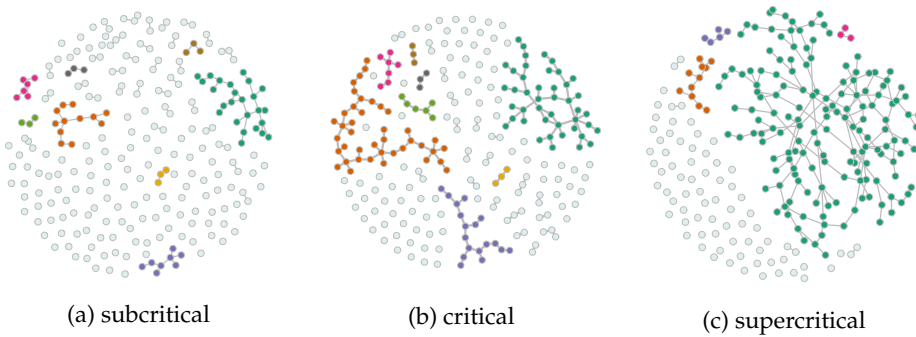


Figure 1.7: Illustration of the subcritical, critical and supercritical components of a configuration model with $\tau > 3$ under percolation.

connected after the random failures. This regime is also called the supercritical regime. When $p < p_c$ on the other hand, the fraction of vertices in the largest component tends to zero in the large-network limit, so that the network has broken down into many small pieces. This regime is also called the subcritical regime.

For $p \approx p_c$, the behavior of the percolated network is often substantially different from the behavior in the supercritical and the subcritical regimes. In the configuration model and the rank-1 inhomogeneous random graph, the component sizes then scale as n^α for some $\alpha \in (0, 1)$ depending on τ in the critical regime [22, 23, 71, 72]. Thus, the component sizes are of intermediate (or mesoscopic) size in the critical regime, as illustrated in Figure 1.7, and grow as a function of the network size n .

On the other hand, for $\tau < 3$ a component containing a positive fraction of vertices exists in the configuration model and the rank-1 inhomogeneous random graph for any $p > 0$ in the large-network limit [55, 117], as illustrated in Figure 1.6. Thus, in the scale-free regime, the largest component of a configuration model never breaks down under a random edge attack. This phenomenon has also been observed in real-world networks with scale-free degree distributions [5, 51].

1.4 Main results and outline

Network structures such as clustering, degree-correlations, motifs, graphlets and community structures influence the behavior of various network processes. Therefore, it is important to understand these network structures in random graph models, as well as in real-world network data. This thesis addresses several questions regarding these network structures, which we categorize below in three parts.

1.4.1 Part 1: Clustering and correlations

The first part of this thesis considers the presence of degree correlations and clustering in real-world networks as well as in scale-free network models, and aims to answer the following research questions:

What is the degree-correlation structure of scale-free random graph models? (Chapter 2)

We investigate the behavior of the average nearest-neighbor degree $a(k)$ defined in (1.2.1) in three simple (single-edge constrained) scale-free random graph models: the erased configuration model, the rank-1 inhomogeneous random graph and the hyperbolic random graph. We conclude that $a(k)$ follows a universal curve across these three random graph models. For small values of k , $a(k)$ remains flat, so that the degree-degree correlations are not visible yet for small degrees. For larger degrees, $a(k)$ starts to decay as a power of k . This power of k as well as the point where $a(k)$ starts to decay is universal for all three scale-free network models, even though these models are very different in nature. The universally decaying curves theoretically support the claim that simple scale-free networks contain negative degree-degree correlations.

How does the clustering vanish in locally tree-like random graph models? (Chapter 3)

The rank-1 inhomogeneous random graph is known to be locally-tree like. Therefore, its average clustering coefficient tends to zero in the large-network limit. We determine the rate at which the average clustering coefficient decays to zero as a function of the network size for a wide class of rank-1 inhomogeneous random graphs. Interestingly, for τ close to 2, the decay is extremely slow, so that the average clustering coefficient only starts to vanish for extremely large networks.

Can we characterize clustering in scale-free networks? (Chapters 4, 5 and 6)

We first show in Chapter 4 that several real-world scale-free networks display similar clustering spectra $k \mapsto c(k)$. We then show that the clustering spectrum of the rank-1 inhomogeneous random graph (in Chapter 4) as well as the erased configuration model and the uniform random graph (in Chapter 5) are similar in shape to those of the real-world network data, apart from the fact that $c(k)$ vanishes for small k which is often not the case for real-world networks. We then generalize our method to analyze the clustering spectrum of several other scale-free random graph models in Chapter 6, including the hyperbolic random graph and the preferential attachment model. While the method to analyze clustering in these models is universal, the shape of the clustering spectrum crucially depends on the specifics of the random graph model.

1.4.2 Part 2: Network subgraphs

The second part of this thesis studies the presence of larger network patterns in the form of subgraphs (also called motifs and graphlets) in random graph models and real-world networks. We consider the rank-1 inhomogeneous random graph and the erased configuration model as well as real-world networks, and aim to answer the following questions:

What is the subgraph structure of scale-free random graphs and real-world networks? (Chapters 7-9)

We analyze the subgraph structure of the erased configuration model (Chapter 7), the rank-1 inhomogeneous random graph (Chapter 8) and the preferential attachment model (Chapter 9). We obtain asymptotic (induced) subgraph counts for all subgraphs. We further identify the degrees of the vertices where a specific (induced) subgraph is most likely to be present, by introducing a variational principle. This variational principle resolves the trade-off between the high connectivity of high-degree vertices and their rareness. Interestingly, the variational principle shows that most subgraphs in these random graph models are concentrated on vertices with degrees in very specific ranges. We later show in Chapter 8 that the ordering from the most frequently occurring subgraph to the least frequently occurring subgraph in the erased configuration model and the rank-1 inhomogeneous random graph shows strong resemblance with those in several real-world networks.

How do the fluctuations of subgraph counts behave? (Chapter 8)

We investigate the fluctuations in subgraph counts across different samples of rank-1 inhomogeneous random graphs. Interestingly, some subgraph counts turn out to be self-averaging, i.e., the variance of the subgraph counts is small compared to its second moment. This means that one large network sample suffices to investigate the typical number of such subgraphs in random graph models. When considering the random graph model as a null model for a real-world network, this also allows one to compare subgraph counts in model and reality in a more effective way. Other subgraphs are non-self-averaging, showing wild fluctuations of subgraph counts between different samples of inhomogeneous random graphs.

Can we efficiently find an induced subgraph in a large scale-free network? (Chapter 10)

We first propose an algorithm that finds a small induced subgraph of a rank-1 inhomogeneous random graph in linear time. The algorithm exploits the subgraph structure of the rank-1 inhomogeneous random graph by only looking for subgraphs on vertices of a specific degree-range. We then show that this algorithm may also find subgraphs of real-world scale-free networks efficiently when the algorithm searches for subgraphs on vertices of degrees within some network-dependent degree-range.

1.4.3 Part 3: Community structures

The third part of this thesis considers networks with a mesoscopic community structure, and aims to answer the following questions:

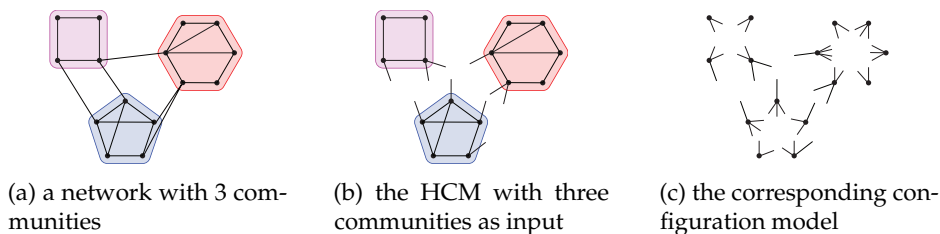


Figure 1.8: Illustration of the HCM. The half-edges in (b) and (c) are connected randomly.

How can we model networks with general community structures? (Chapter 11)

We introduce a random graph model that is able to incorporate a general community structure, which we call the hierarchical configuration model (HCM), illustrated in Figure 1.8. Where the configuration model is able to investigate the influence of the degree sequence on network properties, the HCM enables us to study the influence of community structures on network properties. The HCM behaves as a configuration model on the community level, making the model mathematically tractable.

What do real-world communities look like? (Chapter 12)

We investigate real-world networks through the lens of the HCM model, which reveals several interesting previously unobserved insights in the community structure of real-world networks. We find that there is a power-law relation between the community sizes and their densities. We observe a second power-law relation between the community sizes and the number of edges leaving a community. We show that for densely connected communities, a power-law shift is present. When the degree distribution follows a power-law with exponent τ , the community sizes follow a power-law with degree exponent $\tau - 1$.

How do communities affect epidemic processes? (Chapters 11, 13 and 14)

We first focus on the behavior of the HCM under bond percolation. By comparing HCM under bond percolation to the behavior of bond percolation on a configuration model with the same degree sequence, we find how community structures affect bond percolation. In Chapter 11 we show that the HCM undergoes a similar phase transition under percolation as the configuration model. Furthermore, compared to a configuration model with the same degrees, community structures may either increase, or decrease the critical percolation probability p_c , depending on the exact community shapes. We then focus on the critical regime, where the network is at the point of breaking down, in Chapter 14. We show that the community structure does not significantly affect the component sizes at criticality as long as the communities are smaller than $n^{2/3}$. We then investigate numerically the behavior of several other epidemic processes on the HCM in Chapter 13, using real-world community structures as input for the HCM model. Interestingly, some community structures speed up the spread of an epidemic, but other community structures slow it down. This illustrates the importance of using community structures in network models.

Summary. In this thesis, we analyze degree correlations, local clustering and the presence of subgraphs in several random graph models. We compare the behavior of these properties with those on real-world network data. We further introduce a new random graph model to include mesoscopic community sizes, and investigate the effect of these community structures on several network properties.

Part I

Clustering and correlations

2 Degree-degree correlations

Based on:

Degree-degree correlations in scale-free network models
C. Stegehuis, arXiv:1709.01085

In this chapter, we study the average nearest-neighbor degree $a(k)$ of vertices with degree k . In many real-world networks with power-law degree distribution $a(k)$ falls off in k , a property ascribed to the constraint that any two vertices are connected by at most one edge. We show that $a(k)$ indeed decays in k in three simple random graph models with power-law degrees: the erased configuration model, the rank-1 inhomogeneous random graph and the hyperbolic random graph. We find that for all three random graph models $a(k)$ starts to decay beyond $n^{(\tau-2)/(\tau-1)}$ and then settles on a power law $a(k) \sim k^{\tau-3}$ in the large-network limit, with τ the degree exponent.

2.1 Introduction

There exists a vast array of papers, empirical, non-rigorous and rigorous, on the function $k \mapsto a(k)$ [16, 17, 35, 36, 57, 148, 175, 176, 210, 226] which describes the correlation between the degrees on the two sides of an edge. This classifies the network into one of the following three categories [160]. When $a(k)$ increases with k , the network is said to be *assortative*: vertices with high degrees mostly connect to other vertices with high degrees. When $a(k)$ decreases in k , the network is said to be *disassortative*. Then high-degree vertices typically connect to low-degree vertices. When $a(k)$ forms a flat curve in k , the network is said to be *uncorrelated*. In this case, the degrees on the two different sides of an edge can be viewed as independent of each other, a desirable property when studying the mathematical properties of networks. But the fact is that the majority of real-world networks with power-law degrees and unbounded degree fluctuations ($\tau \in (2, 3)$) show a clear decay of $a(k)$ as k grows large [146, 176], as illustrated in Figure 1.4. Hence, scale-free networks are inherently disassortative, and large-degree vertices (hubs) are predominantly connected to small-degree vertices. In complex network theory, such a well-established empirical fact then asks for a theoretical explanation. Typically, this explanation comes in the form of a null model that matches the degree distribution and has the empirical observation as a property, in this case disassortivity, or more specifically, the essential features of the curve $k \mapsto a(k)$.

The configuration model generates random networks with any prescribed degree distribution, but only results in uncorrelated networks when including self-loops and multiple edges. Hence, the configuration model can never explain the $a(k)$ fall-off.

We therefore resort to different null models that, contrary to the configuration model, generate random networks without self-loops and multiple edges. The resulting *simple* random networks are therefore prone to the structural correlations that come with the presence of hubs. We study $a(k)$ for three widely used random graph models: the erased configuration model, the rank-1 inhomogeneous random graph (also called hidden-variable model) and the hyperbolic random graph. We show that these models display universal $a(k)$ -behavior: For k sufficiently small, $a(k)$ is independent of k . Thus, in simple scale-free networks, neighbors of small-degree vertices are similar. We then identify the value of k for which $a(k)$ starts decaying. An intuitive explanation for the $a(k)$ fall-off is that in simple networks, high-degree vertices have so many neighbors that they must reach out to lower-degree vertices, because networks typically only contain a small amount of high-degree vertices. This causes the average degree of a neighbor of a high-degree vertex to be smaller. Thus, single-edge constraints may cause the decay of $a(k)$.

According to several studies the degree-degree correlation measure $a(k)$ can largely explain the fall-off of the clustering spectrum $k \mapsto c(k)$ [35, 57, 198]. In this chapter, we provide support for this statement, by identifying an explicit relation between $a(k)$ and $c(k)$ for large k . But the main goal of this chapter is to explain the full spectrum $k \mapsto a(k)$ for all k , and to provide theoretical underpinning for the widely observed $a(k)$ fall-off.

2.2 Main results

We first define the average nearest-neighbor degree $a(k, G)$ of a graph G in more detail. Let $(D_i)_{i \in [n]}$ be the degree sequence of the graph, where $[n] = \{1, \dots, n\}$. Furthermore, let N_k denote the total number of degree k vertices in the graph, and \mathcal{N}_i denote the neighborhood of vertex i . The average nearest-neighbor degree of graph G is then defined as

$$a(k, G) = \frac{1}{kN_k} \sum_{i:D_i=k} \sum_{j \in \mathcal{N}_i} D_j. \quad (2.2.1)$$

It is possible that no vertex of degree k exists in the graph. We therefore analyze

$$a_\varepsilon(k, G) = \frac{1}{k|M_\varepsilon(k)|} \sum_{i \in M_\varepsilon(k)} \sum_{j \in \mathcal{N}_i} D_j, \quad (2.2.2)$$

where $M_\varepsilon(k) = \{i \in [n] : D_i \in [k(1 - \varepsilon), k(1 + \varepsilon)]\}$. When no vertex in $M_\varepsilon(k)$ exists, we set $a_\varepsilon(k, G) = 0$. We will show that in the models we analyze, $M_\varepsilon(k)$ is non-empty with high probability, so that $a_\varepsilon(k, G)$ is well defined. Note that $a(k, G) = a_0(k, G)$. We now analyze $a_\varepsilon(k, G)$, first for the erased configuration model in Subsection 2.2.1 and then for the rank-1 inhomogeneous random graph and the hyperbolic random graph in Sections 2.2.3 and 2.2.4.

2.2.1 The erased configuration model

We first investigate $a(k, G)$ for the erased configuration model defined in Section 1.1.3. In particular, we take the original degree sequence (D_1, \dots, D_n) to be an i.i.d. sample

from the distribution

$$\mathbb{P}(D = k) = ck^{-\tau}, \quad \text{when } k \rightarrow \infty, \quad (2.2.3)$$

where $\tau \in (2, 3)$ so that $\mathbb{E}[D^2] = \infty$. We denote $\mathbb{E}[D] = \mu$. When this sample constructs a degree sequence such that the sum of the degrees is odd, we add an extra half-edge to the last vertex. This does not affect our computations. We denote the actual degree sequence of the graph after merging the multiple edges and self-loops by $(D^{(er)})_{i \in [n]}$, and we call these the resulting degrees.

Stable random variables. The limit theorem of $a(k, G_n)$ for the erased configuration model contains stable random variables. Stable random variables can be parametrized by four parameters, and are usually denoted by $\mathcal{S}_\alpha(\sigma, \beta, \mu)$ (see for example [217, Chapter 4]). Throughout this chapter, we will only use stable distributions with $\sigma = 1, \beta = 1, \mu = 0$ and we denote $\mathcal{S}_\alpha = \mathcal{S}_\alpha(1, 1, 0)$ to ease notation. The probability density functions of general stable random variables cannot be written analytically in general. The characteristic function of \mathcal{S}_α can be written as [217, Chapter 4]

$$\log \left(\mathbb{E} \left[e^{it\mathcal{S}_\alpha} \right] \right) = \begin{cases} |t|^\alpha (1 - i \operatorname{sign} \theta) \tan \left(\frac{\pi\alpha}{2} \right) & \alpha \neq 1 \\ |t| \left(1 + \frac{2i}{\pi} \operatorname{sign} \theta \right) \log (|\theta|) & \alpha = 1. \end{cases} \quad (2.2.4)$$

We now state the main result for the erased configuration model:

Theorem 2.1 ($a_{\varepsilon_n}(k, G_n)$ in the erased configuration model). *Let $(G_n)_{n \geq 1}$ be a sequence of erased configuration models on n vertices, where the degrees are an i.i.d. sample from (2.2.3). Take ε_n such that $\lim_{n \rightarrow \infty} \varepsilon_n = 0$ and $\lim_{n \rightarrow \infty} nk^{-(\tau-1)\varepsilon_n} = \infty$ and let Γ denote the Gamma function.*

(i) For $1 \ll k \ll n^{(\tau-2)/(\tau-1)}$,

$$\frac{a_{\varepsilon_n}(k, G_n)}{n^{(3-\tau)/(\tau-1)}} \xrightarrow{d} \frac{1}{\mu} \left(\frac{2c\Gamma(\frac{5}{2} - \frac{1}{2}\tau)}{(\tau-1)(3-\tau)} \cos \left(\frac{\pi(\tau-1)}{4} \right) \right)^{2/(\tau-1)} \mathcal{S}_{(\tau-1)/2}, \quad (2.2.5)$$

where $\mathcal{S}_{(\tau-1)/2}$ is a stable random variable.

(ii) For $n^{(\tau-2)/(\tau-1)} \ll k \ll n^{1/(\tau-1)}$,

$$\frac{a_{\varepsilon_n}(k, G_n)}{n^{3-\tau}k^{\tau-3}} \xrightarrow{\mathbb{P}} -c\mu^{2-\tau}\Gamma(2-\tau). \quad (2.2.6)$$

Remark 2.1. The convergence in (2.2.5) also holds jointly in k and n , so that for fixed $m \geq 1$ and $1 \leq k_1 < k_2 < \dots < k_m \ll n^{(\tau-2)/(\tau-1)}$,

$$\frac{(a_{\varepsilon_n}(k_i, G_n))_{i \in [m]}}{n^{(3-\tau)/(\tau-1)}} \xrightarrow{d} \frac{1}{\mu} \left(\frac{2c\Gamma(\frac{5}{2} - \frac{1}{2}\tau)}{(\tau-1)(3-\tau)} \cos \left(\frac{\pi(\tau-1)}{4} \right) \right)^{2/(\tau-1)} \mathcal{S}_{(\tau-1)/2} \mathbf{1}, \quad (2.2.7)$$

where $\mathbf{1} \in \mathbb{R}^m$ is a vector with m entries equal to 1.

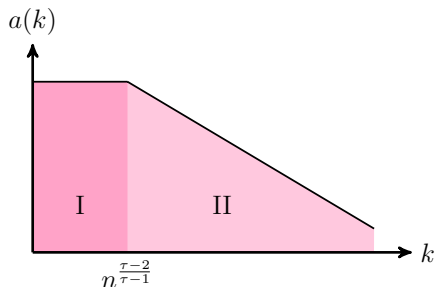


Figure 2.1: Illustration of the behavior of $a_{\varepsilon_n}(k, G_n)$ in the erased configuration model

Figure 2.1 illustrates the behavior of $a_{\varepsilon_n}(k, G_n)$. First, it stays flat and does not depend on k . After that, $a_{\varepsilon_n}(k, G_n)$ starts decreasing in k , which shows that the erased configuration model indeed is a disassortative random graph. Theorem 2.1 shows that $n^{(\tau-2)/(\tau-1)}$ serves as a threshold. Thus, the negative degree-degree correlations due to the single-edge constraint only affect vertices of degrees at least $n^{(\tau-2)/(\tau-1)}$. This can be understood as follows. In the erased configuration model the maximum contribution to $a_{\varepsilon_n}(k, G)$ (see Propositions 2.1 and 2.2) comes from vertices with degrees proportional to n/k . The maximal degree in an observation of n i.i.d. power-law distributed samples is proportional to $n^{1/(\tau-1)}$ w.h.p. Therefore, if $k \ll n^{(\tau-2)/(\tau-1)}$, such vertices with degree proportional to n/k do not exist w.h.p. This explains the two regimes.

For k small, $a_{\varepsilon_n}(k, G_n)$ converges to a stable random variable, as was also shown in [226] for k fixed. Thus, for k small, different instances of the erased configuration model show wild fluctuations. The joint convergence in k explained in Remark 2.1 shows that $a(k, G_n)$ still forms a flat curve in k for one realization of an erased configuration model when k is small. In contrast, $a_{\varepsilon_n}(k, G_n)$ converges to a constant for large k -values, so that different realizations of erased configuration models result in similar $a_{\varepsilon_n}(k, G_n)$ -values for k large.

2.2.2 Sketch of the proof

We now give a heuristic proof of Theorem 2.1. Conditionally on the degrees, the probability that vertices with degrees D_i and D_j are connected in the erased configuration model can be approximated by [108] $1 - \exp(-D_i D_j / \mu n)$. Let $v \in M_{\varepsilon_n}(k)$, and let X_{iv} denote the indicator that vertex i is connected to v . The expected degree of a neighbor of v can then be approximated by

$$a_{\varepsilon_n}(k, G_n) \approx k^{-1} \sum_{i \in [n]} D_i \mathbb{P}(X_{iv} = 1) \approx k^{-1} \sum_{i \in [n]} D_i (1 - e^{-D_i k / (\mu n)}). \quad (2.2.8)$$

The maximum degree in an i.i.d. sample from (2.2.3) scales as $n^{1/(\tau-1)}$ w.h.p.. Thus, as long as $k \ll n^{(\tau-2)/(\tau-1)}$, we can Taylor expand the exponential so that

$$a_{\varepsilon_n}(k, G_n) \approx \frac{1}{\mu n} \sum_{i \in [n]} D_i^2. \quad (2.2.9)$$

Because $(D_i)_{i \in [n]}$ are samples from a power-law distribution with infinite second moment, the Stable Law Central Limit Theorem gives Theorem 2.1(i).

When $k \gg n^{(\tau-2)/(\tau-1)}$, we approximate the sum in (2.2.8) by the integral

$$\begin{aligned} a_{\varepsilon_n}(k, G_n) &\approx cnk^{-1} \int_1^\infty x^{1-\tau} (1 - e^{-xk/(\mu n)}) dx \\ &= c\mu^{2-\tau} \left(\frac{n}{k}\right)^{3-\tau} \int_{k/(\mu n)}^\infty y^{1-\tau} (1 - e^{-y}) dy, \end{aligned} \quad (2.2.10)$$

using the degree distribution (2.2.3) and the change of variables $y = xk/(\mu n)$. When $k \ll n$, we can approximate this by

$$a_{\varepsilon_n}(k, G_n) \approx c\mu^{2-\tau} \left(\frac{n}{k}\right)^{3-\tau} \int_0^\infty y^{1-\tau} (1 - e^{-y}) dy = -c\mu^{2-\tau} \left(\frac{n}{k}\right)^{3-\tau} \Gamma(2 - \tau).$$

The proof of Theorem 2.1(ii) then consists of showing that the above approximations are indeed valid. We prove Theorem 2.1 in detail in Sections 2.3.2 and 2.3.3.

2.2.3 Rank-1 inhomogeneous random graphs

We now turn to the rank-1 inhomogeneous random graph, defined in Section 1.1.4, which constructs simple graphs with soft constraints on the degree sequence. We take the weight sequence of the rank-1 inhomogeneous random graph to be an i.i.d. sample from the power-law distribution (2.2.3). We denote the average value of the weights by μ . For the probability that vertices with weights h and h' connect, we take

$$p(h, h') = \min(hh'/(\mu n), 1), \quad (2.2.11)$$

which is the Chung-Lu version of the rank-1 inhomogeneous random graph [61]. This connection probability ensures that the degree of a vertex with weight h is close to h [35]. We show the following result:

Theorem 2.2 ($a_{\varepsilon_n}(k, G_n)$ in the rank-1 inhomogeneous random graph). *Let $(G_n)_{n \geq 1}$ be a sequence of rank-1 inhomogeneous random graphs on n vertices, where the weights are an i.i.d. sample from (2.2.3). Take ε_n such that $\lim_{n \rightarrow \infty} \varepsilon_n = 0$ and $\lim_{n \rightarrow \infty} n^{-1/(\tau-1)} k \varepsilon_n = \infty$ and let Γ denote the Gamma function.*

(i) For $1 \ll k \ll n^{(\tau-2)/(\tau-1)}$,

$$\frac{a_{\varepsilon_n}(k, G_n)}{n^{(3-\tau)/(\tau-1)}} \xrightarrow{d} \frac{1}{\mu} \left(\frac{2c\Gamma(\frac{5}{2} - \frac{1}{2}\tau)}{(\tau-1)(3-\tau)} \cos\left(\frac{\pi(\tau-1)}{4}\right) \right)^{2/(\tau-1)} \mathcal{S}_{(\tau-1)/2}, \quad (2.2.12)$$

where $\mathcal{S}_{(\tau-1)/2}$ is a stable random variable.

(ii) For $n^{(\tau-2)/(\tau-1)} \ll k \ll n^{1/(\tau-1)}$,

$$\frac{a_{\varepsilon_n}(k, G_n)}{n^{3-\tau} k^{\tau-3}} \xrightarrow{\mathbb{P}} \frac{c\mu^{2-\tau}}{(3-\tau)(\tau-2)}. \quad (2.2.13)$$

Theorem 2.2 is almost identical to Theorem 2.1. The proof of Theorem 2.2 exploits the deep connection between both models, and essentially carries over the results for the erased configuration model to the rank-1 inhomogeneous random graph. The similarity can be understood by noticing that in the erased configuration model the probability that vertices i and j with degrees D_i and D_j are connected can be approximated by $1 - \exp(-D_i D_j / (\mu n))$ which is close to $\min(1, D_i D_j / (\mu n))$, the connection probability in the rank-1 inhomogeneous random graph. Similar arguments that lead to (2.2.8) show that $a_{\varepsilon_n}(k, G_n)$ can be approximated by

$$a_{\varepsilon_n}(k, G_n) \approx k^{-1} \sum_{i \in [n]} h_i \min(h_i k / \mu n, 1). \quad (2.2.14)$$

This sum behaves very similarly to the sum in (2.2.8), so that the only difference between Theorems 2.1 and 2.2 hides in the limiting constants in (2.2.6) and (2.2.13). The main difference between both models is that in the rank-1 inhomogeneous random graph the presence of all edges is independent as soon as the weights are sampled. This is not true in the erased configuration model, because we know that a vertex with sampled degree D_i cannot have more than D_i neighbors, creating dependence between the presence of edges incident to vertex i . We show that these correlations between the presence of different edges in the erased configuration model are small enough for $a_{\varepsilon_n}(k, G_n)$ to behave similarly in the erased configuration model and the rank-1 inhomogeneous random graph.

2.2.4 Hyperbolic random graphs

The third random graph model we consider is the hyperbolic random graph, defined in Section 1.1.6. The hyperbolic random graph creates simple sparse random graphs with power-law degrees, but in contrast to the erased configuration model and the rank-1 inhomogeneous random graph, creates many triangles at the same time due to its geometric nature [56, 130]. In both the rank-1 inhomogeneous random graph and the erased configuration model, the connection probabilities of different pairs of vertices are (almost) independent. In the hyperbolic random graph, this is not true. When u is connected to both v and w , then v and w should also be close to one another by the triangle inequality. However, if we define the type of a vertex as

$$t(u) = e^{(R-r_u)/2} \quad (2.2.15)$$

then we show that we can approximate the probability that vertices u and v are connected as

$$\mathbb{P}(X_{uv} = 1 \mid t(u), t(v)) = \begin{cases} \frac{2}{\pi} \sin^{-1}(vt(u)t(v)/n) & vt(u)t(v)/n < 1, \\ 1 & vt(u)t(v)/n \geq 1, \end{cases} \quad (2.2.16)$$

which behaves similarly as the connection probability in the rank-1 inhomogeneous random graph. Furthermore, by [33, Lemma 1.3], the density of $2 \ln(t(u))$ can be written as

$$f_{2 \ln(t(u))}(x) = \frac{\tau-1}{2} e^{-(\tau-1)x/2} (1 + o(1)), \quad (2.2.17)$$

where the $o(1)$ term is with respect to the network size n . Therefore,

$$\mathbb{P}(t(u) > x) = \mathbb{P}(2 \ln(t(u)) > 2 \ln(x)) = x^{-\tau+1}(1 + o(1)), \quad (2.2.18)$$

so that on a high level the hyperbolic random graph can be interpreted as a rank-1 inhomogeneous random graph with $(t(u))_{u \in [n]}$ as weights (see [33, Section 1.1.1] for a more elaborate discussion).

The next theorem shows that indeed the behavior of $a_{\varepsilon_n}(k, G_n)$ in the hyperbolic random graph is similar as in the rank-1 inhomogeneous random graph:

Theorem 2.3 ($a_{\varepsilon_n}(k, G_n)$ in the hyperbolic random graph). *Let $(G_n)_{n \geq 1}$ be a sequence of hyperbolic random graphs on n vertices with power-law degrees with exponent τ and parameter ν . Take ε_n such that $\lim_{n \rightarrow \infty} \varepsilon_n = 0$ and $\lim_{n \rightarrow \infty} nk^{-(\tau-1)}\varepsilon_n = \infty$ and let Γ denote the Gamma function.*

(i) For $1 \ll k \ll n^{(\tau-2)/(\tau-1)}$,

$$\frac{a_{\varepsilon_n}(k, G_n)}{n^{(3-\tau)/(\tau-1)}} \xrightarrow{d} \frac{2\nu}{\pi} \left(\frac{2}{3-\tau} \Gamma\left(\frac{5}{2} - \frac{1}{2}\tau\right) \cos\left(\frac{\pi(\tau-1)}{4}\right) \right)^{2/(\tau-1)} \mathcal{S}_{(\tau-1)/2}, \quad (2.2.19)$$

where $\mathcal{S}_{(\tau-1)/2}$ is a stable random variable.

(ii) For $n^{(\tau-2)/(\tau-1)} \ll k \ll n^{1/(\tau-1)}$,

$$\frac{a_{\varepsilon_n}(k, G_n)}{n^{3-\tau}k^{\tau-3}} \xrightarrow{\mathbb{P}} \frac{(\tau-1)\nu\sqrt{\pi}\Gamma\left(\frac{3}{2} - \frac{\tau}{2}\right)}{2(\tau-2)\Gamma\left(2 - \frac{\tau}{2}\right)} \left(\frac{2(\tau-1)}{\pi(\tau-2)} \right)^{3-\tau}. \quad (2.2.20)$$

2.2.5 Discussion

Universality. The behavior of $a_{\varepsilon_n}(k, G_n)$ is universal across the three null models we consider. The erased configuration model and the rank-1 inhomogeneous random graph are closely related. They are known to behave similarly for example under critical percolation [24, 26], in terms of distances [78] when $\tau > 3$, and in terms of clustering when $\tau \in (2, 3)$ (see Chapters 4 and 5). The hyperbolic random graph typically shows different behavior, for example in terms of clustering [56, 101], or connectivity [33, 34]. Still, the behavior of $a_{\varepsilon_n}(k, G_n)$ is similar in the hyperbolic random graph and the other two null models. In all three null models, the main contribution for $k \gg n^{(\tau-2)/(\tau-1)}$ comes from vertices with degrees proportional to n/k (see Propositions 2.1 and 2.2). In the hyperbolic random graph, we can relate this maximum contribution to the geometry of the hyperbolic sphere. A vertex i of degree k has radius $r_i \approx R - 2 \log(k)$. Similarly, a vertex j of degree $n/(vk)$ has radius $r_j \approx R - 2 \log(n/(vk)) = 2 \log(k)$. Then, $r_j \approx R - r_i$, so that the major contributing vertices have radial coordinate proportional to $R - r_i$.

Expected average nearest-neighbor degree. In Theorems 2.1-2.3 we show that $a_{\varepsilon_n}(k, G_n)$ converges in probability to a stable random variable when k is small. Thus, when we generate many samples of random graphs, for fixed k , the distribution of the values of $a_{\varepsilon_n}(k, G_n)$ across the different samples will look like a stable random

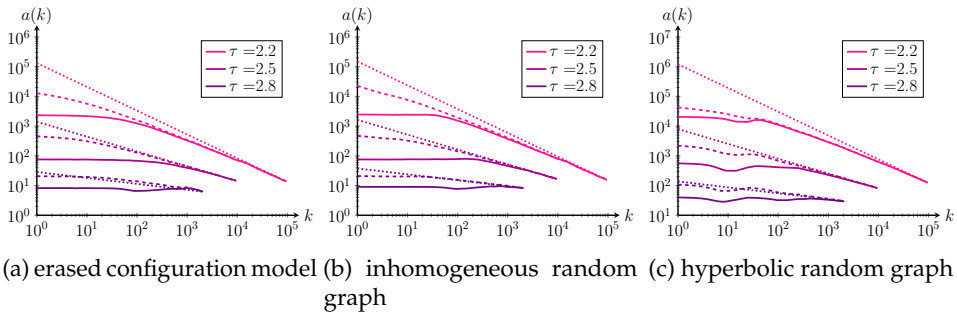


Figure 2.2: $a(k, G_n)$ for different random graph models with $n = 10^6$. The solid line is the median of $a(k, G_n)$ over 10^4 realizations of the random graph, and the dashed line is the average over these realizations. The dotted line is the asymptotic slope $k^{\tau-3}$.

variable. We can also study the expected value of $a(k, G_n)$ across the different samples. For the erased configuration model for example, we can show that (see Section 2.3.4)

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E}[a(k, G_n)]}{(n/k)^{3-\tau}} = -c\mu^{2-\tau}\Gamma(2-\tau). \quad (2.2.21)$$

The difference between the scaling of the expected value of $a(k, G_n)$ and the typical behavior of $a(k, G_n)$ in Theorem 2.1(i) is caused by high-degree vertices. In typical degree sequences, the maximum degree is proportional to $n^{1/(\tau-1)}$. It is unlikely that vertices with higher degrees are present, but if they are, they have a high impact on the average nearest-neighbor degree of low degree vertices, causing the difference between the expected average nearest-neighbor degree and the typical average nearest-neighbor degree. Thus, the expected value of $a(k, G_n)$ is not very informative when k is small, since Theorem 2.1 shows that $a(k, G_n)$ will almost always be smaller than its expected value when k is small.

Figure 2.2 illustrates this difference in terms of the mean and median value of $a(k, G_n)$ over many realizations of the erased configuration model, the rank-1 inhomogeneous random graph and the hyperbolic random graph. Here indeed we see that the expected average neighbor degree scales as a power of k over the entire range of k , where the median shows the straight part of the curve from Theorem 2.1. Thus, it is important to distinguish between mean and median of $a(k, G_n)$ when simulating random graphs.

Vertices of degree k . Definition (2.2.1) assumes that a vertex of degree k is present. For large values of k , this is a rare event, by (2.2.3). Indeed, vertices of degree at most $n^{1/\tau}$ are present with high probability in the erased configuration model, whereas the probability that a vertex of degree $k \gg n^{1/\tau}$ is present tends to zero in the large network limit [226]. We avoid this problem by averaging $a(k, G_n)$ over a small range of degrees. Another option is to condition on the event that a vertex of degree k is present. Our proofs for $k \ll n^{(\tau-2)/(\tau-1)}$ for the erased configuration model can easily be adjusted to condition on this event. For k larger, we leave the behavior of $a(k, G_n)$ conditionally on a vertex of degree k being present open for further research.

Fixed degrees. In the proof of Theorem 2.1 we show that for small k , the fluctuations that come with the stable laws are not present when we condition on the degree sequence. Thus, the large fluctuations in $a_{\varepsilon_n}(k, G_n)$ for small k are caused by fluctuations of the i.i.d. degrees, weights or radii. For a given real-world network, its degrees are often preserved, and many erased configuration models or inhomogeneous random graphs with the same observed degree sequence are created. In this fixed-degree setting, the sample-to-sample fluctuations of $a_{\varepsilon_n}(k, G_n)$ are relatively small.

Relation with local clustering. The local clustering coefficient $c(k)$ of vertices of degree k measures the probability that two randomly chosen neighbors of a randomly chosen vertex of degree k are connected. Chapters 4-6 show that in many real-world networks as well as simple null models, $c(k)$ decreases as a function of k . The relation between the decay rate of $c(k)$ and the decay rate of $a(k)$ has been investigated for the rank-1 inhomogeneous random graph, where it was shown that $c(k) < a(k)/k$ [198]. Using our results for $c(k)$ on the erased configuration model and the rank-1 inhomogeneous random graph that are presented in Chapters 4 and 5, we can make the relation between $c(k)$ and $a(k)$ more precise. In Chapter 5 we show that when $k \gg \sqrt{n}$, $c(k)$ in the erased configuration model satisfies

$$c(k) = c^2 \Gamma(2 - \tau)^2 \mu^{3-2\tau} n^{5-2\tau} k^{2\tau-6} (1 + o_{\mathbb{P}}(1)). \quad (2.2.22)$$

Then, by Theorem 2.1, when $k \gg \sqrt{n}$,

$$c(k) = \frac{a^2(k)}{\mu n} (1 + o_{\mathbb{P}}(1)). \quad (2.2.23)$$

Intuitively, we can see this relationship in the following way. Pick two neighbors of a vertex with degree k . By definition, these vertices have degree $a(k)$ on average. Since $k \gg \sqrt{n}$, by Theorem 2.2 $a(k) \ll \sqrt{n}$. Therefore, the probability of two vertices with weight $a(k)$ to be connected is approximately $1 - \exp(-a(k)^2/\mu n) \approx a(k)^2/\mu n$. Since the clustering coefficient can be interpreted as the probability that two randomly chosen neighbors are connected, the clustering coefficient should satisfy $c(k) \approx a(k)^2/\mu n$ when $k \gg \sqrt{n}$. In particular, the decay of the clustering coefficient should be twice as fast as the decay of the average neighbor degree. Analytical results on $c(k)$ on the rank-1 inhomogeneous random graph (see Chapter 4) show that (2.2.23) is also the correct relation between clustering and degree correlations in the rank-1 inhomogeneous random graph. In Chapter 6 we show that the difference between expectation and typical behavior that is present in $a(k)$ also occurs for the local clustering coefficient $c(k)$.

Correlations in the hyperbolic random graph. The relation in (2.2.23) is based on the fact that in the erased configuration model and the rank-1 inhomogeneous random graph the connection probabilities of pairs of vertices (i, j) , (i, k) and (j, k) are (almost) independent. In the hyperbolic random graph, the geometry causes a strong dependence between these connection probabilities. If vertices j and k are neighbors of i , then they are likely to be geometrically close to one another due to the triangle inequality. This makes the probability that j and k are connected larger

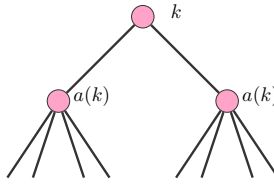


Figure 2.3: The neighbors of a vertex of degree k have average degree $a(k)$

than in the rank-1 inhomogeneous random graph or the erased configuration model. These correlations do not play a role when computing $a(k, G_n)$, since it only involves the connection probability of two different vertices. When computing statistics of the hyperbolic random graph that include three-point correlations, the equivalence between the hyperbolic random graph and the rank-1 inhomogeneous random graph may fail to hold, as in the example of $c(k)$.

Interestingly, the number of large cliques was shown to be similar in the hyperbolic random graph, the rank-1 inhomogeneous random graph and the erased configuration model [85], even though cliques clearly involve three-point correlations. Large cliques in the hyperbolic random graph are typically formed between vertices at radius proportional to $R/2$ [85], so that their degrees are proportional to \sqrt{n} [33]. These vertices form a dense core, which is very similar to what happens in the erased configuration model and the rank-1 inhomogeneous random graph [118]. In the erased configuration model, many other small subgraphs typically occur between vertices of degrees proportional to \sqrt{n} (see Chapter 7). It would be interesting to see whether the number of these small subgraphs behaves similarly in the hyperbolic random graph.

2.3 Average nearest-neighbor degree in the ECM

In this section, we prove Theorem 2.1. For $k \ll n^{(\tau-2)/(\tau-1)}$, we couple the degrees of neighbors of a uniformly chosen vertex of degree k to i.i.d. samples of the size-biased degree distribution in Section 2.3.2. When $k \gg n^{(\tau-2)/(\tau-1)}$, this coupling is no longer valid. We then show in Section 2.3.3 that a specific range of degrees contributes most to $a_{\varepsilon_n}(k, G_n)$.

2.3.1 Preliminaries

We say that $X_n = O_{\mathbb{P}}(b_n)$ for a sequence of random variables $(X_n)_{n \geq 1}$ if $|X_n|/b_n$ is a tight sequence of random variables, and $X_n = o_{\mathbb{P}}(b_n)$ if $X_n/b_n \xrightarrow{\mathbb{P}} 0$. Let L_n denote the total number of half-edges, so that $L_n = \sum_i D_i$. We define the events

$$J_n = \{|L_n - \mu n| \leq n^{2/\tau}\}, \quad A_n = \{|M_{\varepsilon_n}(k)| \geq 1\}. \quad (2.3.1)$$

By [111, Lemma 2.3], $\mathbb{P}(J_n) \rightarrow 1$ as $n \rightarrow \infty$. By [50, Theorem 2.1]

$$\mathbb{E}[|M_{\varepsilon_n}(k)|] = cn \int_{k(1-\varepsilon_n)}^{k(1+\varepsilon_n)} x^{-\tau} dx (1 + o(1)) = \tilde{C} n^{-1} k^{1-\tau} \varepsilon_n (1 + o(1)), \quad (2.3.2)$$

for some $\tilde{C} > 0$, so that $\mathbb{P}(A_n) \rightarrow 1$ for $k \ll n^{1/(\tau-1)}$ by the choice of ε_n in Theorem 2.1.

We will often condition on the degree sequence. For some event \mathcal{E} , we use the notation $\mathbb{P}_n(\mathcal{E}) = \mathbb{P}_n(\mathcal{E} \mid (D_i)_{i \in [n]})$, and we define \mathbb{E}_n and Var_n similarly. We often want to interchange the sampled degree of a vertex i , D_i and its resulting degree $D_i^{(\text{er})}$. The next lemma shows that D_i and $D_i^{(\text{er})}$ are close:

Lemma 2.1. *Let G be an erased configuration model where the degrees are i.i.d. samples from a power-law distribution with $\tau \in (2, 3)$. Then, for $D_i = o(n)$,*

$$D_i^{(\text{er})} = D_i(1 + o_{\mathbb{P}}(1)). \quad (2.3.3)$$

Proof. By [108, Eq. (4.9)]

$$\mathbb{P}_n(X_{ij} = 0) \leq \prod_{s=0}^{D_i-1} \left(1 - \frac{D_i}{L_n - 2D_i - 1}\right) + \frac{D_i^2 D_j}{(L_n - 2D_i)^2} \leq e^{-D_i D_j / L_n} + \frac{D_i^2 D_j}{(L_n - 2D_i)^2}.$$

Let $\psi(x) = x - 1 + e^{-x}$. The expected number of erased edges at vertex i satisfies

$$\begin{aligned} \mathbb{E}_n \left[D_i - D_i^{(\text{er})} \right] &= D_i - \sum_{j \in [n]} (1 - \mathbb{P}_n(X_{ij} = 0)) \\ &\leq \sum_{j \in [n]} \left(\frac{D_i D_j}{L_n} - 1 + e^{-D_i D_j / L_n} - \frac{D_i^2 D_j}{(L_n - 2D_i)^2} \right) \\ &= \sum_{j \in [n]} \psi \left(\frac{D_i D_j}{L_n} \right) + O \left(D_i^2 / L_n \right) \\ &= \sum_{j \in [n]} \psi \left(\frac{D_i D_j}{\mu n} \right) (1 + o_{\mathbb{P}}(1)) + O_{\mathbb{P}} \left(D_i^2 / n \right). \end{aligned} \quad (2.3.4)$$

By [112, Theorem 3(ii)], for D distributed as in (2.2.3), $\mathbb{E}[\psi(D/t)] = O(t^{-(\tau-1)})$ so that

$$\mathbb{E}[\psi(D_i D_j / (\mu n)) \mid D_i] = O(D_i^{\tau-1} n^{1-\tau}). \quad (2.3.5)$$

This shows that $\sum_{j=1}^n \psi(D_i D_j / L_n) = O_{\mathbb{P}}(D_i^{\tau-1} n^{2-\tau})$ and therefore.

$$\mathbb{E}_n \left[D_i - D_i^{(\text{er})} \right] = O_{\mathbb{P}} \left(n^{-\tau+2} D_i^{\tau-1} \right) + O_{\mathbb{P}} \left(D_i^2 / n \right). \quad (2.3.6)$$

Thus, we obtain using the Markov inequality

$$\mathbb{P}_n \left(D_i - D_i^{(\text{er})} > \varepsilon D_i \right) \leq O_{\mathbb{P}} \left(\varepsilon^{-1} (D_i / n)^{\tau-2} \right) + O_{\mathbb{P}} \left(\varepsilon^{-1} D_i / n \right), \quad (2.3.7)$$

so that for $D_i = o(n)$, $D_i^{(\text{er})} = D_i(1 + o_{\mathbb{P}}(1))$. \square

2.3.2 Small k : coupling to i.i.d. random variables

In this section we investigate the behavior of $a_{\varepsilon_n}(k, G_n)$ when $k = o(n^{(\tau-2)/(\tau-1)})$. We first pick a random vertex $v \in M_{\varepsilon_n}(k)$. We couple the degrees of the neighbors of v to i.i.d. copies of the size-biased degree distribution D_n^* , where

$$\mathbb{P}_n(D_n^* = j) = \frac{j}{L_n} \sum_{i \in [n]} \mathbb{1}_{\{D_i = j\}}. \quad (2.3.8)$$

We then use this coupling to compute $a_{\varepsilon_n}(k, G_n)$.

Proof of Theorem 2.1(i). We first condition on the degree sequence $(D_i)_{i \in [n]}$. Furthermore, we condition on the event J_n and A_n defined in (2.3.1), since $\mathbb{P}(J_n \cap A_n) \rightarrow 1$. Let v be a vertex of degree d . In the configuration model, neighbors of v are constructed by pairing the half-edges of v uniformly to other half-edges. The distribution of the degree of a vertex attached to a uniformly chosen half-edge is given by D_n^* . However, the degrees of the neighbors of v in the erased configuration model are not an i.i.d. sample of D_n^* due to the fact that the half-edges should attach to distinct vertices that are different from v . We now show that we can still approximate the degrees of the neighbors of v by an i.i.d. sample of D_n^* using a coupling argument. Denote the neighbors of v by $(v_i)_{i \in [d]}$ and their degrees by B_1, \dots, B_d . Let Y_1, \dots, Y_d be i.i.d. samples of D_n^* . These samples can be obtained by sampling uniform half-edges with replacement and setting $Y_i = D_{v'_i}$, where v'_i denotes the vertex incident to the i th drawn half-edge. We use a similar coupling as in [21, Construction 4.2] to couple B_i to Y_i . Let $(v'_i)_{i \in [d]}$ denote vertices attached to d uniformly chosen half-edges (with replacement) and set $Y_i = D_{v'_i}$ for $i \in [d]$ and $V_0 = v$. Then for $i \in [d]$ the coupling is defined in the following way:

- If $v'_i \notin V_{i-1}$, then $B_i = Y_i$ and $v_i = v'_i$. Set $V_i = V_{i-1} \cup v'_i$. We say that B_i and Y_i are successfully coupled.
- If $v'_i \in V_{i-1}$, we redraw a uniformly chosen half-edge from the set of half-edges not incident to V_{i-1} . Let v_i denote the vertex incident to the chosen half-edge. Set $B_i = D_{v_i}$ and $V_i = V_{i-1} \cup v_i$. We then say that B_i and Y_i are miscoupled.

Thus, informally, at every step we sample a uniformly chosen half-edge from all half-edges, and select the vertex incident to it. If this vertex is different from all previously selected vertices and unequal to v , we declare the vertex to be a neighbor of v . In this case, we have successfully coupled this neighbor of v to an i.i.d. sample of D_n^* . If not, we need to redraw the selected half-edge to ensure that all neighbors of v are distinct and unequal to v . In this case the neighbor of v and the i.i.d. sample of D_n^* are miscoupled. We now show that the coupling is successful with high probability. By [21, Lemma 4.3], the probability of a miscoupling at step i can be bounded as

$$\mathbb{P}_n(B_i \neq Y_i \mid \mathcal{F}_{i-1}) \leq L_n^{-1} \left(d + \sum_{s=1}^{i-1} B_s \right), \quad (2.3.9)$$

where $\mathcal{F}_i = \sigma(B_j, Y_j)_{j \in [i]}$ denotes the sigma-algebra containing all information about the Y_j and B_j variables encountered up to step i . Thus, the expected number of miscouplings up to time t , $N_{\text{mis}}(t)$, satisfies

$$\mathbb{E}_n [N_{\text{mis}}(t)] \leq \frac{dt}{L_n} + \frac{1}{L_n} \sum_{i=1}^t \sum_{s=1}^{i-1} \mathbb{E}_n [B_s]. \quad (2.3.10)$$

When B_s is successfully coupled,

$$\mathbb{E}_n [B_s \mid B_s \text{ successfully coupled}] = \mathbb{E}_n [D_n^*] = \sum_i D_i^2 / L_n. \quad (2.3.11)$$

When B_s is not successfully coupled, it is drawn in a size-biased manner from the vertices that are not chosen yet. Then, because $D_i \geq 0$ for all i ,

$$\begin{aligned} \mathbb{E}_n [B_s \mid \mathcal{F}_{s-1}, B_s \text{ miscoupled}] &= \frac{\sum_{i \notin V_s} D_i^2}{\sum_{i \notin V_s} D_i} \leq \frac{\sum_{i \in [n]} D_i^2}{\sum_{i \in [n]} D_i - \sum_{i \in V_s} D_i} \\ &= \frac{\sum_{i \in [n]} D_i^2}{\sum_{i \in [n]} D_i} \left(1 + \frac{\sum_{i \in V_s} D_i}{\sum_{i \in [n]} D_i - \sum_{i \in V_s} D_i} \right). \end{aligned} \quad (2.3.12)$$

Since $D_{\max} = O_{\mathbb{P}}(n^{1/(\tau-1)})$, $\sum_{i \in V_s} D_i \leq s D_{\max} = O_{\mathbb{P}}(sn^{1/(\tau-1)})$ for all possible V_s , so that for $s \ll n^{(\tau-2)/(\tau-1)}$

$$\mathbb{E}_n [B_s \mid B_s \text{ miscoupled}] = \frac{\sum_{i \in [n]} D_i^2}{L_n} (1 + o_{\mathbb{P}}(1)). \quad (2.3.13)$$

For t large, we obtain from (2.2.3) that

$$\mathbb{P}(D^2 > t) = \mathbb{P}(D > \sqrt{t}) = \frac{c}{\tau-1} t^{(1-\tau)/2} (1 + o(1)). \quad (2.3.14)$$

Using (2.3.14) we can use the Stable Law Central Limit Theorem (see for example [217, Theorem 4.5.2]) to conclude that

$$\frac{\sum_{i \in [n]} D_i^2}{n^{2/(\tau-1)} \left(\frac{2c}{(\tau-1)(3-\tau)} \Gamma\left(\frac{5}{2} - \frac{1}{2}\tau\right) \cos\left(\frac{\pi(\tau-1)}{4}\right) \right)^{2/(\tau-1)}} \xrightarrow{d} \mathcal{S}_{(\tau-1)/2}, \quad (2.3.15)$$

where $\mathcal{S}_{(\tau-1)/2}$ is a stable random variable. Thus, as long as $s = o(n^{(\tau-2)/(\tau-1)})$,

$$\mathbb{E}_n [B_s] = L_n^{-1} \sum_{i \in [n]} D_i^2 (1 + o_{\mathbb{P}}(1)) = O_{\mathbb{P}} \left(n^{(3-\tau)/(\tau-1)} \right). \quad (2.3.16)$$

By (2.3.10), for $d = o(n^{(\tau-2)/(\tau-1)})$

$$\mathbb{E}_n [N_{\text{mis}}(d)] = \frac{d^2}{L_n} + \frac{1}{L_n} O_{\mathbb{P}} \left(n^{(3-\tau)/(\tau-1)} \right) \sum_{i=1}^d (i-1) = O_{\mathbb{P}} \left(d^2 n^{2\frac{\tau-2}{\tau-1}} \right). \quad (2.3.17)$$

Thus, as long as $d = o(n^{(\tau-2)/(\tau-1)})$, $\mathbb{E}_n [N_{\text{mis}}(d)] = o_{\mathbb{P}}(1)$. Then, by the Markov inequality

$$\mathbb{P}_n (N_{\text{mis}}(d) = 0) = 1 - \mathbb{P}_n (N_{\text{mis}}(d) \geq 1) \geq 1 - \mathbb{E}_n [N_{\text{mis}}(d)] = 1 - o_{\mathbb{P}}(1). \quad (2.3.18)$$

Hence, when $d = o(n^{(\tau-2)/(\tau-1)})$, we can approximate the sum of the degrees of the neighbors of a vertex with degree d by i.i.d. samples of the size-biased degree distribution.

Let $i \in M_{\varepsilon_n}(k)$ for $k = o(n^{(\tau-2)/(\tau-1)})$. Then, $D_i^{(\text{er})} = D_i(1 + o_{\mathbb{P}}(1)) = k(1 + o_{\mathbb{P}}(1))$. Thus, conditionally on the degree sequence

$$\begin{aligned} a_{\varepsilon_n}(k, G_n) &= \frac{1}{k|M_{\varepsilon_n}(k)|} \sum_{i \in M_{\varepsilon_n}(k)} \sum_{j \in \mathcal{N}_i} D_j^{(\text{er})} = \frac{1}{k} \mathbb{E}_n \left[\sum_{j \in \mathcal{N}_{V_k}} D_j^{(\text{er})} \right] \\ &= \mathbb{E}_n [D_{\mathcal{N}_{V_k}}^{(\text{er})}(U)] = (1 + o_{\mathbb{P}}(1)) \mathbb{E}_n [D_{\mathcal{N}_{V_k}}(U)], \end{aligned} \quad (2.3.19)$$

where V_k denotes a uniformly chosen vertex in $M_{\varepsilon_n}(k)$, and $\mathcal{N}_{V_k}(U)$ is a uniformly chosen neighbor of vertex V_k . Here the second equality holds because the average nearest-neighbor degree averages over all neighbors of vertex j . The third equality holds because it also averages over all vertices in $M_{\varepsilon_n}(k)$, together with the fact that $D_{V_k} = k(1 + o(1))$ and $D_i^{(\text{er})} = D_i(1 + o_{\mathbb{P}}(1))$ uniformly in i . With high probability, we can couple the degrees of neighbors of a uniformly chosen vertex of degree in $[k(1 - \varepsilon_n), k(1 + \varepsilon_n)]$ to i.i.d. copies of D_n^* . Then, conditionally on the degree sequence,

$$a_{\varepsilon_n}(k, G_n) = (1 + o_{\mathbb{P}}(1)) \mathbb{E}_n [D_n^*] = (1 + o_{\mathbb{P}}(1)) L_n^{-1} \sum_{i \in [n]} D_i^2. \quad (2.3.20)$$

Combining this with (2.3.15) results in

$$\frac{a_{\varepsilon_n}(k, G_n)}{n^{(3-\tau)/(\tau-1)}} \xrightarrow{d} \frac{1}{\mu} \left(\frac{2c\Gamma(\frac{5}{2} - \frac{1}{2}\tau)}{(\tau-1)(3-\tau)} \cos \left(\frac{\pi(\tau-1)}{4} \right) \right)^{2/(\tau-1)} \mathcal{S}_{(\tau-1)/2}. \quad (2.3.21)$$

To prove the joint convergence of Remark 2.1, note that for fixed m , the probability that $M_{\varepsilon_n}(k_i)$ is non-empty for all $i \in [m]$ tends to 1, by a similar proof as the proof that A_n defined in (2.3.1) satisfies $\mathbb{P}(A_n) \rightarrow 1$. Thus, we may condition on the event that $M_{\varepsilon_n}(k_i)$ is non-empty for all $i \in [m]$. Then, the fact that (2.3.20) is the same for all $k_i \ll n^{(\tau-2)/(\tau-1)}$ proves the joint convergence. \square

2.3.3 Large k

Now we study the value of $a_{\varepsilon_n}(k, G_n)$ when $k \gg n^{(\tau-2)/(\tau-1)}$. We show that there exists a range of degrees $W_n^k(\delta)$ which gives the largest contribution to $a_{\varepsilon_n}(k, G_n)$. For ease of notation, we write $a_{\varepsilon_n}(k)$ for $a_{\varepsilon_n}(k, G_n)$ in this section. We define

$$W_n^k(\delta) = \{u : D_u \in [\delta\mu n/k, \mu n/(\delta k)]\}, \quad (2.3.22)$$

and we write

$$\begin{aligned} a_{\varepsilon_n}(k) &= \frac{1}{k|M_{\varepsilon_n}(k)|} \sum_{i \in M_{\varepsilon_n}(k)} \sum_{j \in W_n^k(\delta)} D_j^{(\text{er})} + \frac{1}{k|M_{\varepsilon_n}(k)|} \sum_{i \in M_{\varepsilon_n}(k)} \sum_{j \notin W_n^k(\delta)} D_j^{(\text{er})} \\ &=: a_{\varepsilon_n}(k, W_n^k(\delta)) + a_{\varepsilon_n}(k, \bar{W}_n^k(\delta)), \end{aligned} \quad (2.3.23)$$

where $a_{\varepsilon_n}(k, W_n^k(\delta))$ denotes the contribution to $a_{\varepsilon_n}(k)$ from vertices in $W_n^k(\delta)$, and $a_{\varepsilon_n}(k, \bar{W}_n(\varepsilon))$ the contribution from vertices not in $W_n^k(\delta)$. In the rest of this section, we prove the following two propositions, which together show that the largest contribution to $a_{\varepsilon_n}(k)$ indeed comes from vertices in $W_n^k(\delta)$.

Proposition 2.1 (Minor contributions). *There exists $\kappa > 0$ such that for $k \ll n^{1/(\tau-1)}$,*

$$\limsup_{n \rightarrow \infty} \frac{\mathbb{E} \left[a_{\varepsilon_n}(k, \bar{W}_n^k(\delta)) \right]}{(n/k)^{3-\tau}} = O(\delta^\kappa). \quad (2.3.24)$$

Proposition 2.2 (Major contributions). *For $k \gg n^{(\tau-2)/(\tau-1)}$,*

$$\frac{a_{\varepsilon_n}(k, W_n^k(\delta))}{(n/k)^{3-\tau}} \xrightarrow{\mathbb{P}} c\mu^{2-\tau} \int_{\delta}^{1/\delta} x^{1-\tau} (1 - e^{-x}) dx. \quad (2.3.25)$$

We now show how these propositions prove part (ii) of Theorem 2.1.

Proof of Theorem 2.1 (ii). By the Markov inequality and Proposition 2.1,

$$\frac{a_{\varepsilon_n}(k, \bar{W}_n^k(\delta))}{(n/k)^{3-\tau}} = O_{\mathbb{P}}(\delta^\kappa). \quad (2.3.26)$$

Combining this with Proposition 2.2 results in

$$\frac{a_{\varepsilon_n}(k)}{(n/k)^{3-\tau}} \xrightarrow{\mathbb{P}} c\mu^{2-\tau} \int_{\delta}^{1/\delta} x^{1-\tau} (1 - e^{-x}) dx + O_{\mathbb{P}}(\delta^\kappa). \quad (2.3.27)$$

Taking the limit of $\delta \rightarrow 0$ then proves the theorem. \square

The rest of this section is devoted to proving Propositions 2.1 and 2.2.

Conditional expectation.

We first compute the expectation of $a_{\varepsilon_n}(k, W_n^k(\delta))$ conditionally on the degree sequence.

Lemma 2.2. *When $k \gg n^{(\tau-2)/(\tau-1)}$,*

$$\mathbb{E}_n \left[a_{\varepsilon_n}(k, W_n^k(\delta)) \right] = \frac{1}{k} \sum_{u \in W_n^k(\delta)} D_u (1 - e^{-D_u k/L_n}) (1 + o_{\mathbb{P}}(1)). \quad (2.3.28)$$

Proof. It suffices to prove the lemma under the event J_n from (2.3.1), since $\mathbb{P}(J_n) \rightarrow 1$. Thus we may assume that $L_n = \mu n(1 + o(1))$. Let X_{ij} denote the indicator that i and j are connected. By (2.3.23)

$$\mathbb{E}_n \left[a_{\varepsilon_n}(k, W_n^k(\delta)) \right] = \frac{1}{k |M_{\varepsilon_n}(k)|} \sum_{v \in M_{\varepsilon_n}(k)} \sum_{u \in W_n^k(\delta)} D_u^{(\text{er})} \mathbb{P}_n(X_{uv} = 1). \quad (2.3.29)$$

By [108, Eq. (4.9)]

$$\mathbb{P}_n(X_{uv} = 1) = 1 - e^{-D_u D_v / L_n} + O\left(\frac{D_v^2 D_u + D_u^2 D_v}{L_n^2}\right). \quad (2.3.30)$$

Thus, when $D_u \in \mu n/k[\delta, 1/\delta]$ and $v \in M_{\varepsilon_n}(k)$,

$$\mathbb{P}_n(X_{uv} = 1) = (1 - e^{-D_u k / L_n})(1 + o_{\mathbb{P}}(1)). \quad (2.3.31)$$

Using (2.3.6), we obtain

$$\mathbb{E}_n\left[\sum_{u \in W_n^k(\delta)} D_v^{(\text{er})} - D_v\right] = \sum_{u \in W_n^k(\delta)} O_{\mathbb{P}}\left(D_v^{\tau-1} n^{2-\tau}\right) = |W_n^k(\delta)| O_{\mathbb{P}}\left(n k^{1-\tau}\right). \quad (2.3.32)$$

Since $\sum_{u \in W_n^k(\delta)} D_u = |W_n^k(\delta)| O(n/k)$, also $\sum_{u \in W_n^k(\delta)} D_u^{(\text{er})} = (1 + o_{\mathbb{P}}(1)) \sum_{u \in W_n^k(\delta)} D_u$. Because $1 - e^{-D_u k / L_n} \in [1 - e^{-\delta}, 1 - e^{-1/\delta}]$ when $u \in W_n^k(\delta)$, this also shows that

$$\sum_{u \in W_n^k(\delta)} D_u^{(\text{er})} (1 - e^{-D_u k / L_n}) = (1 + o_{\mathbb{P}}(1)) \sum_{u \in W_n^k(\delta)} D_u (1 - e^{-D_u k / L_n}). \quad (2.3.33)$$

Thus, we obtain

$$\begin{aligned} \mathbb{E}_n\left[a_{\varepsilon_n}(k, W_n^k(\delta))\right] &= \frac{1}{k} \sum_{u \in W_n^k(\delta)} D_u^{(\text{er})} (1 - e^{-D_u k / L_n}) (1 + o_{\mathbb{P}}(1)) \\ &= \frac{1}{k} \sum_{u \in W_n^k(\delta)} D_u (1 - e^{-D_u k / L_n}) (1 + o_{\mathbb{P}}(1)). \end{aligned} \quad (2.3.34)$$

□

Convergence of conditional expectation.

We now show that $\mathbb{E}_n[a_{\varepsilon_n}(k, W_n^k(\delta))]$ of Lemma 2.2 converges to a constant.

Lemma 2.3. *When $k \gg n^{(\tau-2)/(\tau-1)}$,*

$$\frac{\mathbb{E}_n\left[a_{\varepsilon_n}(k, W_n^k(\delta))\right]}{n^{3-\tau} k^{\tau-3}} \xrightarrow{\mathbb{P}} c \mu^{2-\tau} \int_{\delta}^{1/\delta} x^{1-\tau} (1 - e^{-x}) dx. \quad (2.3.35)$$

Proof. Define the random measure

$$M^{(n)}[a, b] = \frac{1}{\mu^{1-\tau} n^{2-\tau} k^{\tau-1}} \sum_{u \in [n]} \mathbb{1}_{\{D_u \in [a, b] \mu n/k\}}. \quad (2.3.36)$$

Since the degrees are i.i.d. samples from (2.2.3), the number of vertices with degrees in the interval $[a, b]$ is binomially distributed. Then,

$$\begin{aligned} M^{(n)}[a, b] &= \frac{1}{\mu^{1-\tau} n^{2-\tau} k^{\tau-1}} |\{u : D_u \in [a, b] \mu n/k\}| \xrightarrow{\mathbb{P}} \frac{\mathbb{P}(D \in [a, b] \mu n/k)}{(\mu n)^{1-\tau} k^{\tau-1}} \\ &= \frac{1}{(\mu n)^{1-\tau} k^{\tau-1}} \int_{a \mu n/k}^{b \mu n/k} c x^{-\tau} dx = \int_a^b c y^{-\tau} dy =: \lambda[a, b], \end{aligned} \quad (2.3.37)$$

where we have used the change of variables $y = xk/(\mu n)$. By Lemma 2.2,

$$\begin{aligned} \mathbb{E}_n \left[a_{\varepsilon_n}(k, W_n^k(\delta)) \right] &= \frac{\sum_{u \in W_n^k(\delta)} D_u (1 - e^{-D_u k / L_n})}{k} (1 + o_{\mathbb{P}}(1)) \\ &= \frac{\mu n \sum_{u \in W_n^k(\delta)} \frac{D_u k}{\mu n} (1 - e^{-D_u k / (\mu n)})}{k} (1 + o_{\mathbb{P}}(1)) \\ &= \frac{\mu^{2-\tau} n^{3-\tau}}{k^{3-\tau}} \int_{\delta}^{1/\delta} t(1 - e^{-t}) dM^{(n)}(t) (1 + o_{\mathbb{P}}(1)). \end{aligned} \quad (2.3.38)$$

Fix $\eta > 0$. Since $t(1 - e^{-t})$ is bounded and continuous on $[\delta, 1/\delta]$, we can find $m < \infty$, disjoint intervals $(B_i)_{i \in [m]}$ and constants $(b_i)_{i \in [m]}$ such that $\bigcup B_i = [\delta, 1/\delta]$

$$\left| t(1 - e^{-t}) - \sum_{i=1}^m b_i \mathbb{1}_{\{t \in B_i\}} \right| < \eta / \lambda([\delta, 1/\delta]), \quad (2.3.39)$$

for all $t \in [\delta, 1/\delta]$. Because $M^{(n)}(B_i) \xrightarrow{\mathbb{P}} \lambda(B_i)$ for all i ,

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(|M^{(n)}(B_i) - \lambda(B_i)| > \eta / (mb_i) \right) = 0. \quad (2.3.40)$$

Furthermore,

$$\begin{aligned} & \left| \int_{\delta}^{1/\delta} t(1 - e^{-t}) dM^{(n)}(t) - \int_{\delta}^{1/\delta} t(1 - e^{-t}) d\lambda(t) \right| \\ & \leq \left| \int_{\delta}^{1/\delta} t(1 - e^{-t}) - \sum_{i=1}^m b_i \mathbb{1}_{\{t \in B_i\}} dM^{(n)}(t) \right| \\ & \quad + \left| \int_{\delta}^{1/\delta} t(1 - e^{-t}) - \sum_{i=1}^m b_i \mathbb{1}_{\{t \in B_i\}} d\lambda(t) \right| \\ & \quad + \left| \int_{\delta}^{1/\delta} \sum_{i=1}^m b_i \mathbb{1}_{\{t \in B_i\}} dM^{(n)}(t) - \int_{\delta}^{1/\delta} \sum_{i=1}^m b_i \mathbb{1}_{\{t \in B_i\}} d\lambda(t) \right|. \end{aligned} \quad (2.3.41)$$

Using that $\int_{\delta}^{1/\delta} \mathbb{1}_{\{t \in B_i\}} dM^{(n)}(t) = M^{(n)}(B_i)$ yields

$$\begin{aligned} & \left| \int_{\delta}^{1/\delta} \sum_{i=1}^m b_i \mathbb{1}_{\{t \in B_i\}} dM^{(n)}(t) - \int_{\delta}^{1/\delta} \sum_{i=1}^m b_i \mathbb{1}_{\{t \in B_i\}} d\lambda(t) \right| \\ & = \left| \sum_{i=1}^m b_i (M^{(n)}(B_i) - \lambda(B_i)) \right| = \left| \sum_{i=1}^m o_{\mathbb{P}}(\eta/m) \right| = o_{\mathbb{P}}(\eta). \end{aligned} \quad (2.3.42)$$

Thus, (2.3.41) results in

$$\left| \int_{\delta}^{1/\delta} t(1 - e^{-t}) dM^{(n)}(t) - \int_{\delta}^{1/\delta} t(1 - e^{-t}) d\lambda(t) \right| \leq \eta \frac{M^{(n)}([\delta, 1/\delta])}{\lambda([\delta, 1/\delta])} + \eta + o_{\mathbb{P}}(\eta).$$

Using that $M^{(n)}([\delta, 1/\delta]) = O_{\mathbb{P}}(\lambda([\delta, 1/\delta]))$ proves that

$$\int_{\delta}^{1/\delta} t(1 - e^{-t}) dM^{(n)}(t) \xrightarrow{\mathbb{P}} \int_{\delta}^{1/\delta} t(1 - e^{-t}) d\lambda(t) = c \int_{\delta}^{1/\delta} x^{1-\tau} (1 - e^{-x}) dx, \quad (2.3.43)$$

which together with (2.3.38) proves the lemma. \square

Conditional variance of $a(k)$.

We now show that the variance of $a_{\varepsilon_n}(k, W_n^k(\delta))$ is small when conditioning on the degree sequence, so that $a_{\varepsilon_n}(k, W_n^k(\delta))$ concentrates around its expected value computed in Lemma 2.2.

Lemma 2.4. *When $n^{(\tau-2)/(\tau-1)} \ll k \ll n^{1/(\tau-1)}$,*

$$\frac{\text{Var}_n \left(a_{\varepsilon_n}(k, W_n^k(\delta)) \right)}{\mathbb{E}_n \left[a_{\varepsilon_n}(k, W_n^k(\delta)) \right]^2} \xrightarrow{\mathbb{P}} 0. \quad (2.3.44)$$

Proof. Again, it suffices to prove the lemma under the event J_n and A_n from (2.3.1). We write the variance of $a_{\varepsilon_n}(k, W_n^k(\delta))$ as

$$\begin{aligned} \text{Var}_n \left(a_{\varepsilon_n}(k, W_n^k(\delta)) \right) &= \frac{1}{k^2 |M_{\varepsilon_n}(k)|^2} \sum_{i,j \in M_{\varepsilon_n}(k)} \sum_{u,v \in W_n^k(\delta)} D_u^{(\text{er})} D_v^{(\text{er})} \\ &\quad \times (\mathbb{P}_n(X_{iu} = X_{jv} = 1) - \mathbb{P}_n(X_{iu} = 1) \mathbb{P}_n(X_{jv} = 1)) \\ &= \frac{(1 + o_{\mathbb{P}}(1))}{k^2 |M_{\varepsilon_n}(k)|^2} \sum_{i,j \in M_{\varepsilon_n}(k)} \sum_{u,v \in W_n^k(\delta)} D_u D_v \\ &\quad \times (\mathbb{P}_n(X_{iu} = X_{jv} = 1) - \mathbb{P}_n(X_{iu} = 1) \mathbb{P}_n(X_{jv} = 1)). \end{aligned} \quad (2.3.45)$$

Equation (2.3.45) splits into various cases, depending on the size of $\{i, j, u, v\}$. We denote the contribution of $|\{i, j, u, v\}| = r$ to the variance by $V^{(r)}(k)$. We first consider $V^{(4)}(k)$. We can write

$$\mathbb{P}_n(X_{iu} = X_{jv} = 0) = \mathbb{P}_n(X_{iu} = 0) \mathbb{P}_n(X_{jv} = 0 \mid X_{iu} = 0). \quad (2.3.46)$$

For the second term, we first pair all half-edges adjacent to vertex i , conditionally on not pairing to vertex u . Then the second term can be interpreted as the probability that vertex j does not pair to vertex v in a new configuration model with $\hat{L}_n = L_n - D_i = L_n(1 + o_{\mathbb{P}}(n^{-(\tau-2)/(\tau-1)}))$ half-edges, where the new degree of vertex j is reduced by the amount of half-edges from vertex i that paired to j . Similarly, the new degree of vertex v is reduced by the amount of half-edges from vertex i that paired to v . Since the expected number of half-edges from i that pair to vertex j is $O(D_i D_j / L_n) = D_j o_{\mathbb{P}}(n^{-(\tau-1)/(\tau-1)})$ [74], the new degree of vertex j is $\hat{D}_j = D_j(1 + o_{\mathbb{P}}(n^{-(\tau-2)/(\tau-1)}))$, and a similar statement holds for vertex v . Thus, by (2.3.30)

$$\begin{aligned} \mathbb{P}_n(X_{iu} = X_{jv} = 0) &= e^{-D_i D_u / L_n} e^{-\hat{D}_j \hat{D}_v / \hat{L}_n} + o_{\mathbb{P}}(n^{-(\tau-2)/(\tau-1)}) \\ &= e^{-D_i D_u / L_n} e^{-D_j D_v / L_n} (1 + o_{\mathbb{P}}(n^{-(\tau-2)/(\tau-1)})), \end{aligned} \quad (2.3.47)$$

using that $D_i D_u, D_j D_v = O(n)$ and $D_i, D_u, D_j, D_v \ll n^{1/(\tau-1)}$. This results in

$$\begin{aligned} \mathbb{P}_n(X_{iu} = X_{jv} = 1) &= 1 - \mathbb{P}_n(X_{iu} = 0) - \mathbb{P}_n(X_{jv} = 0) + \mathbb{P}_n(X_{iu} = X_{jv} = 0) \\ &= 1 + (-e^{-\frac{D_u k}{L_n}} - e^{-\frac{D_v k}{L_n}} + e^{-\frac{D_u k}{L_n} - \frac{D_v k}{L_n}}) (1 + o_{\mathbb{P}}(n^{-\frac{\tau-2}{\tau-1}})) \end{aligned}$$

$$= (1 - e^{-D_u k/L_n})(1 - e^{-D_v k/L_n})(1 + o_{\mathbb{P}}(1)), \quad (2.3.48)$$

where the last equality holds because $D_u k = \Theta(n)$ and $D_v k = \Theta(n)$ for $u, v \in W_n^k(\delta)$. Therefore

$$\begin{aligned} V^{(4)}(k) &= \frac{1}{|M_{\varepsilon_n}(k)|^2 k^2} \sum_{i,j \in M_{\varepsilon_n}(k)} \sum_{u,v \in W_n^k(\delta)} D_u D_v (1 - e^{-D_u k/L_n})(1 - e^{-D_v k/L_n})(1 + o_{\mathbb{P}}(1)) \\ &\quad - D_u D_v (1 - e^{-D_u k/L_n})(1 - e^{-D_v k/L_n})(1 + o_{\mathbb{P}}(1)) \\ &= \sum_{u,v \in W_n^k(\delta)} o_{\mathbb{P}} \left(k^{-2} D_u D_v (1 - e^{-D_u k/L_n})(1 - e^{-D_v k/L_n}) \right) \\ &= o_{\mathbb{P}} \left(\mathbb{E}_n \left[a_{\varepsilon_n}(k, W_n^k(\delta)) \right]^2 \right), \end{aligned}$$

where the last equality follows from Lemma 2.2. Since there are no overlapping edges when $\{i, j, u, v\} = 3$, $V^{(3)}(k)$ can be bounded similarly.

We then consider the contribution from $V^{(2)}$, which is the contribution where the two edges are the same. By Lemma 2.3, we have to show that this contribution is small compared to $n^{6-2\tau} k^{2\tau-6}$. We bound the summand in (2.3.45) as

$$D_u^2 \left(\mathbb{P}_n(X_{iu} = 1) - \mathbb{P}_n(X_{iu} = 1)^2 \right) \leq D_u^2. \quad (2.3.49)$$

Thus, using that on A_n , $|M_{\varepsilon_n}(k)| \geq 1$, $V^{(2)}$, can be bounded as

$$\begin{aligned} V^{(2)} &\leq \frac{1}{k^2 |M_{\varepsilon_n}(k)|^2} \sum_{i \in M_{\varepsilon_n}(k)} \sum_{u \in W_n^k(\delta)} D_u^2 = \frac{1}{k^2 |M_{\varepsilon_n}(k)|} \sum_{u \in W_n^k(\delta)} D_u^2 \\ &= O(n^2 k^{-4}) |W_n^k(\delta)|. \end{aligned} \quad (2.3.50)$$

Since the degrees are i.i.d. samples from (2.2.3), $|W_n^k(\delta)|$ is distributed as a binomial with parameters $(n, C(n/k)^{1-\tau})$ for some constant C . Therefore, $|W_n^k(\delta)| = O_{\mathbb{P}}(n(n/k)^{1-\tau})$. This results in

$$V^{(2)} = O_{\mathbb{P}}(n^{4-\tau} k^{\tau-5}), \quad (2.3.51)$$

which is smaller than $n^{6-2\tau} k^{2\tau-6}$ when $k \gg n^{(\tau-2)/(\tau-1)}$, as required. \square

Proof of Proposition 2.2. Lemma 2.4 together with the Chebyshev inequality show that

$$\frac{a_{\varepsilon_n}(k, W_n^k(\delta))}{\mathbb{E}_n [a_{\varepsilon_n}(k, W_n^k(\delta))]} \xrightarrow{\mathbb{P}} 1. \quad (2.3.52)$$

Combining this with Lemmas 2.2 and 2.3 yields

$$\frac{a_{\varepsilon_n}(k, W_n^k(\delta))}{n^{3-\tau} k^{\tau-3}} \xrightarrow{\mathbb{P}} c \mu^{2-\tau} \int_{\delta}^{1/\delta} x^{1-\tau} (1 - e^{-x}) dx. \quad (2.3.53)$$

\square

Contributions outside $W_n^k(\delta)$.

In this section, we prove Proposition 2.1 and show that the contribution to $a_{\varepsilon_n}(k)$ outside of the major contributing regimes as described in (2.3.22) is negligible.

Proof of Proposition 2.1. We use that $\mathbb{P}_n(X_{ij} = 1) \leq \min(1, \frac{D_i D_j}{L_n})$. This yields

$$\begin{aligned} \mathbb{E} \left[a_{\varepsilon_n}(k, \bar{W}_n^k(\delta)) \right] &= \mathbb{E} \left[\mathbb{E}_n \left[a_{\varepsilon_n}(k, \bar{W}_n^k(\delta)) \right] \right] \leq \frac{n}{k} \mathbb{E} \left[D \min \left(1, \frac{kD}{L_n} \right) \mathbb{1}_{\{D \in \bar{W}_n^k(\delta)\}} \right] \\ &= \frac{n}{k} \int_0^{\delta \mu n / k} x^{1-\tau} \min \left(1, \frac{kx}{\mu n} \right) dx + \frac{n}{k} \int_{\mu n / (\delta k)}^{\infty} x^{1-\tau} \min \left(1, \frac{kx}{\mu n} \right) dx. \end{aligned} \quad (2.3.54)$$

For ease of notation, we assume that $\mu = 1$ in the rest of this section. We have to show that the contribution to (2.3.54) from vertices u such that $D_u < \delta n/k$ or $D_u > n/(\delta k)$ is small. First, we study the contribution to (2.3.54) for $D_u < \delta n/k$. We can bound this contribution by taking the second term of the minimum, which bounds the contribution as

$$\int_0^{\delta n/k} x^{2-\tau} dx = \frac{\delta^{3-\tau}}{\tau-3} (k/n)^{\tau-3}. \quad (2.3.55)$$

Then, we study the contribution for $D_u > n/(k\varepsilon)$. This contribution can be bounded very similarly by taking 1 for the minimum in (2.3.54) as

$$\frac{n}{k} \int_{n/(\delta k)}^{\infty} x^{1-\tau} dx = \frac{\delta^{\tau-2}}{\tau-2} (k/n)^{\tau-3}. \quad (2.3.56)$$

Taking $\kappa = \min(\tau-2, 3-\tau) > 0$ then proves the proposition. \square

2.3.4 Expected average nearest-neighbor degree

Similarly as in (2.3.23), we can write

$$\mathbb{E} [a(k, G_n)] = \mathbb{E} [a(k, W_n^k(\delta))] + \mathbb{E} [a(k, \bar{W}_n^k(\delta))]. \quad (2.3.57)$$

By Proposition 2.1, $\mathbb{E} [a(k, \bar{W}_n^k(\delta))] / (n/k)^{\tau-3} = O(\delta^\kappa)$. We now focus on the first term. The expected degree of a neighbor of a randomly chosen vertex of degree k can be written as

$$\begin{aligned} \mathbb{E} [a(k, W_n^k(\delta))] &= \mathbb{E}_n \left[D_{\mathcal{N}_{V_k}^{(er)}(U)} \mathbb{1}_{\{D_{\mathcal{N}_{V_k}(U)} \in [\delta, 1/\delta] \mu n/k\}} \right] \\ &= \mathbb{E}_n \left[D_{\mathcal{N}_{V_k}(U)} \mathbb{1}_{\{D_{\mathcal{N}_{V_k}(U)} \in [\delta, 1/\delta] \mu n/k\}} \right] (1 + o(1)) \end{aligned} \quad (2.3.58)$$

where $\mathcal{N}_{V_k}(U)$, denotes a uniformly chosen neighbor of a vertex of degree k . By (2.3.30), we can write the connection probability between a vertex of degree k and a neighbor of degree $d \in [\delta, 1/\delta] \mu n/k$ as $1 - e^{-kd/(\mu n)} (1 + o(1))$. Therefore

$$\begin{aligned} \mathbb{E} \left[a(k, W_n^k(\delta)) \right] &= (1 + o(1)) \int_{\delta \mu n/k}^{\mu n/(\delta k)} c x^{1-\tau} (1 - e^{-xk/(\mu n)}) dx \\ &= (1 + o(1)) (n/k)^{3-\tau} c \mu^{2-\tau} \int_{\delta}^{1/\delta} x^{1-\tau} (1 - e^{-x}) dx. \end{aligned} \quad (2.3.59)$$

Combining this with (2.3.57) and Proposition 2.1, and letting first $n \rightarrow \infty$ and then $\delta \rightarrow 0$ proves (2.2.21).

2.4 Proofs of Theorem 2.2 and 2.3

We now briefly show how the proof of Theorem 2.1 can be adapted for the rank-1 inhomogeneous random graph and the hyperbolic random graph to prove Theorems 2.2 and 2.3. We denote by \mathbb{P}_n the probability conditioned on the weights in the rank-1 inhomogeneous random graph or conditioned on the radial coordinates in the hyperbolic model. Again, for ease of notation, we drop the dependence of $a(k, G)$ on the graph G .

2.4.1 Inhomogeneous random graph

First, we show how to prove Theorem 2.2(i). In the rank-1 inhomogeneous random graph, the degree of vertex i with weight $h_i \gg 1$ satisfies $D_i = h_i(1 + o_{\mathbb{P}}(1))$ (see Section 4.C). Furthermore, the largest weight is of order $n^{1/(\tau-1)}$ with high probability. Thus, when $h \ll n^{(\tau-2)/(\tau-1)}$, w.h.p. $p(h, h') = hh'/(\mu n)$ for all vertices. When $u \in M_{\varepsilon_n}(k)$, $h_u = k(1 + o_{\mathbb{P}}(1))$, so that conditionally on the weight sequence

$$\begin{aligned} a_{\varepsilon_n}(k) &= \frac{1}{k|M_{\varepsilon_n}(k)|} \sum_{u \in M_{\varepsilon_n}(k)} \sum_{i \in [n]} D_i \mathbb{P}_n(X_{iu} = 1) \\ &= (1 + o_{\mathbb{P}}(1)) \frac{1}{k} \sum_{i \in [n]} h_i \frac{h_i k}{\mu n} = (1 + o_{\mathbb{P}}(1)) \sum_{i \in [n]} \frac{h_i^2}{\mu n}, \end{aligned} \quad (2.4.1)$$

which is equivalent to (2.3.20) because the weights are also sampled from (2.2.3). This proves Theorem 2.2(i).

Similarly to (2.3.22), we define for the rank-1 inhomogeneous random graph

$$W_n^{k, \text{HVM}}(\delta) = \{u : h_u \in [\delta \mu n/k, \mu n/(\delta k)]\}. \quad (2.4.2)$$

Then it is easy to show that Proposition 2.1 also holds for the rank-1 inhomogeneous random graph with (2.4.2) instead of $W_n^k(\delta)$. Because the weights are sampled from (2.2.3) and $\mathbb{P}_n(X_{ij} = 1) = \min(h_i h_j / (\mu n), 1)$, (2.3.54) also holds for the rank-1 inhomogeneous random graph, so that Proposition 2.1 indeed holds for the rank-1 inhomogeneous random graph.

We now sketch how to adjust the proof of Proposition 2.2 to prove an analogous version for the rank-1 inhomogeneous random graph, which states that

$$\frac{a_{\varepsilon_n}(k, W_n^{k, \text{HVM}}(\delta))}{(n/k)^{3-\tau}} \xrightarrow{\mathbb{P}} c \mu^{2-\tau} \int_{\delta}^{1/\delta} x^{1-\tau} \min(x, 1) dx. \quad (2.4.3)$$

Following the proofs of Lemmas 2.2-2.4, we see that these lemmas also hold for the rank-1 inhomogeneous random graph if we replace the connection probability of the erased configuration model of $1 - e^{-D_i D_j / L_n}$ by $\min(h_i h_j / \mu n, 1)$. For the rank-1 inhomogeneous random graph the contribution to (2.3.45) from 3 or 4 different vertices is 0, because the edge probabilities in the rank-1 inhomogeneous random graph conditioned on the weights are independent. From these lemmas, (2.4.3) follows. This then shows similarly to (2.3.27) that

$$\frac{a_{\varepsilon_n}(k)}{(n/k)^{3-\tau}} \xrightarrow{\mathbb{P}} c\mu^{2-\tau} \int_0^\infty x^{1-\tau} \min(x, 1) dx = \frac{c\mu^{2-\tau}}{(3-\tau)(\tau-2)}. \quad (2.4.4)$$

which proves Theorem 2.2(ii).

2.4.2 Hyperbolic random graph

We first provide a lemma that gives the connection probabilities conditioned on the radial coordinates in the hyperbolic random graph. Denote

$$g(x) = \begin{cases} \frac{2}{\pi} \sin^{-1}(x) & x < 1 \\ 1 & x \geq 1. \end{cases} \quad (2.4.5)$$

Lemma 2.5. *The probability that u and v are connected in a hyperbolic random graph conditionally on the radial coordinates can be written as*

$$\mathbb{P}_n(X_{uv} = 1) = g(vt(u)t(v)/n)(1 + o_{\mathbb{P}}(1)). \quad (2.4.6)$$

Proof. Suppose $vt(u)t(v)/n \geq 1$. Then,

$$1 \leq \frac{vt(u)t(v)}{n} = \frac{ve^R e^{-(r_u+r_v)/2}}{n} = \frac{n}{v} e^{-(r_u+r_v)/2}, \quad (2.4.7)$$

so that $r_u + r_v \leq 2 \log(n/v) = R$. Thus, by the definition of hyperbolic distance in (1.1.7)

$$\begin{aligned} \cosh(d(u, v)) &= \cosh(r_u) \cosh(r_v) - \sinh(r_u) \sinh(r_v) \cos(\theta_{uv}) \\ &\leq \cosh(r_u + r_v) \leq \cosh(R), \end{aligned} \quad (2.4.8)$$

so that the distance between u and v is less than R and u and v are connected.

Now suppose that $vt(u)t(v)/n < 1$, so that $r_u + r_v > R$. We calculate the maximal value of θ_{uv} such that u and v are connected, which we denote by θ_{uv}^* . When the angle between u and v equals θ_{uv}^* , the hyperbolic distance between u and v is precisely R . Thus, we obtain, using the definition of the hyperbolic sine and cosine

$$\frac{e^R - e^{-R}}{2} = \frac{e^{r_u} - e^{-r_u}}{2} \frac{e^{r_v} - e^{-r_v}}{2} - \frac{e^{r_u} + e^{-r_u}}{2} \frac{e^{r_v} + e^{-r_v}}{2} \cos(\theta_{uv}^*). \quad (2.4.9)$$

Because $t(u)$ is distributed as (2.2.18), the maximal type scales as $O_{\mathbb{P}}(n^{1/(\tau-1)})$. Therefore, $e^{r_u - r_v} = (t(v)/t(u))^2 = O_{\mathbb{P}}(n^{2/(\tau-1)})$. Similarly, $e^{r_v - r_u} = O_{\mathbb{P}}(n^{2/(\tau-1)})$ and

$e^{-r_u-r_v} \leq e^{r_u-r_v} = O_{\mathbb{P}}(n^{2/(\tau-1)})$. Furthermore, $e^{-R} = O(n^{-2})$ so that (2.4.9) becomes

$$\frac{1}{2}e^R + O(n^{-2}) = \frac{1}{4}e^{r_u+r_v}(1 - \cos(\theta_{uv}^*)) + O_{\mathbb{P}}\left(n^{2/(\tau-1)}\right). \quad (2.4.10)$$

We then use that by the definitions of $t(u), t(v)$ and R

$$e^{r_u+r_v} = e^R \left(e^{(r_u+r_v-R)/2} \right)^2 = e^R \left(\frac{n}{vt(u)t(v)} \right)^2. \quad (2.4.11)$$

This yields for (2.4.10) that

$$\begin{aligned} 1 - \cos(\theta_{uv}^*) &= 2 \left(\frac{vt(u)t(v)}{n} \right)^2 + O_{\mathbb{P}} \left(n^{\frac{2}{\tau-1}-2} \frac{v^2 t(u)^2 t(v)^2}{n^2} \right) \\ &= 2 \left(\frac{vt(u)t(v)}{n} \right)^2 + O_{\mathbb{P}} \left(n^{-2(\tau-2)/(\tau-1)} \right), \end{aligned} \quad (2.4.12)$$

for $vt(u)t(v)/n \leq 1$ so that

$$\begin{aligned} \theta_{uv}^* &= \cos^{-1}(1 - 2(vt(u)t(v)/n)^2)(1 + o_{\mathbb{P}}(1)) \\ &= 2 \sin^{-1}(vt(u)t(v)/n)(1 + o_{\mathbb{P}}(1)). \end{aligned} \quad (2.4.13)$$

Because u and v are connected if their relative angle is at most θ_{uv}^* and the angular coordinates of u and v are sampled uniformly, we obtain that

$$\mathbb{P}_n(X_{uv} = 1) = \frac{2}{\pi} \sin^{-1}(vt(u)t(v)/n)(1 + o_{\mathbb{P}}(1)). \quad (2.4.14)$$

□

Using this lemma, we now prove Theorem 2.3:

Proof of Theorem 2.3. We first focus on $k \ll n^{(\tau-2)/(\tau-1)}$. By [101, Section 4.3], for vertex u with radial coordinate r_u

$$\begin{aligned} \mathbb{E}_n[D_u] &= (n-1) \frac{2\alpha e^{-r_u/2}}{\pi(\alpha-1/2)} (1 + O(e^{-r_u})) \\ &= \frac{2v(\tau-1)}{\pi(\tau-2)} t(u) (1 + O((t(u)/n)^2)), \end{aligned} \quad (2.4.15)$$

where we have used that $\alpha = (\tau-1)/2$, $t(u) = e^{-(R-r_u)/2}$ and $R = \log(n/v)$. By [49, Theorem 2.7], for every τ and v , we can interpret the hyperbolic random graph as a variant of the geometric inhomogeneous random graph, defined in [49] where every vertex u has weight $w_u = t(u)$. By [49, Lemma 3.5(ii)], $D_u = \mathbb{E}_n[D_u] (1 + o_{\mathbb{P}}(1))$ in geometric inhomogeneous random graphs when $w_u = t(u) \gg 1$. Combining this with (2.4.15) we obtain that in the hyperbolic random graph

$$D_u = \frac{2v(\tau-1)}{\pi(\tau-2)} t(u) (1 + o_{\mathbb{P}}(1)) \quad (2.4.16)$$

when $1 \ll t(u) \ll n$. When $u \in M_{\varepsilon_n}(k)$, $D_u = k(1 + o(1))$. Using (2.4.16) then shows that

$$t(u) = \frac{\pi(\tau-2)}{2v(\tau-1)}k(1 + o_{\mathbb{P}}(1)) \quad (2.4.17)$$

when $k \gg 1$ and $u \in M_{\varepsilon_n}(k)$. Since the types are distributed as (2.2.18), the largest type is $O_{\mathbb{P}}(n^{1/(\tau-1)})$. Therefore, if $u \in M_{\varepsilon_n}(k)$, then $t(u)t(v)/n = o_{\mathbb{P}}(1)$ for all v . Applying that $\sin^{-1}(x) = x + O(x^2)$ to (2.4.6) then shows that for $u \in M_{\varepsilon_n}(k)$

$$\mathbb{P}_n(X_{uv} = 1) = \frac{2vt(u)t(v)}{\pi n}(1 + o_{\mathbb{P}}(1)) = \frac{\tau - 2}{\tau - 1} \frac{kt(v)}{n}(1 + o_{\mathbb{P}}(1)). \quad (2.4.18)$$

Thus, conditionally on the types,

$$\begin{aligned} a_{\varepsilon_n}(k) &= \frac{1}{k|M_{\varepsilon_n}(k)|} \sum_{u \in M_{\varepsilon_n}(k)} \sum_{v \in [n]} D_v \mathbb{P}_n(X_{uv} = 1) \\ &= (1 + o_{\mathbb{P}}(1)) \frac{2v(\tau - 1)}{\pi(\tau - 2)k} \sum_{v \in [n]} t(v) \frac{t(v)k(\tau - 2)}{(\tau - 1)n} = (1 + o_{\mathbb{P}}(1)) \sum_{v \in [n]} \frac{2vt(v)^2}{\pi n}. \end{aligned}$$

Combining this with the power-law distribution of the types (2.2.18) proves Theorem 2.3(i) (which is the same as Theorem 2.2(i) where μ is replaced by $\pi/(2v)$ and $c/(\tau - 1)$ by 1).

We now investigate the case $k \gg n^{(\tau-2)/(\tau-1)}$. Similarly to (2.3.22), we define for the hyperbolic random graph

$$W_n^{\text{HRG}}(\delta) = \{u : t(u) \in [\delta\zeta n/k, \zeta n/(\delta k)]\} \quad (2.4.19)$$

with $\zeta = 2(\tau - 1)/(\pi(\tau - 2))$. Using that $2 \sin^{-1}(x)/\pi \leq x$ combined with Lemma 2.5, we obtain

$$\mathbb{P}_n(X_{uv} = 1) \leq \min(2vt(u)t(v)/(\pi n), 1). \quad (2.4.20)$$

Combining this with the fact that the $t(u)$'s are sampled from a distribution similar to (2.2.3) shows that (2.3.54) also holds for the hyperbolic random graph, apart from a multiplicative constant. From there we can follow the same lines as the proof of Proposition 2.1, so that Proposition 2.1 also holds for the hyperbolic random graph.

We follow the lines of the proof of Lemma 2.2, replacing $1 - e^{-D_u D_v / L_n}$ by $g(vt(u)t(v)/n)$ and using (2.4.16) to show that

$$\begin{aligned} \mathbb{E}_n \left[a_{\varepsilon_n}(k, W_n^k(\delta)) \right] &= \frac{(1 + o_{\mathbb{P}}(1))}{k|M_{\varepsilon_n}(k)|} \sum_{v \in M_{\varepsilon_n}(k)} \sum_{u \in W_n^k(\delta)} D_u g(vt(u)t(v)/n) \\ &= \frac{2v(\tau - 1)}{k\pi(\tau - 2)} \sum_{u \in W_n^k(\delta)} t(u) g\left(\frac{t(u)k\pi(\tau - 2)}{2n(\tau - 1)}\right) (1 + o_{\mathbb{P}}(1)) \\ &= \frac{v\zeta}{k} \sum_{u \in W_n^k(\delta)} t(u) g\left(\frac{t(u)k}{\zeta n}\right) (1 + o_{\mathbb{P}}(1)). \end{aligned} \quad (2.4.21)$$

We analyze this expression following the lines of the proof of Lemma 2.3. We define

$$M^{(n)}[a, b] = \frac{1}{\zeta^{1-\tau} n^{2-\tau} k^{\tau-1}} \sum_{u \in [n]} \mathbb{1}_{\{D_u \in [a, b]\zeta n/k\}}, \quad (2.4.22)$$

similarly to (2.3.36). From there, we follow the lines of the proof of Lemma 2.3, again replacing the connection probability $1 - \exp(-D_i D_j / (\mu n))$ of the erased configuration model by $g(vt(u)t(v)/n)$ and replacing the constant c from (2.2.3) by its equivalent constant for the hyperbolic model of $\tau - 1$ (see (2.2.18)) and μ by ζ . Then (2.4.21) results in

$$\begin{aligned} \mathbb{E}_n \left[a_{\varepsilon_n}(k, W_n^{k, \text{HRG}}(\delta)) \right] &= \frac{vn\zeta^2}{k^2} \sum_{u \in W_n^k(\delta)} \frac{t(u)k}{n\zeta} g\left(\frac{t(u)k}{\zeta n}\right) (1 + o_{\mathbb{P}}(1)) \\ &= v \left(\frac{n\zeta}{k}\right)^{3-\tau} \int_{\delta}^{1/\delta} t g(t) dM^{(n)}(t) (1 + o_{\mathbb{P}}(1)). \end{aligned} \quad (2.4.23)$$

Similar steps that prove (2.3.43) then show that

$$\frac{\mathbb{E}_n \left[a_{\varepsilon_n}(k, W_n^{k, \text{HRG}}(\delta)) \right]}{(n/k)^{3-\tau}} \xrightarrow{\mathbb{P}} (\tau - 1) v \zeta^{3-\tau} \int_{\delta}^{1/\delta} x^{1-\tau} g(x) dx. \quad (2.4.24)$$

Furthermore, because conditionally on the radial coordinates, the probabilities that two distinct edges are present are independent, Lemma 2.4 also holds for the hyperbolic random graph. This proves an analogous proposition to Proposition 2.2 which states that

$$\frac{a_{\varepsilon_n}(k, W_n^{k, \text{HRG}}(\delta))}{(n/k)^{3-\tau}} \xrightarrow{\mathbb{P}} (\tau - 1) v \zeta^{3-\tau} \int_{\delta}^{1/\delta} x^{1-\tau} g(x) dx. \quad (2.4.25)$$

Similar steps that lead to (2.3.27) then show that

$$\frac{a_{\varepsilon_n}(k)}{(n/k)^{3-\tau}} \xrightarrow{\mathbb{P}} (\tau - 1) v \zeta^{3-\tau} \int_0^{\infty} x^{1-\tau} g(x) dx. \quad (2.4.26)$$

Finally,

$$\begin{aligned} \int_0^{\infty} x^{1-\tau} g(x) dx &= \frac{2}{\pi} \int_0^1 x^{1-\tau} \sin^{-1}(x) dx + \int_1^{\infty} x^{1-\tau} dx \\ &= \frac{2}{\pi} \left[\frac{x^{2-\tau} \sin^{-1}(x)}{2-\tau} \right]_0^1 + \frac{1}{\tau-2} \int_0^1 \frac{x^{2-\tau}}{\sqrt{1-x^2}} dx + \frac{1}{\tau-2} \\ &= \frac{1}{\tau-2} \int_0^{1/(2\pi)} \sin(t)^{2-\tau} dt, \end{aligned} \quad (2.4.27)$$

where the last equation uses the substitution $t = \sin(x)$. By [96, Eq. 3.621.5]

$$\frac{1}{\tau-2} \int_0^{1/(2\pi)} \sin(t)^{2-\tau} dt = \frac{\Gamma(\frac{3-\tau}{2})\Gamma(\frac{1}{2})}{2(\tau-2)\Gamma(\frac{4-\tau}{2})} = \frac{\sqrt{\pi}\Gamma(\frac{3-\tau}{2})}{2(\tau-2)\Gamma(\frac{4-\tau}{2})}, \quad (2.4.28)$$

where Γ denotes the Gamma-function, which finishes the proof of Theorem 2.3(ii). \square

3 Global clustering in inhomogeneous random graphs

Based on:

Local clustering in scale-free networks with hidden variables
R. van der Hofstad, A.J.E.M. Janssen, J.S.H. van Leeuwen, C. Stegehuis
Physical Review E 95 p. 022307 (2017)

In this chapter, we investigate the presence of triangles in a wide class of scale-free rank-1 inhomogeneous random graphs. We first show that the clustering spectrum $k \mapsto c(k)$ decreases in k . We then determine how the average clustering coefficient C scales with the network size n . For scale-free networks with exponent $2 < \tau < 3$ this gives $C \sim n^{2-\tau} \ln n$ for the universality class at hand. We characterize the extremely slow decay of C when $\tau \approx 2$ and show that for $\tau = 2.1$, say, clustering only starts to vanish for networks as large as $n = 10^9$.

3.1 Introduction

In rank-1 inhomogeneous random graphs, vertices are characterized by weights that influence the creation of edges between pairs of vertices. All topological properties, including correlations and clustering, then become functions of the distribution of the weights and the probability of connecting vertices [35, 42]. The independence between edges makes rank-1 inhomogeneous random graphs analytically tractable. However, Chapter 2 showed that these rank-1 inhomogeneous random graphs introduce disassortative degree-degree correlations when $\tau < 3$: high-degree vertices tend to be connected to low-degree vertices. This negative correlation can have a strong influence on topological network properties, including clustering, which measures the presence of triangles in the network [145, 174].

In [201] it was shown that rank-1 inhomogeneous random graphs with a non-restrictive cutoff scheme can generate nearly size-independent levels of clustering, and can thus generate networks with high levels of clustering, particularly for τ close to 2. In the configuration model, without banning large-degree vertices by installing a cutoff, the long tail of the power law makes it quite likely that pairs of high-degree vertices share more than one edge. But rank-1 inhomogeneous random graphs allow at most one edge between pairs of vertices, so that large-degree vertices must inevitably connect to small-degree vertices due to lack of available large-degree vertices. This phenomenon is related to the difference in scaling between the so-called structural cutoff and natural cutoff. The structural cutoff is defined as the largest

possible upper bound on the degrees required to avoid degree correlations, while the natural cutoff characterizes the maximal degree in a sample of n vertices. For scale-free networks with $\tau \in (2, 3]$ the structural cutoff scales as $n^{1/2}$ while the natural cutoff scales as $n^{1/(\tau-1)}$ (see Section 3.2), which gives rise to structural negative correlations and possibly other finite-size effects.

Clustering can be measured in various ways, as explained in Section 1.2.2. The local clustering coefficient of vertex i is given by $c_i = 2T_i/k_i(k_i - 1)$ with k_i the degree of vertex i and T_i the number of triangles that vertex i is part of. This can be interpreted as the probability that two randomly chosen neighbors of a vertex are neighbors themselves. The average clustering coefficient C is then defined as the average (over vertices of degree ≥ 2) of the local clustering coefficient of single vertices.

In the absence of high-degree vertices (for $\tau > 3$), the average clustering coefficient of rank-1 inhomogeneous random graphs is given by [162]

$$C = \frac{1}{n} \sum_{i=1}^n c_i = \frac{\mathbb{E}[D(D-1)]^2}{n\mathbb{E}[D]^3} (1 + o(1)), \quad (3.1.1)$$

where D denotes the degree of a uniformly chosen vertex in the network. This shows that clustering vanishes very fast in the large network limit $n \rightarrow \infty$ in support of the tree-like approximations of complex networks. However, for scale-free distributions with $\tau < 3$, the natural cutoff that scales as $n^{1/(\tau-1)}$ together with (3.1.1) gives $C \sim n^{(7-3\tau)/(\tau-1)}$. The diverging C for $\tau < 7/3$ is caused by the fact that the approximation in (3.1.1) allows many edges between high-degree vertices, and can be judged as anomalous or nonphysical behavior if one wants C to be smaller than 1 and interpret it as a probability or proportion. If a structural cutoff of order $n^{1/2}$ is imposed, hence banning the largest-degree vertices, formula (3.1.1) predicts the correct (in the sense that it matches simulations) scaling $n^{2-\tau}$ [57, 201].

In a power-law setting, infinite variance is essential for describing scale-free network behavior, which makes the banning of large-degree vertices unnatural. In this chapter we investigate average clustering for a class of scale-free random graphs that allows for an interplay between structural correlations and large-degree vertices. The clustering coefficient in this ensemble turns out to depend on the network size, the structural cutoff that arises when conditioning on simplicity and the natural cutoff that accounts for large degrees.

Outline. We first describe the wide class of inhomogeneous random graphs that we study throughout this chapter in Section 3.2. We then show in Section 3.3 that the local clustering coefficient is a decaying function of the vertex weight across the entire class of rank-1 inhomogeneous random graphs and relate the decay of the local clustering coefficient to the average clustering coefficient C . In Section 3.4 we then compute C as a function of n and τ for one of the members of the class of rank-1 inhomogeneous random graphs and show that it serves as a bound for the clustering coefficient of all members of the class of inhomogeneous random graphs. After that, we show in Section 3.5 that for τ close to 2, the clustering coefficient decays extremely slowly in the network size n . Section 3.6 concludes the chapter. We present all mathematical derivations of the main results in the appendix.

3.2 Hidden variables and cutoffs

In this chapter, we focus on rank-1 inhomogeneous random graphs as defined in Section 1.1.4. Given n vertices, each vertex is equipped with a weight h drawn from a given probability distribution function $\rho(h)$. Then each pair of vertices is joined independently according to a given probability $p(h, h')$ with h and h' the weights associated to the two vertices. The probability $p(h, h')$ can be any function of the weights, as long as $p(h, h') \in [0, 1]$. Chung and Lu [61] introduced this model in the form

$$p(h, h') \sim \frac{hh'}{n\mathbb{E}[h]}, \quad (3.2.1)$$

so that the expected degree of a vertex equals its weight. For (3.2.1) to make sense the product hh' should never exceed $n\mathbb{E}[h]$. This can be guaranteed by the assumption that the weight h is smaller than the structural cutoff $h_s = \sqrt{n\mathbb{E}[h]}$. While this restricts $p(h, h')$ within the interval $[0, 1]$, the structural cutoff strongly violates the reality of scale-free networks. Regarding the weights as the desired degrees, the structural cutoff conflicts with the fact that the natural cutoff for the degree scales as $n^{1/(\tau-1)}$.

In [35, 42, 174] more general rank-1 inhomogeneous random graphs were introduced, constructed to preserve the conditional independence between edges, while making sure there is only one edge between every vertex pair and that the natural extreme values of power-law degrees are not neglected. Within that large spectrum of models, we focus on the subset of models for which the fraction of vertices of degree k , $P(k)$, satisfies

$$P(k) \sim \rho(k), \quad (3.2.2)$$

so that the degrees and the weights in the network are similar. The class of models considered in this chapter starts from the ansatz $p(h, h') \approx hh'/n\mathbb{E}[h]$, but like [35, 42, 50, 166, 174] adapts this setting to incorporate extreme values.

3.2.1 Class of random graphs

Within the wide class of rank-1 inhomogeneous random graphs we consider probabilities of the form

$$p(h, h') = r(u) = uf(u) \quad \text{with} \quad u = hh'h_s^{-2} \quad (3.2.3)$$

with functions $f: [0, \infty) \rightarrow (0, 1]$ that belong to the F-class spanned by the properties

F1 $f(0) = 1$, $f(u)$ decreases to 0 as $u \rightarrow \infty$.

F2 $r(u) = uf(u)$ increases to 1 as $u \rightarrow \infty$.

F3 f is continuous and there are $0 = u_0 < u_1 < \dots < u_K < \infty$ such that f is twice differentiable on each of the intervals $[u_{k-1}, u_k]$ and on $[u_K, \infty)$, where

$$f'(u_k) = \frac{1}{2}f'(u_k + 0) + \frac{1}{2}f'(u_k - 0) \quad (3.2.4)$$

for $k = 1, \dots, K$ and $f'(0) = f'(+0)$.

F4 $-uf'(u)/f(u)$ is increasing in $u \geq 0$.

The class of rank-1 inhomogeneous random graphs considered in this chapter is completely specified by all functions f that satisfy F1-F4. Here are important classical members of the F-class:

(i) (*maximally dense graph*) The Chung-Lu setting

$$r(u) = \min\{u, 1\}. \quad (3.2.5)$$

This is the default choice in [42] and leads within the F-class to the densest random graphs.

(ii) (*Poisson graph*) A simple exponential form gives

$$r(u) = 1 - e^{-u}. \quad (3.2.6)$$

Here we take u to define the intensities of Poisson processes of edges, and ignore multiple edges, so that (3.2.6) gives the probability that there is an edge between two vertices. Variants of this form are covered in e.g. [22, 23, 42, 166].

(iii) (*maximally random graph*) The next function was considered in [174, 201, 203]:

$$r(u) = \frac{u}{1+u}. \quad (3.2.7)$$

This connection probability ensures that the entropy of the ensemble is maximal [201]. This random graph is also known in the literature as the generalized random graph [50, 106].

The conditions F1-F4 will prove to be the minimally required conditions for the results that we present for the clustering coefficient. The F-class is constructed so that it remains amenable to analysis; the technique developed in [201] to characterize the average clustering coefficient despite the presence of correlations can be applied to our class. Notice that the technical condition F3 allows to consider piecewise smooth functions with jumps in their derivatives, such as $\min(1, 1/u)$ that comes with (3.2.5). It can be shown that F4 is slightly stronger than the condition of concavity of $r(u)$. It appears in Section 3.3 that F4 is necessary and sufficient for monotonicity in a general sense of the local clustering coefficient $c(h)$.

3.2.2 Cutoffs and correlation

The structural cutoff h_s marks the point as of which correlations imposed by the network structure arise. All pairs of vertices with weights smaller than this cutoff are connected with probability close to $u = hh'/h_s^2$ and do not show degree-degree correlations. Since the natural cutoff describing the maximal degree is of the order $h_c \sim n^{1/(\tau-1)}$, for $\tau \geq 3$, correlations are avoided. For $\tau < 3$, however, the structural cutoff is smaller than the natural cutoff, resulting in a network with a structure that can only be analyzed by considering non-trivial degree-degree correlations. The extent to which the network now shows correlation is determined by the gap between

the natural cutoff h_c and the structural cutoff h_s . A fully uncorrelated network arises when $h_c < h_s$, while correlation will be present when $h_c > h_s$. Let $\mathbb{E}[h]$ denote the average value of the random variable h with density $\rho(h) = Ch^{-\tau}$ on $[h_{\min}, n]$, so that

$$\mathbb{E}[h] = \frac{\int_{h_{\min}}^n h^{1-\tau} dh}{\int_{h_{\min}}^n h^{-\tau} dh} = \frac{\tau - 1}{\tau - 2} \frac{h_{\min}^{2-\tau} - n^{2-\tau}}{h_{\min}^{1-\tau} - n^{1-\tau}}. \quad (3.2.8)$$

With the default choices

$$h_s = \sqrt{n\mathbb{E}[h]}, \quad h_c = (n\mathbb{E}[h])^{1/(\tau-1)} \quad (3.2.9)$$

in mind, the regime in terms of cutoffs we are interested in is, just as in [201],

$$h_s \leq h_c \ll h_s^2, \quad (3.2.10)$$

where we regard these cutoffs as indexed by n and consider what happens as $n \rightarrow \infty$, with emphasis on the asymptotic regime $h_s \ll h_c$ for n large.

Choice of the natural cutoff. We now show that h_c as given in (3.2.9) is an accurate approximation of $\mathbb{E}[\max(\underline{h}_1, \dots, \underline{h}_n)]$, where the \underline{h}_i are i.i.d. with $\rho(h)$ as density. We have

$$\mathbb{E}[\max_i \underline{h}_i] = h_{\min} \Gamma\left(\frac{\tau-2}{\tau-1}\right) \frac{\Gamma(n+1)}{\Gamma\left(n + \frac{\tau-2}{\tau-1}\right)} \approx h_{\min} \Gamma\left(\frac{\tau-2}{\tau-1}\right) n^{\frac{1}{\tau-1}}. \quad (3.2.11)$$

The first identity in (3.2.11) is exact, and follows from

$$\begin{aligned} \mathbb{E}[\max_i \underline{h}_i] &= \int_{h_{\min}}^{\infty} h d[\mathbb{P}^n(\underline{h} \leq h)] = \int_{h_{\min}}^{\infty} h d[(1 - \mathbb{P}(\underline{h} > h))^n] \\ &= \int_{h_{\min}}^{\infty} h d\left[1 - \left(\frac{h}{h_{\min}}\right)^{1-\tau}\right]^n = nh_{\min} \int_0^1 t^{\frac{1}{\tau-1}} (1-t)^{n-1} dt, \end{aligned} \quad (3.2.12)$$

using the substitution $t = (h/h_{\min})^{1-\tau} \in (0, 1]$ and the expression of the Beta function in terms of the Γ -function. The approximate identity in (3.2.11) follows from $\Gamma(n+a)/\Gamma(n+b) \approx n^{a-b}$, which is quite accurate when $a, b \in [0, 1]$ and n large.

Furthermore, for $2 < \tau < 3$,

$$\left(\frac{\tau-1}{\tau-2}\right)^{\frac{1}{\tau-1}} \leq \Gamma\left(\frac{\tau-2}{\tau-1}\right) \leq \frac{4}{3} \left(\frac{\tau-1}{\tau-2}\right)^{\frac{1}{\tau-1}}. \quad (3.2.13)$$

This inequality follows from

$$u^u \leq \Gamma(1+u) = u\Gamma(u) \leq \frac{4}{3}u^u, \quad (3.2.14)$$

using $u = (\tau-2)/(\tau-1) \in (0, 1/2]$ that can be shown by considering the concave function $\ln(\Gamma(1+u)) - u \ln u$ which vanishes at $u = 0, 1$ and is positive at $u = 1/2$ (the upper bound in (3.2.14) follows from a numerical inspection of this function). Then from (3.2.13) we get, using $\mathbb{E}[h] = h_{\min}(\tau-1)/(\tau-2)(1+o(1))$,

$$h_{\min}^{\frac{\tau-2}{\tau-1}} (n\mathbb{E}[h])^{\frac{1}{\tau-1}} \leq h_{\min} \Gamma\left(\frac{\tau-2}{\tau-1}\right) n^{\frac{1}{\tau-1}} \leq \frac{4}{3} h_{\min}^{\frac{\tau-2}{\tau-1}} (n\mathbb{E}[h])^{\frac{1}{\tau-1}}, \quad (3.2.15)$$

showing that the order of magnitude of $\mathbb{E}[\max_i \underline{h}_i]$ is $(n\mathbb{E}[h])^{1/(\tau-1)}$. This motivates our choice of h_c in (3.2.9).

3.3 Universal clustering properties

We now characterize the large-network asymptotics of the local clustering coefficient $c(h)$ and average clustering coefficient C for the class of rank-1 inhomogeneous random graphs described in Section 3.2.1. For a class of uncorrelated random scale-free networks with a cutoff of $n^{1/2}$ [57] it was shown that C scales as $n^{2-\tau}$, a decreasing function of the network size for $\tau > 2$. In [201] the more general setup discussed in Section 3.2 was used, with the specific choice of $r(u) = u/(1+u)$. After involved calculations with Lerch's transcendent, [201] revealed the scaling relation

$$C \sim h_s^{-2(\tau-2)} \ln(h_c/h_s). \quad (3.3.1)$$

For the default choices in (3.2.9) this predicts $C \sim n^{2-\tau} \ln n$ (ignoring the constant).

We adopt the weights formalism developed in [35] that leads, among other things, to explicit expressions for the local clustering coefficient $c(h)$ of a vertex with weight h and for the average clustering coefficient C . The clustering coefficient of a vertex with weight h can be interpreted as the probability that two randomly chosen edges from h are neighbors. The clustering of a vertex of degree one or zero is defined as zero. Then, if vertex h has degree at least two,

$$c(h) = \int_{h_{\min}}^{h_c} \int_{h_{\min}}^{h_c} p(h'|h)p(h',h'')p(h''|h)dh'dh'', \quad (3.3.2)$$

with $p(h'|h)$ the conditional probability that a randomly chosen edge from an h vertex is connected to an h' vertex given by

$$p(h'|h) = \frac{\rho(h')p(h,h')}{\int_{h''} \rho(h'')p(h,h'')dh''}. \quad (3.3.3)$$

The degree of a vertex conditioned on its weight h is asymptotically distributed as a Poisson random variable with parameter h [35] and [106, Chapter 6]. Therefore, the probability that a vertex with weight h has degree at least two is given by

$$\mathbb{P}(k \geq 2 | h) = \sum_{k=2}^{\infty} \frac{h^k e^{-h}}{k!} = 1 - e^{-h} - he^{-h}. \quad (3.3.4)$$

Therefore, for $\rho(h) \sim h^{-\tau}$ [35, Eq. (29)]

$$c(h) = (1 - e^{-h} - he^{-h}) \frac{\int_{h_{\min}}^{h_c} \int_{h_{\min}}^{h_c} \rho(h')p(h,h')\rho(h'')p(h,h'')p(h',h'')dh'dh''}{\left[\int_{h_{\min}}^{h_c} \rho(h')p(h,h')dh'\right]^2}, \quad (3.3.5)$$

and hence

$$C = \int_{h_{\min}}^{h_c} \rho(h)c(h)dh. \quad (3.3.6)$$

The Poisson distribution is sharply peaked around $k = h$, which for large k yields

$$P(k) \sim \rho(k) \quad \text{and} \quad \bar{c}(k) \sim c(k), \quad (3.3.7)$$

where $\bar{c}(k)$ denotes the average clustering coefficient over all vertices of degree k , so that the weights become soft constraints on the degrees.

We make the change of variables

$$a = 1/h_s, \quad b = h_c/h_s \quad (3.3.8)$$

and assume henceforth, in line with (3.2.10), that

$$0 < ah_{\min} \leq ah_{\min}b \leq 1 \leq b < \infty, \quad 2 < \tau < 3. \quad (3.3.9)$$

This gives $c(h) = (1 - e^{-h} - he^{-h})c_{ab}(h)$ with

$$c_{ab}(h) = \frac{\int_{ah_{\min}}^b \int_{ah_{\min}}^b (xy)^{-\tau} r(ahx)r(ahy)r(xy) dx dy}{\left[\int_{ah_{\min}}^b x^{-\tau} r(ahx) dx \right]^2}. \quad (3.3.10)$$

Within the domain of integration $[ah_{\min}, b]$ in (3.3.10) the arguments ahx and ahy do not exceed a maximum value $O(ab)$ as long as $h < h_s^2/h_c$, which tends to zero under assumption (3.2.10). Therefore, since $r(u) \approx u$, $c_{ab}(h) \approx c_{ab}(0)$ for $h < h_s^2/h_c$. When choosing h_s as in (3.2.9), this means that $c_{ab}(h) \approx c_{ab}(0)$ for $h \leq n\mathbb{E}[h]/h_c$. In Proposition 3.1 below we prove that $h \mapsto c_{ab}(h)$ is a bounded monotonically decreasing function for the class of models at hand. Furthermore, the density $\rho(h) \sim h^{-\tau}$ with $\tau \geq 2$, decays sufficiently rapidly for the integral in (3.3.6) for C to have converged already before $c_{ab}(h)$ starts to drop significantly below its value $c_{ab}(0)$ at $h = 0$. Thus, C can be approximated with

$$C_{ab}(\tau) = c_{ab}(0) \int_{h_{\min}}^n \rho(h) (1 - (1+h)e^{-h}) dh := c_{ab}(0) A(\tau), \quad (3.3.11)$$

where we have conveniently extended the integration range to the τ -independent interval $[h_{\min}, n]$ at the expense of a negligible additional error.

We now state our results on the properties of the local clustering coefficient for the class of rank-1 inhomogeneous random graphs described in Section 3.2.1:

Proposition 3.1. *Assume that f satisfies F1 - F3. Then $c_{ab}(h)$ is decreasing in $h \geq 0$ for all a, b with $0 < a < b$ if and only if f satisfies F4.*

Proposition 3.1 shows that for large enough h local clustering decreases with the weight. For the default choices (3.2.5), (3.2.9), local clustering $c(h)$ is plotted in Figure 3.1, which shows both exact formulas and extensive simulations. The proof of Proposition 3.1 can be found in Appendix 3.B.

Proposition 3.2. *Assume that f is positive, and satisfies F3. Then $c_{ab}(0)$ is decreasing in $\tau > 0$ for all a, b with $0 < a \leq b < \infty$ if and only if f satisfies F2.*

We prove Proposition 3.2 in Appendix 3.A. Proposition 3.2 gives evidence for the fact that clustering increases as τ decreases, as confirmed in Figure 3.3. More precisely, Proposition 3.2 shows the monotonicity of $c_{ab}(0)$, which is one of the factors of $C_{ab}(\tau)$ in (3.3.11). The issue of monotonicity of $C_{ab}(\tau)$ is more delicate, since a and b are function of τ themselves. In Appendix 3.F we present several other monotonicity

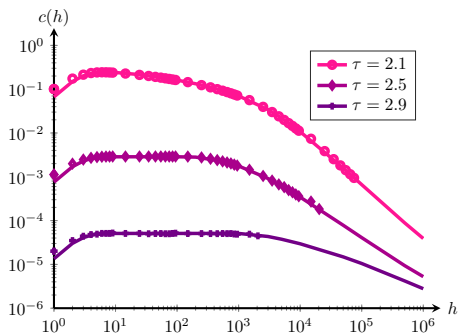


Figure 3.1: $c(h)$ for $\tau = 2.1, 2.5, 2.9$ and networks of size $n = 10^6$, using $h_{\min} = 1$. The markers indicate the average of 10^5 simulations, and the solid lines follow from (3.3.5).

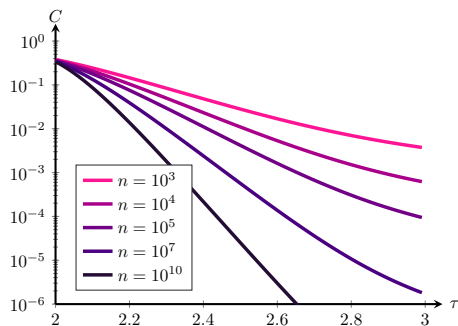


Figure 3.2: $C_{ab}^{\max}(\tau)$ as a function of τ , with r as in (3.2.5), using $h_{\min} = 1$.

properties of the remaining building blocks that together give $C_{ab}(\tau)$. It follows that for $\tau > 2$, $C_{ab}(\tau)$ is bounded from above by an envelope function of τ that is very close to $C_{ab}(\tau)$ and that is decreasing in τ . Figure 3.2 provides empirical evidence for the monotonicity of $C_{ab}(\tau)$ in τ . This monotonicity seems to conflict observations in [201], where the clustering coefficient of a rank-1 inhomogeneous random graph first increases in τ when τ is close to 2, and then starts decreasing. The difference is caused by the choice of the structural cutoff. Where we take $h_s = \sqrt{n\mathbb{E}[h]}$ with $\mathbb{E}[h]$ as in (3.2.8), in [201] $h_s = \sqrt{n(\tau-1)/(\tau-2)}$ was used. Thus, in [201], the structural cutoff includes the infinite system size limit of $\mathbb{E}[h]$, where we use the size-dependent version of $\mathbb{E}[h]$.

Figure 3.3 suggests that C falls off with n according to a function n^δ where δ depends on τ . In Proposition 3.3 below, we show that for the F-class of rank-1 inhomogeneous random graphs and the standard cutoff levels, C decays as $n^{\tau-2} \ln n$. On a log-scale, moreover, the clustering coefficient of different rank-1 inhomogeneous random graphs in the F-class only differs by a constant, which is confirmed in Figure 3.3 and substantiated in Proposition 3.4. Then, we focus in Section 3.5 on $\tau \approx 2$, for which Figure 3.3 suggests that the clustering remains nearly constant as a function of n , and characterize how large a network should be for C to start showing decay. This again will depend on τ .

3.4 Universal bounds

We next compute the clustering coefficient $C_{ab}^{\max}(\tau) = C_{ab}(\tau)$ for the maximally dense graph with $f(u) = f_{\max}(u) = \min(1, 1/u)$, $u \geq 0$ and a, b satisfying (3.3.9). In this case we have $c_{ab}(h) = c_{ab}(0)$ for $h \leq 1/(ab) = h_s^2/h_c = n\mathbb{E}[h]/h_c$. It is easy to see that f_{\max} is the maximal element in the F-class in the sense that $f(u) \leq f_{\max}(u)$ for all $u \geq 0$ and all $f \in F$. For a general $f \in F$ we shall also bound $C_{ab}^f(\tau)$ in terms of $C_{ab}^{\max}(\tau)$. This yields a scaling relation similar to (3.3.1), but then for the whole F-class.

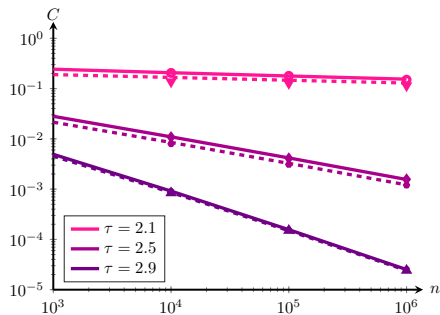


Figure 3.3: $C_{ab}(\tau)$ for $\tau = 2.1, 2.5, 2.9$, choices (3.2.5) (solid line) and (3.2.7) (dashed line) and networks of size $n = 10^k$ for $k = 4, 5, 6$, using $h_{\min} = 1$. The markers indicate the average of 10^5 simulations, and the solid lines follow from (3.4.1) and (3.4.5).

We start from an explicit representation for $C_{ab}^{\max}(\tau)$:

Proposition 3.3.

$$C_{ab}^{\max}(\tau) = \frac{A(\tau)(\tau - 2)^2}{((ah_{\min})^{2-\tau} - b^{2-\tau})^2} \times I_{ab}^{\max}(\tau), \quad (3.4.1)$$

with $A(\tau)$ given in (3.3.11) and

$$I_{ab}^{\max}(\tau) = \frac{\ln(b^2)}{(\tau - 2)(3 - \tau)} - \frac{1 - b^{2(2-\tau)}}{(\tau - 2)^2} + \frac{1 - 2(ah_{\min}b)^{3-\tau} + (ah_{\min})^{2(3-\tau)}}{(3 - \tau)^2}. \quad (3.4.2)$$

When τ is away from 2 and 3, b is large and a is small, we can ignore the $b^{2-\tau}$ in the front factor of (3.4.1) and the second term in (3.4.2). Furthermore, $ab = O(n^{(2-\tau)/(\tau-1)})$, so that we may also ignore this factor in the third term of (3.4.2). In this case we get the approximation

$$C_{ab}^{\max}(\tau) \approx A(\tau) \frac{\tau - 2}{3 - \tau} (ah_{\min})^{2(\tau-2)} \ln(b^2). \quad (3.4.3)$$

Using the default choices for a and b from (3.2.9) and (3.3.8) then shows that $C_{ab}^{\max} \sim n^{2-\tau} \ln(n)$ (ignoring the constant).

Maximally random graph. For the maximally random graph (3.2.7) the counterpart of (3.4.1) has been derived in [201]. Define Lerch's transcendent

$$\Phi(z, s, v) = \sum_{k=0}^{\infty} \frac{z^k}{(k+v)^s}. \quad (3.4.4)$$

In [201] it was shown that for the maximally random graph (3.2.7) with $h_{\min} = 1$

s	$\frac{\pi}{\sin(\pi s)}$	$\frac{1}{s(1-s)}$	$\frac{\pi^2 \cos(\pi s)}{(\sin(\pi s))^2}$	$\frac{1}{s^2} - \frac{1}{(1-s)^2}$
0.1	10.1664	11.1111	98.2972	98.7654
0.2	5.3448	6.2500	23.1111	23.4375
0.3	3.8832	4.7619	8.8635	9.0703
0.4	3.3033	4.1666	3.3719	3.4722
0.5	3.1416	4.0000	0.0000	0.0000

Table 3.1: Dominant terms in (3.4.5) and (3.4.1) for several values of $s = \tau - 2$.

$$\begin{aligned}
C_{ab}(\tau) = & \frac{A(\tau)(\tau-2)^2}{(a^{2-\tau} - b^{2-\tau})^2} \left\{ \frac{\pi \ln(b^2)}{\sin(\pi(\tau-2))} - \frac{\pi^2 \cos(\pi(\tau-2))}{(\sin(\pi(\tau-2)))^2} \right. \\
& + b^{-2(\tau-2)} \Phi(-b^{-2}, 2, \tau-2) + a^{2(3-\tau)} \Phi(-a^2, 2, 3-\tau) \\
& \left. - 2(ab)^{3-\tau} \Phi(-ab, 2, 3-\tau) \right\}. \tag{3.4.5}
\end{aligned}$$

(The expression is slightly simplified compared to [201, Eq. (5)].) Comparing (3.4.5) and (3.4.1) shows that the front factor is identical, and that the terms in between brackets differ. Table 3.1 compares the two dominant terms in (3.4.5) and (3.4.1) and shows that these terms are of comparable magnitude for $\tau - 2$ small.

Thus, on a log-scale the leading asymptotics of the maximally dense graph and the maximally random graph differ only by a constant, so that the decay exponent describing how the clustering decays with network size is the same. This can also be seen in Figure 3.3. In fact, for all functions f in the F-class we show below that the decay exponent is universal, and that the difference in constants can be bounded.

The entire F-class. Proposition 3.3 can be used to find upper and lower bounds for $C_{ab}^f(\tau)$ with general $f \in F$. Since $f(u) \leq f_{\max}(u)$, $u \geq 0$, it follows from (3.3.11) that

$$C_{ab}^f(\tau) \leq C_{ab}^{\max}(\tau). \tag{3.4.6}$$

The following proposition shows that $C_{ab}^f(\tau)$ can also be lower bounded by $C_{ab}^{\max}(\tau)$:

Proposition 3.4. For all $u_0 \geq 1$,

$$C_{ab}^f(\tau) \geq u_0 f(u_0) C_{a_0 b_0}^{\max}(\tau), \tag{3.4.7}$$

with $a_0 = a / \sqrt{u_0}$ and $b_0 = b / \sqrt{u_0}$.

In particular, the choice $u_0 = 1$ yields

$$C_{ab}^f(\tau) \geq f(1) C_{ab}^{\max}(\tau), \tag{3.4.8}$$

which together with (3.4.3) and (3.4.6) gives the large-network behavior of $C_{ab}^f(\tau)$, when $\tau \in (2, 3)$ and away from 2 and 3, and a is small and b is large (up to a multiplicative constant). In particular, this shows that $C_{ab}^f \sim n^{2-\tau} \ln(n)$ (again ignoring the constant). The proofs of Propositions 3.3 and 3.4 can be found in Appendix 3.C and 3.D, respectively.

3.5 Persistent clustering

In [201] it was observed that for values of the exponent $\tau \approx 2$, clustering remains nearly constant up to extremely large network sizes, which makes the convergence to the thermodynamic limit extremely slow (as also observed in [37, 121] and Figure 3.2). We now use the explicit results for the maximally dense graph to characterize this rate of convergence as a function of the network size n . For convenience, we assume in this section that $h_{\min} = 1$.

In view of the lower and upper bounds obtained in Section 3.4 for $C_{ab}^f(\tau)$ with general $f \in \mathcal{F}$ in terms of $C_{ab}^{\max}(\tau) = C_{ab}(\tau)$, it suffices to consider $C_{ab}^{\max}(\tau)$ for τ close to 2. In Appendix 3.E we show that when τ is close to 2 and $|\ln(ab)/\ln(b^2)|$ is small, $C_{ab}^{\max}(\tau)$ can be approximated by

$$C_{ab}^{\max}(\tau) \approx \frac{A(\tau)(1 - \frac{1}{3}(\tau - 2)\ln(b^2))}{2(1 - \frac{1}{2}(\tau - 2)\ln(ab) + \frac{1}{6}(\tau - 2)^2\ln^2 b)^2}. \quad (3.5.1)$$

The term $-\frac{1}{3}(\tau - 2)\ln(b^2)$ in the numerator and the term $\frac{1}{6}(\tau - 2)^2\ln^2 b$ in the denominator of the right-hand side of (3.5.1) are the main influencers on when $C_{ab}^{\max}(\tau)$ starts to decay. The decay is certainly absent as long as the numerator $1 - \frac{1}{3}(\tau - 2)\ln(b^2)$ is away from zero.

We then apply this reasoning to the canonical choices $h_s = \sqrt{n\mathbb{E}[h]}$ and $h_c = (n\mathbb{E}[h])^{1/(\tau-1)}$, for which

$$b = (n\mathbb{E}[h])^{\frac{3-\tau}{2(\tau-1)}}, \quad ab = (n\mathbb{E}[h])^{-\frac{\tau-2}{\tau-1}}, \quad (3.5.2)$$

ensuring $|\ln(ab)/\ln(b^2)| = (\tau - 2)/(\tau - 3)$ to be small indeed. Then, choosing a threshold $t \in (0, 3)$ and solving n from

$$(\tau - 2)\ln(b^2) = t, \quad (3.5.3)$$

we get

$$n\mathbb{E}[h] = \exp\left(\frac{\tau - 1}{(\tau - 2)(3 - \tau)}t\right). \quad (3.5.4)$$

In Table 3.2 we consider the case $t = 2$ and use that $\mathbb{E}[h]$ can accurately be bounded from above by $\ln n$ when $h_{\min} = 1$ and τ is close to 2, and we let $n_{\tau,2}$ be such that $n \ln n$ equals the right-hand side of (3.5.4). For $\tau = 2.1$, the value of n where the clustering starts to decay is much larger than the typical size of real-world network data sets. This supports the observation that clustering is persistent for τ close to 2. Figure 3.4 shows the decay of the clustering coefficient for $\tau \approx 2$, together with $n_{\tau,2}$, the value where the clustering coefficient is expected to decay. We see that for τ very close to 2, clustering indeed barely decays for n smaller than $n_{\tau,2}$. On the other hand, we see that for $\tau = 2.2$ and $\tau = 2.3$ the decay already starts before $n_{\tau,2}$, so there the approximation is less accurate.

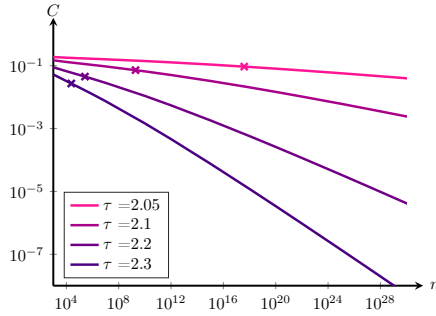


Figure 3.4: $C_{ab}^{\max}(\tau)$ as a function of n for τ close to 2, with $r(u)$ as in (3.2.5) and $h_{\min} = 1$. The marks indicate the value of $n_{\tau,2}$ as calculated in Table 3.2.

τ	$n_{\tau,2}$
2.3	$2.37 \cdot 10^4$
2.2	$2.62 \cdot 10^5$
2.1	$1.93 \cdot 10^9$
2.05	$3.92 \cdot 10^{17}$

Table 3.2: Solution $n_{\tau,t}$ to $(\tau - 2) \ln(b^2) = 2$.

3.6 Conclusion

For rank-1 inhomogeneous random graphs with scale-free degree distributions and connection probabilities in the F-class, we have shown that the local clustering coefficient $c(h)$ decays with the weight h and that the average clustering coefficient $C(\tau)$ roughly decreases with the tail exponent τ according to some function that depends on the structural and natural cutoffs. For the typical cutoff choices \sqrt{n} and $n^{1/(\tau-1)}$ this showed that C decays as $n^{2-\tau} \ln n$, confirming an earlier result in [201] and suggesting universal behavior for the entire F-class introduced in this chapter. By analyzing the special case of maximally dense graphs, a member of the F-class, we estimated the constant $C(\tau)/n^{2-\tau} \ln n$ and the extremely slow decay that occurs when $\tau \downarrow 2$.

3.A Proof of Proposition 3.2

We consider

$$c_{ab}(0) = D_{ab}(\tau) = \frac{\int_a^b \int_a^b f(xy)(xy)^{2-\tau} dx dy}{\left(\int_a^b x^{1-\tau} dx\right)^2}, \quad (3.A.1)$$

where we have written the lower integration limit ah_{\min} in (3.4.1) as a for notational convenience. We fix n , and study the dependence of $D_{ab}(\tau)$ on τ . We assume here that a and b are fixed, and do not depend on τ . Assume that f is positive and satisfies

F2 and F3. We have for $D'_{ab}(\tau) = \frac{d}{d\tau} D_{ab}(\tau)$,

$$D'_{ab}(\tau) = \frac{-\int_a^b \int_a^b f(xy) \ln(xy) (xy)^{2-\tau} dx dy}{\left(\int_a^b x^{1-\tau} dx\right)^2} + \frac{2 \int_a^b \int_a^b f(xy) (xy)^{2-\tau} dx dy \int_a^b x^{1-\tau} \ln x dx}{\left(\int_a^b x^{1-\tau} dx\right)^3}. \quad (3.A.2)$$

Observe that $D'_{ab}(\tau) \leq 0$ if and only if

$$\frac{\int_a^b \int_a^b f(xy) \ln(xy) (xy)^{2-\tau} dx dy}{\int_a^b \int_a^b f(xy) (xy)^{2-\tau} dx dy} \geq 2 \frac{\int_a^b x^{1-\tau} \ln x dx}{\int_a^b x^{1-\tau} dx}. \quad (3.A.3)$$

Symmetry of $f(xy)/(xy)^{\tau-2}$ and $\ln(xy) = \ln x + \ln y$ gives

$$\int_a^b \int_a^b f(xy) \ln(xy) (xy)^{2-\tau} dx dy = 2 \int_a^b \ln x \left(\int_a^b f(xy) (xy)^{2-\tau} dy \right) dx. \quad (3.A.4)$$

Letting

$$W(x) = \frac{\int_a^b f(xy) (xy)^{2-\tau} dy}{\int_a^b \int_a^b f(vy) (vy)^{2-\tau} dv dy}, \quad a \leq x \leq b, \quad (3.A.5)$$

$$V(x) = \frac{x^{1-\tau}}{\int_a^b v^{1-\tau} dv}, \quad a \leq x \leq b, \quad (3.A.6)$$

we thus need to show that

$$\int_a^b \ln x W(x) dx \geq \int_a^b \ln x V(x) dx. \quad (3.A.7)$$

Observe that for $K = \int_a^b v^{1-\tau} dv / \int_a^b \int_a^b f(vy) (vy)^{2-\tau} dv dy$,

$$W(x)/V(x) = Kx^{\tau-1} \int_a^b f(xy) (xy)^{2-\tau} dy = K \int_a^b (xyf(xy)) y^{1-\tau} dy, \quad (3.A.8)$$

which increases in $x > 0$ when f satisfies F2. Therefore, $W(x)/V(x)$ increases in $x > 0$ when f satisfies F2. Furthermore, $\ln x$ increases in $x > 0$, so the inequality in (3.A.7) follows from the following lemma:

Lemma 3.1. *Let $0 < a < b$ and assume that $p(x)$ and $q(x)$ are two positive, continuous probability distribution functions (pdf's) on $[a, b]$ such that $p(x)/q(x)$ is increasing in $x \in [a, b]$. Let $g(x)$ be an increasing function of $x \in [a, b]$. Then*

$$g_p = \int_a^b g(x)p(x)dx \geq \int_a^b g(x)q(x)dx = g_q. \quad (3.A.9)$$

Proof. For any $R \in \mathbb{R}$,

$$g_p - g_q = \int_a^b (g(x) - g_q)(p(x) - Rq(x))dx, \quad (3.A.10)$$

since p and q are pdf's. Let x_q be a point in $[a, b]$ such that $g(x) \leq g_q$ when $x \leq x_q$ and $g(x) \geq g_q$ when $x \geq x_q$. Choose $R = p(x_q)/q(x_q)$, so that by monotonicity of g and p/q ,

$$\begin{aligned} a \leq x \leq x_q &\Rightarrow (g(x) - g(x_q)) \leq 0 \wedge (p(x) - Rq(x) \leq 0), \\ x_q \leq x \leq b &\Rightarrow (g(x) - g(x_q)) \geq 0 \wedge (p(x) - Rq(x) \geq 0). \end{aligned}$$

Hence, the integrand in (3.A.10) is everywhere nonnegative, so that $g_p - g_q \geq 0$ as required. \square

Remark 3.1. The following observation will prove useful later: (i) The inequality in (3.A.9) is strict when both $g(x)$ and $p(x)/q(x)$ are strictly increasing. (ii) When $g(x)$ is (strictly) decreasing and $p(x)/q(x)$ is (strictly) increasing, there is \leq ($<$) rather than \geq ($>$) in (3.A.9).

Now that we have shown F2 to be a sufficient condition for $D_{ab}(\tau)$ to be increasing, we next show that F2 is also a necessary condition. Suppose we have two points u_1 and u_2 with $0 < u_1 < u_2$ such that $u_1 f(u_1) > u_2 f(u_2)$. Since f is continuous and piecewise continuous differentiable, there is a $u_0 \in (u_1, u_2)$ and $\varepsilon > 0$ such that $u f(u)$ is strictly decreasing in $u \in [u_0 - \varepsilon, u_0 + \varepsilon]$. In (3.A.8), take $a = \sqrt{u_0 - \varepsilon}$ and $b = \sqrt{u_0 + \varepsilon}$ so that $xy \in [u_0 - \varepsilon, u_0 + \varepsilon]$ when $xy \in [a, b]$. Therefore, $W(x)/V(x)$ is strictly decreasing in $x \in [a, b]$. By the version of Lemma 3.1 with $g(x) = \ln x$ strictly increasing and $p(x)/q(x) = W(x)/V(x)$ strictly decreasing, we see that $g_p - g_q < 0$. Therefore, we have (3.A.7) with $<$ instead of \geq , and so $D'_{ab}(\tau)$ is positive for all τ with this particular choice of a and b . This completes the proof of Proposition 3.2(i). \square

3.B Proof of Proposition 3.1

We consider for a fixed a, b, τ and $h > 0$,

$$c_{ab}(h) = \frac{\int_a^b \int_a^b (xy)^{2-\tau} f(ahx) f(ahy) f(xy) dx dy}{\left(\int_a^b x^{1-\tau} f(ahx) dx \right)^2}. \quad (3.B.1)$$

Observe that $\frac{d}{dh} c_{ab}(h) \leq 0$ if and only if

$$\frac{\frac{d}{dh} \left[\int_a^b \int_a^b (xy)^{2-\tau} f(ahx) f(ahy) f(xy) dx dy \right]}{\int_a^b \int_a^b (xy)^{2-\tau} f(ahx) f(ahy) f(xy) dx dy} \leq 2 \frac{\frac{d}{dh} \left[\int_a^b f(ahx) x^{1-\tau} dx \right]}{\int_a^b f(ahx) x^{1-\tau} dx}. \quad (3.B.2)$$

Using

$$\frac{d}{dh} [f(ahx) f(ahy)] = ax f'(ahx) f(ahy) + ay f(ahx) f'(ahy) \quad (3.B.3)$$

and the symmetry of the function $f(xy)/(xy)^{\tau-2}$ gives

$$\begin{aligned} & \frac{d}{dh} \left[\int_a^b \int_a^b f(ahx)f(ahy)f(xy)(xy)^{2-\tau} dx dy \right] \\ &= 2 \int_a^b \int_a^b ax f'(ahx)f(ahy)f(xy)(xy)^{2-\tau} dx dy. \end{aligned} \quad (3.B.4)$$

Also,

$$\frac{d}{dh} \left[\int_a^b f(ahx)x^{1-\tau} dx \right] = \int_a^b ax f'(ahx)x^{1-\tau} dx. \quad (3.B.5)$$

So we write the left-hand side of (3.B.2) as

$$2 \int_a^b \frac{ax f'(ahx)}{f(ahx)} T(x) dx, \quad (3.B.6)$$

and the right-hand side of (3.B.2) as

$$2 \int_a^b \frac{ax f'(ahx)}{f(ahx)} U(x) dx, \quad (3.B.7)$$

where the pdf's $T(x)$ and $U(x)$ on $[a, b]$ are defined as

$$T(x) = \frac{f(ahx) \int_a^b f(ahy)f(xy)(xy)^{2-\tau} dy}{\int_a^b \int_a^b f(ahv)f(ahy)f(vy)(vy)^{2-\tau} dv dy} \quad (3.B.8)$$

and

$$U(x) = \frac{f(ahx)x^{1-\tau}}{\int_a^b f(ahv)v^{1-\tau} dv}. \quad (3.B.9)$$

The inequality in (3.B.2) thus becomes

$$\int_a^b \frac{-ahx f'(ahx)}{f(ahx)} T(x) dx \geq \int_a^b \frac{-ahx f'(ahx)}{f(ahx)} U(x) dx, \quad (3.B.10)$$

where we have multiplied by $h > 0$. Assume that f satisfies F2. Then

$$x^{\tau-1} \int_a^b f(ahy)f(xy)(xy)^{2-\tau} dy = \int_a^b xy f(ahy)f(xy)y^{1-\tau} dy \quad (3.B.11)$$

is increasing in $x > 0$. Therefore, see (3.B.8) and (3.B.9), $T(x)/U(x)$ is increasing in $x > 0$. Hence, from Lemma 3.1 we get (3.B.10) when $g(x) = -ahx f'(ahx)/f(ahx)$ is increasing in $x > 0$, i.e. when f satisfies F4.

We have now shown that when f satisfies F1-F3, the condition F4 is sufficient for $c_{ab}(h)$ to be decreasing in $h > 0$. For the result in the converse direction we argue as follows. The function $uf(u)$ is continuous, piecewise smooth, increasing and not constant, and so there is a $u_0 > 0$, $\varepsilon > 0$ such that $uf(u)$ is strictly increasing in

$u \in [u_0 - \varepsilon, u_0 + \varepsilon]$. Let $z(v) = -vf'(v)/f(v)$, and assume there are $0 < v_1 < v_2$ such that $z(v_1) > z(v_2)$. We may assume that z is continuous at $v = v_1, v_2$. Indeed, when z is discontinuous at v_1 say, $z(v_1) = \frac{1}{2}(z(v_1 + 0) + z(v_2 - 0))$ and so at least one of $z(v_1 - 0) = \lim_{v \uparrow v_1} z(v)$ and $z(v_1 + 0) = \lim_{v \downarrow v_1} z(v)$ is larger than $z(v_2)$. Since z has only finitely many discontinuities, it suffices to decrease or increase v_1 somewhat, to a point of continuity, while maintaining $z(v_1) > z(v_2)$. We have to consider two cases.

A. Assume that $z(v)$ is continuous on $[v_1, v_2]$. We can then basically argue as in the proof of the only-if part of Proposition 3.2. Thus, there is a $v_0 > 0$, $\delta > 0$ such that $z(v)$ is strictly decreasing in $v \in [v_0 - \delta, v_0 + \delta]$. We choose a, b such that $xy \in [u_0 - \varepsilon, u_0 + \varepsilon]$ when $x, y \in [a, b]$. This is satisfied when $\sqrt{u_0 - \varepsilon} \leq a < b \leq \sqrt{u_0 + \varepsilon}$, and it guarantees that $T(x)/U(x)$ is strictly increasing in $x \in [a, b]$. Next, we choose h such that $ahx \in [v_0 - \delta, v_0 + \delta]$ when $x \in [a, b]$, so that $z(ahx)$ is strictly decreasing in $x \in [a, b]$. For this, we need to take h such that $a^2h \geq v_0 - \delta$ and $abh \leq v_0 + \delta$. This can be done indeed when $a/b \geq (v_0 - \delta)/(v_0 + \delta)$. Choosing a and b with $a < b$, $a, b \in [\sqrt{u_0 - \varepsilon}, \sqrt{u_0 + \varepsilon}]$ such that this latter condition is satisfied, we can apply the version of Lemma 3.1 with strictly decreasing $g(x) = z(ahx)$ and strictly increasing $p(x)/q(x) = T(x)/U(x)$. Thus we get in (3.B.10) strict inequality $<$ for these a, b and h , and this means that $c'_{ab}(h) < 0$. This proves Proposition 3.1 for this case.

B. Assume that $z(v)$ has discontinuities on $[v_1, v_2]$, say at $c_1 < c_2 < \dots < c_j$ with $v_1 < c_1$ and $v_2 > c_j$. In the case that there is an interval $[v_0 - \delta, v_0 + \delta]$ contained in one of $(v_1, c_1), (c_1, c_2), \dots, (c_j, v_2)$ where z is strictly decreasing, we are in the position of case A, and then we are done. Otherwise, we have by F3 that $z'(v) \geq 0$ for all $v \in [v_1, v_2]$, $v \neq c_1, \dots, c_j$. Then we must have $z(v_0 - 0) > z(v_0 + 0)$ for at least one $v_0 = c_1, \dots, c_j$, for else we would have $z(v_1) \leq z(c_1 - 0) \leq z(c_1 + 0) \leq \dots \leq z(c_j - 0) \leq z(c_j + 0) \leq z(v_2)$.

We want to find a, b such that

$$\int_a^b z(ahx)T(x)dx < \int_a^b z(ahx)U(x)dx \quad (3.B.12)$$

for the case that $z(v)$ has a downward jump at $v = v_0 > 0$ while being increasing to the left and to the right of v_0 . Set

$$\Delta = z(v_0 - 0) - z(v_0 + 0), \quad M = \frac{1}{2}(z(v_0 - 0) + z(v_0 + 0)), \quad (3.B.13)$$

and observe that $M \geq \frac{1}{2}\Delta > 0$ since $z(v) \geq 0$ for all v . We can find $\delta > 0$ such that

$$z(v_0 - 0) \geq z(v) \geq z(v_0 - 0) - \frac{1}{8}\Delta, \quad v_0 - \delta \leq v < v_0 \quad (3.B.14)$$

$$z(v_0 + 0) \leq z(v) \leq z(v_0 + 0) + \frac{1}{8}\Delta, \quad v_0 < v \leq v_0 + \delta. \quad (3.B.15)$$

Next, let

$$l(v) = f(v)v^{1-\tau}, \quad v > 0, \quad (3.B.16)$$

and observe that $l(v)$ is positive and continuous at $v = v_0$. Hence, we can choose $\delta > 0$ such that, in addition to (3.B.14) and (3.B.15),

$$\left| \frac{l(v)}{l(v_0)} - 1 \right| \leq \lambda, \quad v \in [v_0 - \delta, v_0 + \delta], \quad (3.B.17)$$

where λ is any number between 0 and $\frac{5}{16}\Delta/(2M + \frac{7}{16}\Delta)$. As in case A of the proof, we choose a, b and h such that

$$xy \in [u_0 - \varepsilon, u_0 + \varepsilon], \quad (3.B.18)$$

when $x, y \in [a, b]$ and

$$ahx \in [v_0 - \delta, v_0 + \delta], \quad (3.B.19)$$

when $x \in [a, b]$. Thus, we let $\sqrt{u_0 - \varepsilon} \leq a < b \leq \sqrt{u_0 + \varepsilon}$ such that $1 > a/b \geq (v_0 - \delta)/(v_0 + \delta)$. Below, we shall transform the two integrals by the substitution $v = ah_0x$ for a special choice of $h = h_0$ to an integral over an interval $[w_1, w_2]$ having v_0 as midpoint. This h_0 is given by

$$h_0 = \frac{2v_0}{a^2 + ab} \in \left[\frac{v_0 - \delta}{a^2}, \frac{v_0 + \delta}{ab} \right]. \quad (3.B.20)$$

Indeed, this h_0 satisfies (3.B.19) since

$$\begin{aligned} \frac{2v_0}{a^2 + ab} \leq \frac{v_0 + \delta}{ab} &\iff 2bv_0 \leq (b + a)(v_0 + \delta) \\ &\iff \left(1 - \frac{a}{b}\right)v_0 \leq \left(1 + \frac{a}{b}\right)\delta \\ &\iff \frac{a}{b} \geq \frac{v_0 - \delta}{v_0 + \delta}, \end{aligned} \quad (3.B.21)$$

and

$$\begin{aligned} \frac{2v_0}{a^2 + ab} \geq \frac{v_0 - \delta}{a^2} &\iff 2av_0 \leq (b + a)(v_0 - \delta) \\ &\iff \left(\frac{a}{b} - 1\right)v_0 \geq -\left(1 + \frac{a}{b}\right)\delta \\ &\iff \frac{a}{b} \geq \frac{v_0 - \delta}{v_0 + \delta}. \end{aligned} \quad (3.B.22)$$

In the integrals in the inequality in (3.B.12) with $h = h_0$, we substitute $ah_0x = v$, and the inequality to be proved becomes

$$z_t := \int_{w_1}^{w_2} z(v)t(v)dv < \int_{w_1}^{w_2} z(v)u(v)dv =: z_u. \quad (3.B.23)$$

Here

$$w_1 = a^2h_0, \quad w_2 = abh_0 \quad (3.B.24)$$

so that $v_0 = \frac{1}{2}(a^2 + ab)h_0$ is the midpoint of the integration interval $[w_1, w_2] \subset [v_0 - \delta, v_0 + \delta]$, and $t(v)$ and $u(v)$ are the pdf's

$$t(v) = \frac{1}{ah_0}T\left(\frac{v}{ah_0}\right), \quad u(v) = \frac{1}{abh_0}U\left(\frac{v}{abh_0}\right) \quad (3.B.25)$$

for which $t(v)/u(v)$ is strictly increasing in $v \in [w_1, w_2]$, since $T(x)/U(x)$ is strictly increasing for $x \in [a, b]$ by (3.B.11). We shall show that $z_u \in (M - \frac{3}{8}\Delta, M + \frac{3}{8}\Delta)$, and so, by (3.B.14) and (3.B.15),

$$z(v) - z_u > 0, \quad w_1 \leq v < v_0; \quad z(v) - z_u < 0, \quad v_0 < v \leq w_2. \quad (3.B.26)$$

With $R = t(v_0)/u(v_0)$, this implies that

$$z_t - z_u = \int_{w_1}^{w_2} (z(v) - z_u)(t(v) - Ru(v))dv < 0, \quad (3.B.27)$$

since the integrand is negative for all $v \neq v_0$.

To show that $z_u \in (M - \frac{3}{8}\Delta, M + \frac{3}{8}\Delta)$, we note that the pdf $u(v)$ is built from the function $l(v)$ in (3.B.16) via (3.B.9) and (3.B.25). In terms of this $l(v)$ we can write z_u as

$$z_u = \frac{\int_{w_1}^{w_2} z(v)l(v)dv}{\int_{w_1}^{w_2} l(v)dv}. \quad (3.B.28)$$

Now, by (3.B.17),

$$(w_2 - w_1)l(v_0)(1 - \lambda) \leq \int_{w_1}^{w_2} l(v)dv \leq (w_2 - w_1)(1 + \lambda)l(v_0). \quad (3.B.29)$$

Also, by (3.B.14), (3.B.15) and (3.B.17) and the fact that $v_0 = \frac{1}{2}(w_1 + w_2)$,

$$\begin{aligned} \int_{w_1}^{w_2} z(v)l(v)dv &= \int_{w_1}^{v_0} z(v)l(v)dv + \int_{v_0}^{w_2} z(v)l(v)dv \\ &\leq z(v_0 - 0) \int_{w_1}^{v_0} l(v)dv + z(v_0 + \frac{1}{8}\Delta) \int_{v_0}^{w_2} l(v)dv \\ &\leq \frac{1}{2}(w_2 - w_1)(1 + \lambda)l(v_0)[z(v_0 - 0) + z(v_0 + 0) + \frac{1}{8}\Delta] \\ &= (w_2 - w_1)(1 + \lambda)l(v_0)[M + \frac{1}{16}\Delta], \end{aligned} \quad (3.B.30)$$

and in a similar fashion

$$\int_{w_1}^{w_2} z(v)l(v)dv \geq (w_2 - w_1)(1 - \lambda)l(v_0)[M - \frac{1}{16}\Delta]. \quad (3.B.31)$$

From (3.B.29), (3.B.30) and (3.B.31) we then get

$$\frac{1 - \lambda}{1 + \lambda}(M - \frac{1}{16}\Delta) \leq z_u \leq \frac{1 + \lambda}{1 - \lambda}(M + \frac{1}{16}\Delta). \quad (3.B.32)$$

Now

$$\begin{aligned} \frac{1 + \lambda}{1 - \lambda}(M + \frac{1}{16}\Delta) < M + \frac{3}{8}\Delta &\iff \lambda < \frac{\frac{5}{16}\Delta}{2M + \frac{7}{16}\Delta}, \\ \frac{1 - \lambda}{1 + \lambda}(M - \frac{1}{16}\Delta) > M - \frac{3}{8}\Delta &\iff \lambda < \frac{\frac{5}{16}\Delta}{2M - \frac{7}{16}\Delta}. \end{aligned}$$

Then it follows from the choice of λ in (3.B.17) that $z_u \in (M - \frac{3}{8}\Delta, M + \frac{3}{8}\Delta)$ for such λ . \square

3.C Proof of Proposition 3.3

Taking the limit $h \downarrow 0$ in (3.3.10), with $r(u) = u \min(1, 1/u)$, we have

$$c_{ab}^{\max}(0) = \frac{\int_a^b \int_a^b (xy)^{2-\tau} \min(1, (xy)^{-1}) dx dy}{\left[\int_a^b x^{1-\tau} dx \right]^2}, \quad (3.C.1)$$

where we have written a instead of ah_{\min} for ease of notation. The denominator in (3.C.1) is evaluated as

$$\left(\int_a^b x^{1-\tau} dx \right)^2 = \frac{1}{(\tau-2)^2} (a^{2-\tau} - b^{2-\tau})^2. \quad (3.C.2)$$

For the numerator in (3.C.1) we compute

$$\begin{aligned} & \int_a^b \int_a^b \min(1, (xy)^{-1}) (xy)^{2-\tau} dx dy \\ &= \int_a^{1/b} \int_a^b (xy)^{2-\tau} dy dx + \int_{1/b}^b \left(\int_a^{1/x} (xy)^{2-\tau} dy + \int_{1/x}^b (xy)^{1-\tau} dy \right) dx \\ &= \int_a^{1/b} x^{2-\tau} dx \int_a^b y^{2-\tau} dy + \int_{1/b}^b x^{2-\tau} \int_a^{1/x} y^{2-\tau} dy dx \\ & \quad + \int_{1/b}^b x^{1-\tau} \int_{1/x}^b y^{1-\tau} dy dx \\ &= \frac{(b^{\tau-3} - a^{3-\tau})(b^{3-\tau} - a^{3-\tau})}{(3-\tau)^2} + \frac{1}{3-\tau} \int_{1/b}^b x^{2-\tau} (x^{\tau-3} - a^{3-\tau}) dx \\ & \quad + \frac{1}{2-\tau} \int_{1/b}^b x^{1-\tau} (b^{2-\tau} - x^{\tau-2}) dx \\ &= \frac{(b^{\tau-3} - a^{3-\tau})(b^{3-\tau} - a^{3-\tau})}{(3-\tau)^2} + \frac{1}{3-\tau} \left(\ln(b^2) - \frac{a^{3-\tau}(b^{3-\tau} - b^{\tau-3})}{3-\tau} \right) \\ & \quad + \frac{1}{2-\tau} \left(\frac{b^{2-\tau}(b^{2-\tau} - b^{\tau-2})}{2-\tau} - \ln(b^2) \right) \\ &= \frac{\ln(b^2)}{(3-\tau)(\tau-2)} - \frac{1-b^{2(2-\tau)}}{(2-\tau)^2} + \frac{1-2(ab)^{3-\tau} + a^{2(3-\tau)}}{(3-\tau)^2}. \end{aligned} \quad (3.C.3)$$

The last member of (3.C.3) equals $I_{ab}^{\max}(\tau)$ in (3.4.2), and the result follows from (3.3.11), (3.C.1), (3.C.2) and (3.C.3). \square

3.D Proof of Proposition 3.4

Take $u_0 \geq 1$ and note that

$$f(u) \geq u_0 f(u_0) \min(u_0^{-1}, u^{-1}), \quad u \geq 0, \quad (3.D.1)$$

since, for $f \in \mathbb{F}$,

$$f(u) \geq f(u_0), \quad 0 \leq u \leq u_0; \quad uf(u) \geq u_0 f(u_0), \quad u \geq u_0. \quad (3.D.2)$$

Now for any $c > 0$,

$$\begin{aligned} \int_a^b \int_a^b \min(c, (xy)^{-1})(xy)^{2-\tau} dx dy &= c^{\tau-2} \int_{a\sqrt{c}}^{b\sqrt{c}} \int_{a\sqrt{c}}^{b\sqrt{c}} \min(1, (xy)^{-1})(xy)^{2-\tau} dx dy \\ &= c^{\tau-2} I_{\max; a\sqrt{c}, b\sqrt{c}}(\tau). \end{aligned} \quad (3.D.3)$$

Also,

$$\frac{(\tau-2)^2}{(a^{2-\tau} - b^{2-\tau})^2} = c^{\tau-2} \frac{(\tau-2)^2}{((a\sqrt{c})^{2-\tau} - (b\sqrt{c})^{2-\tau})^2}. \quad (3.D.4)$$

The result then follows from combining (3.D.1), (3.D.3) and (3.D.4) into

$$\begin{aligned} C_{ab}^f &= A(\tau) c_{ab}^f(0) \\ &\geq A(\tau) u_0 f(u_0) \frac{\int_a^b \int_a^b (xy)^{2-\tau} \min(u_0^{-1}, (xy)^{-1}) dx dy}{[\int_a^b x^{1-\tau} dx]^2} \\ &= A(\tau) u_0 f(u_0) u_0^{\tau-2} I_{\max; a\sqrt{u_0}, b\sqrt{u_0}}(\tau) \frac{(\tau-2)^2}{(a^{2-\tau} - b^{2-\tau})^2} \\ &= u_0 f(u_0) C_{a_0 b_0}^{\max}(\tau), \end{aligned} \quad (3.D.5)$$

which proves the proposition. \square

3.E Derivation of Equation (3.5.1)

We derive (3.5.1) assuming (3.3.9) and that $|\ln(ab)/\ln(b^2)|$ is small. In the present case, where a and b are given by (3.2.9) and (3.3.8) with $h_{\min} = 1$, this indeed holds since $|\ln(ab)/\ln(b^2)| = (\tau-2)/(3-\tau)$. With $s = \tau-2$ we consider

$$C_{ab}^{\max}(\tau) = \frac{s^2}{(a^{-s} - b^{-s})^2} \left[\frac{\ln(b^2)}{s(1-s)} - \frac{1-b^{-2s}}{s^2} + \frac{1-2(ab)^{1-s} + a^{2(1-s)}}{(1-s)^2} \right], \quad (3.E.1)$$

where we have written a instead of ah_{\min} for notational convenience. We Taylor expand the term within brackets around $s = 0$, assuming $s \ln(b^2)$ of order unity and less,

$$\begin{aligned} &\frac{\ln(b^2)}{s(1-s)} - \frac{1-b^{-2s}}{s^2} + \frac{1-2(ab)^{1-s} + a^{2(1-s)}}{(1-s)^2} \\ &= \frac{1}{2} \ln^2 b^2 \left(1 - \frac{1}{3} s \ln(b^2) + \dots + O((\ln^2 b^2)^{-1}) \right). \end{aligned} \quad (3.E.2)$$

Also, Taylor expanding $(a^{-s} - b^{-s})/s$ around $s = 0$ yields

$$\frac{a^{-s} - b^{-s}}{s} = \ln(b/a) \left[1 - \frac{1}{2} s \ln(ab) + \frac{1}{6} s^2 (\ln^2 b + \ln b \ln a + \ln^2 a) - \dots \right]. \quad (3.E.3)$$

Note that

$$\ln(b/a) = \ln(b^2) - \ln(ab) = \ln(b^2) \left(1 - \frac{\ln(ab)}{\ln(b^2)} \right), \quad (3.E.4)$$

$$\ln^2 b + \ln b \ln a + \ln^2 a = \ln^2 b \left(1 - \frac{\ln(ab)}{\ln b} + \left(\frac{\ln(ab)}{\ln b} \right)^2 \right). \quad (3.E.5)$$

Thus we get

$$\begin{aligned} \frac{a^{-s} - b^{-s}}{s} &= \ln(b^2) \left(1 - \frac{\ln(ab)}{\ln(b^2)} \right) \left(1 - \frac{1}{2}s \ln(ab) + \frac{1}{6}s^2 \ln^2 b \left(1 + O\left(\frac{\ln(ab)}{\ln b} \right) \right) \right) \\ &\approx \ln(b^2) \left[1 - \frac{1}{2}s \ln(ab) + \frac{1}{6}s^2 \ln^2 b \right], \end{aligned} \quad (3.E.6)$$

where we have used the assumption that $|\ln(ab)/\ln b|$ is small. When we insert (3.E.2) and (3.E.6) into (3.E.1) and divide through $\ln^2(b^2)$, we arrive at (3.5.1).

3.F Monotonicity properties for $C_{ab}(\tau)$

In this appendix we show that $C_{ab}(\tau)$ is bounded from above by a closely related function that decreases in τ . Notice that Proposition 3.2 assumes a and b fixed. We have from (3.3.10) and (3.2.9) that

$$a = a(\tau) = (n\mathbb{E}[h])^{-1/2}, \quad b = b(\tau) = (n\mathbb{E}[h])^{\frac{3-\tau}{2(\tau-1)}} \quad (3.F.1)$$

where $\mathbb{E}[h]$ is given in (3.2.8). Below we use that $\mathbb{E}[h]$ decreases in $\tau \in (2, 3]$; this is clear intuitively and can be proved rigorously by using Lemma 3.1. We have

$$C_{ab}(\tau) = A(\tau)G(\tau, a(\tau), b(\tau)), \quad (3.F.2)$$

where

$$A(\tau) = \int_{h_{\min}}^n \rho(h) (1 - (1+h)e^{-h}) dh, \quad (3.F.3)$$

with density $\rho(h) = Ch^{-\tau}$ on $[h_{\min}, n]$ and

$$G(\tau, a, b) = \frac{\int_{ah_{\min}}^b \int_{ah_{\min}}^b (xy)^{2-\tau} f(xy) dx dy}{\left(\int_{ah_{\min}}^b x^{1-\tau} dx \right)^2}. \quad (3.F.4)$$

Proposition 3.2 says that

- (i) G decreases in τ (when a and b are fixed).

With the method of the proof of Proposition 3.2 in Appendix 3.A, we will show that

- (ii) A decreases in τ ,
- (iii) G increases in a and in b .

Showing that $G(\tau, a(\tau), b(\tau))$ decreases in τ is complicated by the facts that $a(\tau)$ increases in τ , see (iii), and that the dependence of $a(\tau)$ and $b(\tau)$ on τ is rather involved. Let m and M be the minimum and maximum, respectively, of $\mathbb{E}[h]$ when $\tau \in [2, 3]$ ($m = \mathbb{E}[h] |_{h=3} \approx 2h_{\min}$, $M = \mathbb{E}[h] |_{\tau \downarrow 2} \approx h_{\min} \ln(n/h_{\min})$) from (3.2.8) and the monotonicity of $\mathbb{E}[h]$). Letting

$$\bar{a} := (nm)^{-1/2}, \quad \bar{b}(\tau) = (nM)^{\frac{3-\tau}{2(\tau-1)}}, \quad (3.F.5)$$

we have $a(\tau) \leq \bar{a}$, $b(\tau) \leq \bar{b}(\tau)$, and so by (iii)

$$G(\tau, a(\tau), b(\tau)) \leq G(\tau, \bar{a}, \bar{b}(\tau)). \quad (3.F.6)$$

The right-hand side of (3.F.6) decreases in τ by (i) and (iii) and the fact that $\bar{b}(\tau)$ decreases in τ . Therefore, $C_{ab}(\tau)$ in (3.F.2) is bounded above by a closely related function that does decrease in τ .

We have shown that $G(\tau, a, b)$ decreases in $\tau \in [2, 3]$. We shall show now that

(ii) A decreases in τ and increases in l ,

(iii) G increases in a and in b .

Proof that A decreases in τ . Since $\frac{d}{d\tau} h^{-\tau} = -h^{-\tau} \ln h$, we have

$$\begin{aligned} \frac{\partial A}{\partial \tau} \leq 0 &\iff \int_{h_{\min}}^n -h^{-\tau}(1 - (1+h)e^{-h}) \ln h dh \int_{h_{\min}}^n h^{-\tau} dh \\ &\quad - \int_{h_{\min}}^n h^{-\tau}(1 - (1+h)e^{-h}) dh \int_{h_{\min}}^n -h h^{-\tau} \ln h dh \leq 0 \\ &\iff \frac{\int_{h_{\min}}^n h^{-\tau}(1 - (1+h)e^{-h}) \ln h dh}{\int_{h_{\min}}^n h^{-\tau}(1 - (1+h)e^{-h}) dh} \geq \frac{\int_{h_{\min}}^n h^{-\tau} \ln h dh}{\int_{h_{\min}}^n h^{-\tau} dh}. \end{aligned} \quad (3.F.7)$$

Consider on $[h_{\min}, n]$ the pdf's

$$p(h) = \frac{h^{-\tau}(1 - (1+h)e^{-h})}{\int_{h_{\min}}^n h_1^{-\tau}(1 - (1+h_1)e^{-h_1}) dh_1}, \quad (3.F.8)$$

$$q(h) = \frac{h^{-\tau}}{\int_{h_{\min}}^n h_1^{-\tau} dh_1} = \rho(h). \quad (3.F.9)$$

Clearly $p(h)/q(h) = C(1 - (1+h)e^{-h})$ with C independent of $h \in [h_{\min}, n]$. Hence, $p(h)/q(h)$ is increasing in $[h_{\min}, n]$. Also, $\ln(h) = g(h)$ is increasing in $[h_{\min}, n]$. Hence, by Lemma 3.1,

$$\int_{h_{\min}}^n g(h)p(h)dh \geq \int_{h_{\min}}^n g(h)q(h)dh, \quad (3.F.10)$$

and this is the last inequality in (3.F.7). \square

Proof that G increases in b . Again, for notational simplicity, we write a and b instead of ah_{\min} and bh_{\min} respectively. Let τ and a be fixed, and set

$$p(x, y) = (xy)^{2-\tau} f(xy), \quad P(b, y) = \int_a^b p(x, y) dx. \quad (3.F.11)$$

We have

$$\frac{d}{db} \left[\int_a^b \int_a^b (xy)^{2-\tau} f(xy) dx dy \right] = \frac{d}{db} \left[\int_a^b P(b, y) dy \right]$$

$$\begin{aligned}
&= P(b, b) + \int_a^b \frac{\partial P}{\partial b}(b, y) dy = \int_a^b p(x, b) dx + \int_a^b p(b, y) dy \\
&= 2 \int_a^b p(x, b) dx = 2 \int_a^b (xb)^{2-\tau} f(xb) dx
\end{aligned} \tag{3.F.12}$$

because of symmetry of $p(x, y)$. Also,

$$\frac{d}{db} \left(\int_a^b x^{1-\tau} dx \right)^2 = 2b^{1-\tau} \int_a^b x^{1-\tau} dx. \tag{3.F.13}$$

Therefore,

$$\begin{aligned}
\frac{\partial G}{\partial b} \geq 0 &\iff 2 \int_a^b (xb)^{2-\tau} f(xb) dx \left(\int_a^b x^{1-\tau} dx \right)^2 \\
&\quad - 2 \int_a^b \int_a^b (xy)^{2-\tau} f(xy) dx dy b^{1-\tau} b^{1-\tau} \int_a^b x^{1-\tau} dx \geq 0 \\
&\iff \frac{\int_a^b (xb)^{2-\tau} f(xb) dx}{\int_a^b \int_a^b (xy)^{2-\tau} f(xy) dx dy} \geq \frac{b^{1-\tau}}{\int_a^b x^{1-\tau} dx} \\
&\iff W(b) \geq V(b),
\end{aligned} \tag{3.F.14}$$

where $W(x)$ and $V(x)$ are the pdf's as defined in (3.A.5) and (3.A.6). Since $W(x)/V(x)$ increases in $x \in [a, b]$, we get

$$1 = \int_a^b W(x) dx = \int_a^b \frac{W(x)}{V(x)} V(x) dx \leq \frac{W(b)}{V(b)} \int_a^b V(x) dx = \frac{W(b)}{V(b)}, \tag{3.F.15}$$

and this gives the last inequality in (3.F.14). \square

Proof that G increases in a . This proof is very similar to the proof that G increases in b . Let τ and b be fixed. We now have

$$\frac{d}{da} \left[\int_a^b \int_a^b (xy)^{2-\tau} f(xy) dx dy \right] = -2 \int_a^b (xa)^{2-\tau} f(xa) dx, \tag{3.F.16}$$

and

$$\frac{d}{da} \left(\int_a^b x^{1-\tau} dx \right)^2 = -2a^{1-\tau} \int_a^b x^{1-\tau} dx. \tag{3.F.17}$$

Then, as in (3.F.14),

$$\begin{aligned}
\frac{\partial G}{\partial a} \geq 0 &\iff \frac{\int_a^b (xa)^{2-\tau} f(xa) dx}{\int_a^b \int_a^b (xy)^{2-\tau} f(xy) dx dy} \leq \frac{a^{1-\tau}}{\int_a^b x^{1-\tau} dx} \\
&\iff W(a) \leq V(a),
\end{aligned} \tag{3.F.18}$$

and the inequality follows again from increasingness of $W(x)/V(x)$. \square

4 Local clustering in inhomogeneous random graphs

Based on:
Clustering spectrum of scale-free networks
C. Stegehuis, R. van der Hofstad, J.S.H. van Leeuwen and A.J.E.M. Janssen
Physical Review E 96 p. 042309 (2017)

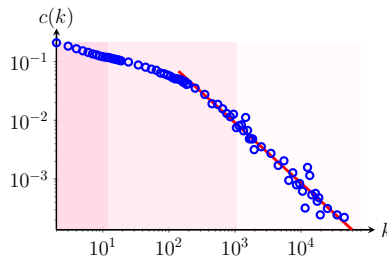
In this chapter, we again study clustering in the rank-1 inhomogeneous random graph, as in Chapter 3. Whereas Chapter 3 focused on the global clustering coefficient by taking the average over all vertices in the network, we now study local clustering to obtain a more detailed understanding of triangle formation in rank-1 inhomogeneous random graphs. That is, we study $c(k)$, the probability that two neighbors of a degree- k vertex are neighbors themselves. We investigate how the clustering spectrum $k \mapsto c(k)$ scales with k and show that $c(k)$ follows a *universal curve* that consists of three k -ranges where $c(k)$ first remains flat, then starts declining, and eventually settles on a power law $c(k) \sim k^{-\alpha}$ with α depending on the power law of the degree distribution. We test these results against ten contemporary real-world networks and explain analytically why the universal curve properties only reveal themselves in large networks.

4.1 Introduction

In *uncorrelated* networks the clustering spectrum $k \mapsto c(k)$ defined in (1.2.4) remains constant and independent of k . However, the majority of real-world networks have spectra that decay in k , as first observed in technological networks including the Internet [176, 187]. Figure 4.1 shows the same phenomenon for a social network: YouTube users as vertices, and edges indicating friendships between them [137].

Close inspection suggests the following properties, not only in Figure 4.1, but also in the nine further networks in Figure 4.2. The right end of the spectrum appears to be of the power-law form $k^{-\alpha}$; approximate values of α give rise to the dashed lines; (ii) The power law is only approximate and kicks in for rather large values of k . In fact, the slope of $c(k)$ decreases with k ; (iii) There exists a transition point: the minimal degree as of which the slope starts to decline faster and settles on its limiting (large k) value.

For scale-free networks a decaying $c(k)$ is taken as an indicator for the presence of modularity and hierarchy [187], collections of subgraphs with dense connections within themselves and sparser ones between them. The existence of clusters of dense

Figure 4.1: $c(k)$ for the YouTube social network

interaction signals hierarchical or nearly decomposable structures. When the function $c(k)$ falls off with k , low-degree vertices have relatively high clustering coefficients, creating small modules that are connected through triangles. In contrast, high-degree vertices have very low clustering coefficients, and therefore act as bridges between the different local modules. This also explains why $c(k)$ is not just a local property, and when viewed as a function of k , measures crucial mesoscopic network properties such as modularity, clusters and communities. The behavior of $c(k)$ also turns out to be a good predictor for the macroscopic behavior of the network. Randomizing real-world networks while preserving the shape of the $c(k)$ curve produces networks with very similar component sizes as well as similar hierarchical structures as the original network [202]. Furthermore, the shape of $c(k)$ strongly influences the behavior of networks under percolation [198]. This places the $c(k)$ -curve among the most relevant indicators for structural correlations in network infrastructures.

In this chapter, we obtain a precise characterization of clustering in the rank-1 inhomogeneous random graph. We start from an explicit form of the $c(k)$ curve [35, 74, 199], and obtain a detailed description of the $c(k)$ -curve in the large-network limit that provides rigorous underpinning of the empirical observations (i)-(iii). We find that the decay rate in the rank-1 inhomogeneous random graph is significantly different from the exponent $c(k) \sim k^{-1}$ that has been found in a hierarchical graph model [187] as well as in preferential attachment models with enhanced clustering [132, 204]. Furthermore, we show that before the power-law decay of $c(k)$ kicks in, $c(k)$ first has a constant regime for small k , and a logarithmic decay phase. This characterizes the entire clustering spectrum of the rank-1 inhomogeneous random graph.

This chapter is structured as follows. Section 4.2 introduces the random graph model and its local clustering coefficient. Section 4.3 presents the main results for the clustering spectrum. Section 4.4 explains the shape of the clustering spectrum in terms of an energy minimization argument, and Section 4.5 quantifies how fast the limiting clustering spectrum arises as a function of the network size. We conclude with a discussion in Section 4.6 and present all mathematical derivations of the main results in the appendix.

4.2 Rank-1 inhomogeneous random graph

As null model we employ the rank-1 inhomogeneous random graph, described in Section 1.1.4. We assume that the vertex weights $(h_i)_{i \in [n]}$ are drawn from the

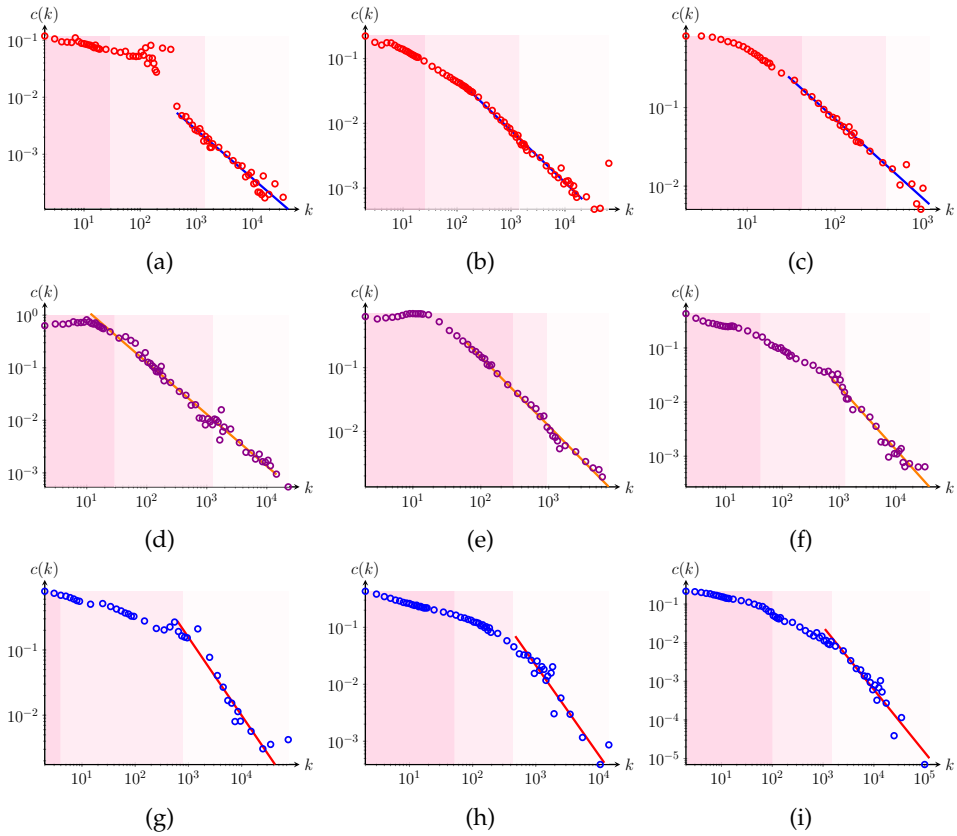


Figure 4.2: $c(k)$ for several information (red), technological (purple) and social (blue) real-world networks. (a) Hudong encyclopedia [165], (b) Baidu encyclopedia [165], (c) WordNet [150], (d) TREC-WT10g web graph [12], (e) Google web graph [137], (f) Internet on the Autonomous Systems level [137], (g) Catster/Dogster social networks [133], (h) Gowalla social network [137], (i) Wikipedia communication network [137]. The different shadings indicate the theoretical boundaries of the regimes as in Figure 4.3, with n and τ as in Table 4.1.

probability density

$$\rho(h) = Ch^{-\tau} \quad (4.2.1)$$

for $h \geq h_{\min}$ and some constant C . Each pair of vertices is joined independently with probability $p(h, h')$ with h and h' the weights associated to the two vertices.

The structural and natural cutoff, also described in Section 3.2.2, both play a crucial role in the description of the clustering spectrum. The structural cutoff h_s is defined as the largest possible upper bound on the degrees required to avoid degree correlations, while the natural cutoff h_c characterizes the maximal degree in a sample of n vertices. We again use the default choices for the structural and natural cutoffs

described in Section 1.1.4:

$$h_s = \sqrt{n\mathbb{E}[h]}, \quad h_c = (n\mathbb{E}[h])^{1/(\tau-1)}. \quad (4.2.2)$$

The fact that $h_s < h_c$ for scale-free networks with exponent $\tau \in (2, 3]$ gives rise to structural negative degree correlations. Throughout this chapter we use the connection probability (although many asymptotically equivalent choices are possible; see Chapter 3 and Section 4.3)

$$p(h, h') = \min\left(1, \frac{hh'}{h_s^2}\right) = \min\left(1, \frac{hh'}{n\mathbb{E}[h]}\right). \quad (4.2.3)$$

In this chapter, we work with $\bar{c}(h)$, the local clustering coefficient of a randomly chosen vertex with weight h . However, when studying local clustering in real-world data sets, we only observe $c(k)$, the local clustering coefficient of a vertex of degree k . In Appendix 4.C we show that the approximation $\bar{c}(h) \approx c(h)$ is highly accurate. We start from the explicit expression for $\bar{c}(h)$ [35], which measures the probability that two randomly chosen edges from h are neighbors, i.e.,

$$\bar{c}(h) = \int_{h'} \int_{h''} p(h'|h)p(h'', h')p(h''|h)dh''dh', \quad (4.2.4)$$

with $p(h'|h)$ the conditional probability that a randomly chosen edge from an h -vertex is connected to an h' -vertex and $p(h, h')$ as in (4.2.3). The goal is now to characterize the $\bar{c}(h)$ -curve (and hence the $c(k)$ -curve).

4.3 Universal clustering spectrum

The asymptotic evaluation of the double integral (4.2.4) in the large- n regime reveals three different ranges, defined in terms of the scaling relation between the weight h and the network size n . The three ranges together span the entire clustering spectrum as shown in Figure 4.3. We present the behavior of $\bar{c}(h)$ in these ranges below, the detailed calculations are deferred to Appendix 4.A.

The first range pertains to the smallest-degree vertices, i.e., vertices with a weight that does not exceed $n^{\beta(\tau)}$ with $\beta(\tau) = (\tau - 2)/(\tau - 1)$. In this case we show that

$$\bar{c}(h) \approx \frac{\tau - 2}{3 - \tau} h_s^{4-2\tau} \ln(h_c^2/h_s^2) \propto n^{2-\tau} \ln n \quad h \leq n^{\beta(\tau)}, \quad (4.3.1)$$

In particular, here the local clustering does not depend on the vertex weight and in fact corresponds with the large- n behavior of the global clustering coefficient of Chapter 3. The interval $[0, \beta(\tau)]$ diminishes when τ is close to 2, a possible explanation for why the flat range associated with Range I is hard to recognize in some of the real-world data sets.

Range II considers vertices with weights (degrees) above the threshold $n^{\beta(\tau)}$, but below the structural cutoff \sqrt{n} . These vertices start experiencing structural correlations, and close inspection of the integral (4.2.4) yields

$$\bar{c}(h) \approx h_s^{4-2\tau} \frac{\ln(h_s^2/h^2)}{(\tau - 2)(3 - \tau)} \propto n^{2-\tau} \ln\left(n/h^2\right), \quad n^{\beta(\tau)} \leq h \leq \sqrt{n}. \quad (4.3.2)$$

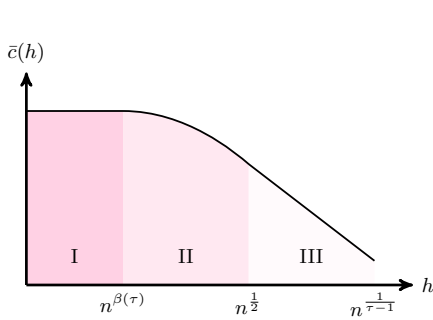


Figure 4.3: Clustering spectrum $h \mapsto \bar{c}(h)$ with three different ranges for h : the flat range, logarithmic decay, and the power-law decay.

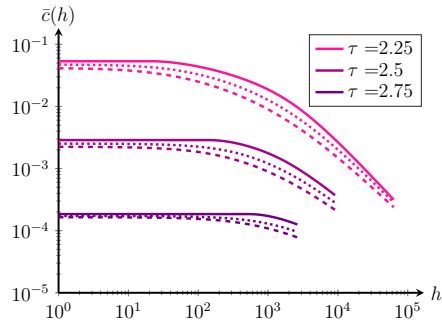


Figure 4.4: $\bar{c}(h)$ for $r(u) = \min(u, 1)$ (line), $r(u) = u/(1 + u)$ (dashed) and $r(u) = 1 - e^{-u}$ (dotted), obtained by calculating (4.A.6) numerically.

This range shows relatively slow, logarithmic decay in the clustering spectrum, and is clearly visible in the ten data sets.

Range III considers weights above the structural cutoff, when the restrictive effect of degree-degree correlations becomes more evident. In this range we find that

$$\bar{c}(h) \approx \frac{1}{(3 - \tau)^2} (h_s/h)^{6-2\tau} h_s^{4-2\tau} \propto n^{5-2\tau} h^{2\tau-6}, \quad h \geq \sqrt{n}, \quad (4.3.3)$$

hence power-law decay with exponent $\alpha = 2(3 - \tau)$. Such power-law decay has been observed in many real-world networks [58, 131, 140, 187, 196, 212], where most networks were found to have power-law exponent close to one. The asymptotic relation (4.3.3) shows that the exponent α decreases with τ and takes values in the entire range $(0, 2)$. Table 4.1 contains estimated values of α for the ten data sets.

Other connection probabilities. In Chapter 3 we have presented a class of functions $r(u) = uf(u)$, $u \geq 0$, so that

$$p(h, h') = r(u) \quad \text{with} \quad u = hh'/h_s^2 \quad (4.3.4)$$

has appropriate monotonicity properties. The maximal member $r(u) = \min(u, 1)$ of this class yields $p(h, h')$ in (4.2.3) and is representative of the whole class, while allowing explicit computation and asymptotic analysis of $\bar{c}(h)$ as in Chapter 3 and this chapter. Figure 4.4 shows that other asymptotically equivalent choices such as $r(u) = u/(1 + u)$ and $r(u) = 1 - e^{-u}$ have comparable clustering spectra. A minor difference is that the choice $r(u) = \min(1, u)$ forces $\bar{c}(h)$ to be constant on the range $h \leq n^{\beta(\tau)}$, while the other two choices show a gentle decrease.

4.4 Energy minimization

We now explain why the clustering spectrum splits into three ranges, using an argument that minimizes the energy needed to create triangles among vertices with specific weights.

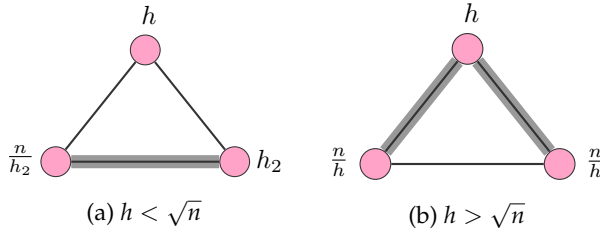


Figure 4.5: Orders of magnitude of the major contributions in the different h -ranges. The highlighted edges are present with asymptotically positive probability.

In all three ranges for h , there is one type of ‘most likely’ triangle, as shown in Figure 4.5. This means that most triangles containing a vertex v with weight h are triangles with two other vertices v' and v'' with weights h' and h'' of specific sizes, depending on h . The probability that a triangle is present between v , v' and v'' can be written as

$$\min\left(1, \frac{hh'}{n\mathbb{E}[h]}\right) \min\left(1, \frac{hh''}{n\mathbb{E}[h]}\right) \min\left(1, \frac{h'h''}{n\mathbb{E}[h]}\right). \quad (4.4.1)$$

While the probability that such a triangle exists among the three vertices thus increases with h' and h'' , the number of such vertices decreases with h' and h'' because vertices with higher weights are rarer. Therefore, the maximum contribution to $\bar{c}(h)$ results from a trade-off between large enough h', h'' for the occurrence of the triangle to be likely, and h', h'' small enough to have enough copies. Thus, having $h' > n\mathbb{E}[h]/h$ is not optimal, since then the probability that an edge exists between v and v' no longer increases with h' . This results in the bound

$$h', h'' \leq n\mathbb{E}[h]/h. \quad (4.4.2)$$

Similarly, $h'h'' > n\mathbb{E}[h]$ is also suboptimal, since then further increasing h' and h'' does not increase the probability of an edge between v' and v'' . This gives as a second bound

$$h'h'' \leq n\mathbb{E}[h]. \quad (4.4.3)$$

In Ranges I and II, $h < \sqrt{n\mathbb{E}[h]}$, so that $n\mathbb{E}[h]/h > \sqrt{n\mathbb{E}[h]}$. In this situation we reach bound (4.4.3) before we reach bound (4.4.2). Therefore, the maximum contribution to $\bar{c}(h)$ comes from $h'h'' \approx n$, where also $h', h'' < n\mathbb{E}[h]/h$ because of the bound (4.4.2). Here the probability that the edge between v' and v'' exists is large, while the other two edges have a small probability to be present, as shown in Figure 4.5a. For h in Range I, the bound (4.4.2) is superfluous, since in this regime $n\mathbb{E}[h]/h > h_c$, while the network does not contain vertices with weights larger than h_c . This bound indicates the minimal value of h' such that an h -vertex is guaranteed to be connected to an h' -vertex. Thus, vertices in Range I are not even guaranteed to have connections to the highest-degree vertices, hence they are not affected by the single-edge constraints. Therefore, the value of $\bar{c}(h)$ in Range I is independent of h .

In Range III, $h > \sqrt{n\mathbb{E}[h]}$, so that $n\mathbb{E}[h]/h < \sqrt{n\mathbb{E}[h]}$. Therefore, we reach bound (4.4.2) before we reach bound (4.4.3). Thus, we maximize the contribution to

	n	τ	g.o.f.	α
Hudong	1.984.484	2,30	0.00	0,85
Baidu	2.141.300	2,29	0.00	0,80
Wordnet	146.005	2,47	0.00	1,01
Google web	875.713	2,73	0.00	1,03
AS-Skitter	1.696.415	2,35	0.06	1,12
TREC-WT10g	1.601.787	2,23	0.00	0,99
Wiki-talk	2.394.385	2,46	0.00	1,54
Catster/Dogster	623.766	2,13	0.00	1,20
Gowalla	196.591	2,65	0.80	1,24
Youtube	1.134.890	2,22	0.00	1,05

Table 4.1: Data sets. n denotes the number of vertices, τ the exponent of the tail of the degree distribution estimated by the method proposed in [66] together with the goodness-of-fit criterion proposed in [66] (when the goodness of fit is at least 0.10, a power-law tail cannot be rejected), and α denotes the exponent of $c(k)$.

the number of triangles by choosing $h', h'' \approx n\mathbb{E}[h]/h$. Then the probability that the edge from v to v' and from v to v'' is present is large, while the probability that the edge between v' and v'' exists is small, as illustrated in Figure 4.5b. We make this energy minimization argument more precise in Chapter 5.

4.5 Convergence rate

We next ask how large networks should be, or become, before they reveal the features of the universal clustering spectrum. In other words, while the results in this chapter are shown for the large- n limit, for what finite n -values can we expect to see the different ranges and clustering decay? To bring networks of different sizes n on a comparable footing, we consider

$$\sigma_n(t) = \frac{\ln(\bar{c}(h)/\bar{c}(h_c))}{\ln(n\mathbb{E}[h])}, \quad h = (n\mathbb{E}[h])^t, \quad (4.5.1)$$

for $0 \leq t \leq \frac{1}{\tau-1}$. The slope of $\sigma_n(t)$ can be interpreted as a measure of the decay of $\bar{c}(h)$ at $h = (n\mathbb{E}[h])^t$, and all curves share the same right end of the spectrum. Figure 4.6 shows this rescaled clustering spectrum for synthetic networks generated with the rank-1 inhomogeneous random graph with $\tau = 2.25$. Already 10^4 vertices reveal the essential features of the spectrum: the decay and the three ranges. Increasing the network size further to 10^5 and 10^6 vertices shows that the spectrum settles on the limiting curve. Here we note that the real-world networks reported in Figures 4.1 and 4.2 are also of order 10^5 - 10^6 vertices, see Table 4.1.

Limiting form of $\sigma_n(t)$ and finite-size effects. Consider $\sigma_n(t)$ of (4.5.1). It follows from the detailed form of (4.3.1) and (4.3.3) that

$$\sigma_n(0) = \frac{\ln(\bar{c}(0)/\bar{c}(h_c))}{\ln(n\mathbb{E}[h])} = \gamma + \frac{\ln(\beta y)}{y}, \quad (4.5.2)$$

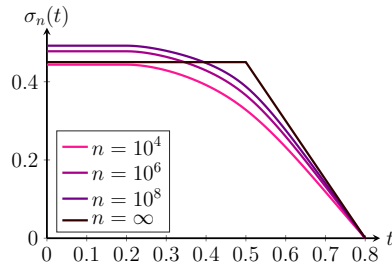


Figure 4.6: $\sigma_n(t)$ for $n = 10^4, 10^6$ and 10^8 together with the limiting function, using $\tau = 2.25$, for which $\frac{1}{\tau-1} = 0.8$.

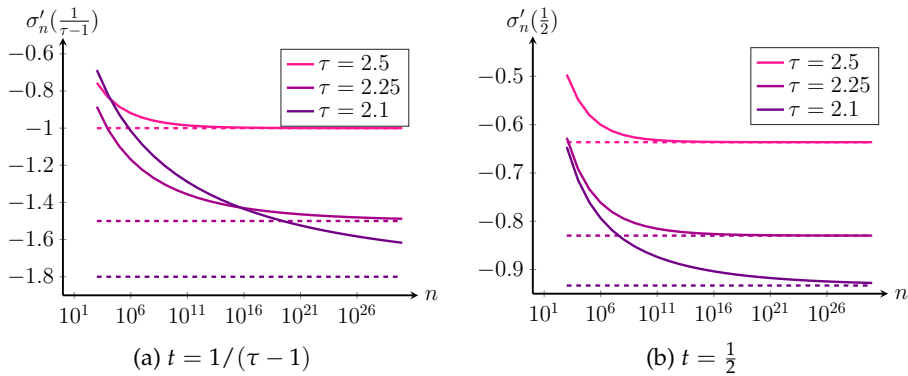


Figure 4.7: $\sigma'_n(t)$ plotted against n for two values of t . The dashed line gives the limiting value of $\sigma'_n(t)$ as $n \rightarrow \infty$.

where

$$\gamma = \frac{(3-\tau)^2}{\tau-1}, \quad \beta = (\tau-2)\gamma, \quad y = \ln(n\mathbb{E}[h]). \quad (4.5.3)$$

We have that $\sigma_n(0) \rightarrow \gamma$ as $n \rightarrow \infty$. The right-hand side of (4.5.2) exceeds this limit γ for $y > 1/\beta$ with a maximum excess β/e for $y = e/\beta$, or equivalently $e^y = n\mathbb{E}[h] = \exp(e/\beta)$. Therefore, n needs to be much larger than $\exp(e/\beta)$ for $\sigma_n(0)$ to be close to its limiting value. This explains why the excess of $\sigma_n(0)$ over its limit value in Figure 4.6 with $e^{e/\beta} = 3 \times 10^{10}$ when $\tau = 9/4$ persists even for large values of n .

Using (4.3.1), (4.3.2) and (4.3.3), we obtain

$$\lim_{n \rightarrow \infty} \sigma_n(t) = \begin{cases} \gamma, & 0 \leq t \leq \frac{1}{2}, \\ \gamma + (3-\tau)(1-2t), & \frac{1}{2} \leq t \leq \frac{1}{\tau-1}. \end{cases} \quad (4.5.4)$$

Hence, some of the detailed information that is present in (4.3.1), (4.3.2) and (4.3.3), disappears when taking the limit as in (4.5.4). This is in particular so for the $\ln n$ -factor in (4.3.1) and the logarithmically decaying factor $\ln(n^2/h)$ in Region II.

Behavior of $\sigma'_n(t)$. By (4.5.1),

$$\bar{c}(h) = \bar{c}(h_c)(n\mathbb{E}[h])^{\sigma_n(t)}, \quad h = (n\mathbb{E}[h])^t. \quad (4.5.5)$$

When we fix a t_0 and linearize $\sigma_n(t)$ around t_0 , we get

$$\bar{c}(h) \approx \bar{c}(h_c)(n\mathbb{E}[h])^{\sigma_n(t_0)+(t-t_0)\sigma'_n(t_0)} = \bar{c}(h_0)(h/h_0)^{\sigma'_n(t_0)} \quad (4.5.6)$$

so that $\sigma'_n(t) = \frac{d}{dt}\sigma_n(t)$ is a measure for the decay rate of $\bar{c}(h)$ at $h = h_0 = (n\mathbb{E}[h])^{t_0}$. In Appendix 4.B we again use the integral expression (4.2.4) to characterize the rate of convergence of $\sigma'_n(t)$ as a function of n .

Figure 4.7 shows the values of $\sigma'_n(1/2)$ and $\sigma'_n(1/(\tau-1))$ for finite-size networks together with its limiting value. Indeed, extreme n -values are required for statistically reliable slope estimates for e.g. t -values of $\frac{1}{2}$ and $\frac{1}{\tau-1}$. For example, when $\tau = 2.25$, Figure 4.7a shows that n needs to be of the order 10^{16} for the slope to be ‘close’ to its limiting value -1.5. When for example $n = 10^6$, the slope is much smaller: approximately -1.1. This makes statistical estimation of the true underlying power-law exponent α of $\bar{c}(h)$ extremely challenging, especially for the relevant regime τ close to 2, because enormous amounts of data should be available to get sufficient statistical accuracy. Most data sets, even the largest available networks used in this chapter, are simply not large enough to have sufficiently many samples from the large-degree region to get a statistically accurate estimate of the power-law part. This also explains why based on smaller data sets it is common to assume that α is roughly one [58, 131, 140, 187, 196, 212]. Comparing Figure 4.7a and Figure 4.7b shows that the convergence to the limiting value is significantly faster at the point $t = 1/2$ than at the point $t = 1/(\tau-1)$.

The extremely slow convergence to the limiting curve for $n = \infty$ is a well documented property of certain clustering measures [37, 121, 201], such as the global clustering coefficient studied in Chapter 3. The behavior of $\sigma_n(t)$ and $\sigma'_n(t)$ show that the local clustering curve $c(k)$ also converges extremely slowly to its limiting value. Therefore, the estimates in Table 4.1 only serve as indicative values of α . Finally, observe that Range II disappears in the limiting curve, due to the rescaling in (4.5.1), but again only for extreme n -values. Because this chapter is about structure rather than statistical estimation, the slow convergence in fact provides additional support for the persistence of Range II in Figures 4.1 and 4.2.

Table 4.1 also shows that the relation $\alpha = -2(3-\tau)$ is inaccurate for the real-world data sets. One explanation for this inaccuracy is that the real-world networks might not follow pure power-law distributions, as measured by the goodness of fit criterion in Table 4.1, and visualized in Appendix 4.D. Furthermore, real-world networks are usually highly clustered and contain community structures, whereas the rank-1 inhomogeneous random graph is locally tree-like. These modular structures may explain, for example, why the power-law decay of the rank-1 inhomogeneous random graph is less pronounced in the three social networks of Figure 4.2. It is remarkable that despite these differences between rank-1 inhomogeneous random graphs and real-world networks, the global shape of the $c(k)$ curve of the rank-1 inhomogeneous random graph is still visible in these heavy-tailed real-world networks.

4.6 Discussion

The decaying clustering spectrum has often been contributed to the restriction that no two vertices have more than one edge connecting them [146, 160, 161, 175, 176]. The physical intuition is that the single-edge constraint leads to fewer connections between high-degree vertices than anticipated based on randomly assigned edges. We have indeed confirmed this intuition for the rank-1 inhomogeneous random graph, not only through analytically revealing the universal clustering curve, but also by providing an alternative derivation of the three ranges based on energy minimization and structural correlations.

In Chapter 5 we will show that the clustering spectrum revealed using the rank-1 inhomogeneous random graph also appears for a second widely studied null model. This second model cannot be the configuration model (CM), which preserves the degree distribution by making connections between vertices in the most random way possible [41, 164]. Indeed, because of the random edge assignment, the CM has no degree correlations, leading in the case of scale-free networks with diverging second moment to uncorrelated networks with non-negligible fractions of self-loops and multiple edges. This picture changes dramatically when self-loops and multiple edges are avoided, a restriction mostly felt by the high-degree vertices, who can no longer establish multiple edges among each other.

We therefore consider the erased configuration model (ECM) instead. While the ECM removes some of the edges in the graph, only a small proportion of the edges is removed, so that the degree of vertex j in ECM is still close to D_j [106, Chapter 7]. In the ECM, the probability that a vertex with degree D_i is connected to a vertex with degree D_j can be approximated by $1 - e^{-D_i D_j / \mathbb{E}[D]^n}$ [108, Eq.(4.9)]. Therefore, we expect the ECM and the rank-1 inhomogeneous random graph to have similar properties (see e.g. Chapter 5) when we choose

$$p(h, h') = 1 - e^{-hh' / n\mathbb{E}[h]} \approx \frac{hh'}{n\mathbb{E}[h]}. \quad (4.6.1)$$

Figure 4.8 illustrates how both null models generate highly similar spectra, which provides additional support for the claim that the clustering spectrum is a universal property of these simple scale-free network models, similarly to the average nearest-neighbor degree $a(k)$ studied in Chapter 2. The ECM is more difficult to deal with compared to rank-1 inhomogeneous random graphs, since edges in ECM are not independent. In Chapter 5, we show that these dependencies vanish for the $k \mapsto c(k)$ curve, so that the ECM indeed has a similar clustering spectrum as the rank-1 inhomogeneous random graph.

The ECM and the rank-1 inhomogeneous random graph are both null models with soft constraints on the degrees. Putting hard constraints on the degrees with uniform random graphs has the nice property that simple graphs generated using this null model are uniform samples of all simple graphs with the same degree sequence. Constructing such uniform random graphs is notoriously hard when the second moment of the degrees is diverging, since the CM will yield many multiple edges [91, 151, 213], but in Chapter 5 we show that our theoretical results for the $k \mapsto c(k)$ curve for soft-constraint models carry over to these uniform simple graphs.

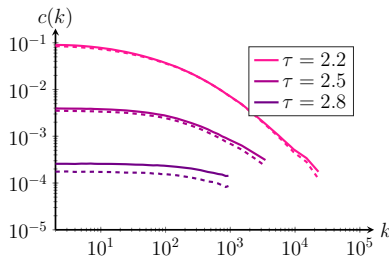


Figure 4.8: $c(k)$ for a rank-1 inhomogeneous random graph with connection probabilities (4.6.1) (solid line) and an erased configuration model (dashed line). The presented values of $c(k)$ are averages over 10^4 realizations of networks of size $n = 10^5$.

In this chapter we have investigated the presence of triangles in the rank-1 inhomogeneous random graph. We have shown that by first conditioning on the vertex degree, there is a unique ‘most likely’ triangle with two other vertices of specific degrees. We have not only explained this insight heuristically, but it is also reflected in the elaborate analysis of the double integral for $\bar{c}(h)$ in Appendix 4.A. As such, we have introduced an intuitive and tractable mathematical method for asymptotic triangle counting. In Chapter 7 we show that this method carries over to counting other subgraphs, such as squares, or complete graphs of larger sizes. For any given subgraph, first conditioning on the vertex degree, we again find specific configurations that are most likely.

4.A Derivation for the three ranges

In this appendix, we compute $\bar{c}(h)$ in (4.2.4), and we show that $\bar{c}(h)$ can be approximated by (4.3.1), (4.3.2), or (4.3.3), depending on the value of h . Throughout the appendix, we assume that $p(h, h') = \min(1, hh'/h_s^2)$ and $\rho(h) = Ch^{-\tau}$. Then, the derivation of $\bar{c}(h)$ in [202] yields

$$\begin{aligned} c(h) &= \frac{\int_1^{h_c} \int_1^{h_c} \rho(h') p(h, h') \rho(h'') p(h, h'') p(h', h'') dh'' dh'}{\left[\int_1^{h_c} \rho(h') p(h, h') dh' \right]^2} \\ &= \frac{\int_1^{h_c} \int_1^{h_c} (h'h'')^{-\tau} \min\left(\frac{hh'}{h_s^2}, 1\right) \min\left(\frac{hh''}{h_s^2}, 1\right) \min\left(\frac{h'h''}{h_s^2}, 1\right) dh'' dh'}{\left[\int_1^{h_c} (h')^{-\tau} \min\left(\frac{hh'}{h_s^2}, 1\right) dh' \right]^2}. \end{aligned} \quad (4.A.1)$$

Adopting the standard choices as in Chapter 3

$$h_s = \sqrt{n\mathbb{E}[h]}, \quad h_c = (n\mathbb{E}[h])^{1/(\tau-1)}, \quad (4.A.2)$$

and setting $h_{\min} = 1$ gives

$$\mathbb{E}[h] = \frac{\tau - 1}{\tau - 2} \frac{1 - n^{2-\tau}}{1 - n^{1-\tau}}. \quad (4.A.3)$$

For ease of notation in the proofs below, we will use

$$a = h_s^{-1} = (n\mathbb{E}[h])^{-1/2}, \quad b = h_c/h_s = (n\mathbb{E}[h])^{\frac{3-\tau}{2(\tau-1)}}, \quad (4.A.4)$$

and

$$r(u) = \min(u, 1). \quad (4.A.5)$$

In this notation, (4.A.1) can be succinctly written as

$$\bar{c}(h) = \frac{\int_a^b \int_a^b (xy)^{-\tau} r(ahx) r(ahy) r(xy) dx dy}{\left[\int_a^b x^{-\tau} r(ahx) dx \right]^2}. \quad (4.A.6)$$

Because of the four min operators in the expression (4.A.1), we have to consider various h -ranges. We compute the value of $\bar{c}(h)$ in these three ranges one by one.

Range I: $h < h_s^2/h_c$. We now show that in this range

$$\bar{c}(h) \approx \frac{\tau-2}{3-\tau} h_s^{4-2\tau} \ln(h_c^2/h_s^2) \propto n^{2-\tau} \ln n, \quad (4.A.7)$$

which proves (4.3.1).

This range corresponds to $h < 1/(ab)$ with a and b as in (4.A.4). In this range, $r(ahx) = ahx$ and $r(ahy) = ah y$ for all $x \in [a, b]$. This yields for $c(h)$

$$c(h) = \frac{\int_a^b \int_a^b (xy)^{1-\tau} r(xy) dx dy}{\left[\int_a^b x^{1-\tau} dx \right]^2}. \quad (4.A.8)$$

For the denominator we compute

$$\int_a^b x^{1-\tau} dx = \frac{a^{2-\tau} - b^{2-\tau}}{\tau-2}. \quad (4.A.9)$$

Since $a \ll b$, this can be approximated as

$$\frac{a^{2-\tau} - b^{2-\tau}}{\tau-2} \approx \frac{a^{2-\tau}}{\tau-2}. \quad (4.A.10)$$

We can compute the numerator of (4.A.8) as

$$\begin{aligned} & \int_a^b \int_a^b (xy)^{1-\tau} r(xy) dx dy \\ &= \int_a^{1/b} \int_a^b (xy)^{2-\tau} dx dy + \int_{1/b}^b \int_a^{1/x} (xy)^{2-\tau} dx dy + \int_{1/b}^b \int_{1/x}^b (xy)^{1-\tau} dx dy \\ &= \frac{(b^{\tau-3} - a^{3-\tau})(b^{3-\tau} - a^{3-\tau})}{(3-\tau)^2} + \frac{1}{3-\tau} \left(\ln(b^2) - \frac{a^{3-\tau}(b^{3-\tau} - b^{\tau-3})}{3-\tau} \right) \\ & \quad + \frac{1}{2-\tau} \left(\frac{b^{2-\tau}(b^{2-\tau} - b^{\tau-2})}{2-\tau} - \ln(b^2) \right) \\ &= \frac{\ln(b^2)}{(3-\tau)(\tau-2)} - \frac{1-b^{4-2\tau}}{(\tau-2)^2} + \frac{1-2(ab)^{3-\tau} + a^{6-2\tau}}{(3-\tau)^2}. \end{aligned} \quad (4.A.11)$$

The first of these three terms dominates when

$$\frac{3 - \tau}{\tau - 1} \frac{\ln(n\mathbb{E}[h])}{(3 - \tau)(\tau - 2)} \gg \frac{1}{(\tau - 2)^2} \quad (4.A.12)$$

and

$$\frac{3 - \tau}{\tau - 1} \frac{\ln(n\mathbb{E}[h])}{(3 - \tau)(\tau - 2)} \gg \frac{1}{(3 - \tau)^2}, \quad (4.A.13)$$

where we have used that $b^2 = (n\mathbb{E}[h])^{(3-\tau)/(\tau-1)}$. Thus, when $\ln(n\mathbb{E}[h])$ is large compared to $(\tau - 1)/(\tau - 2)$ and $(\tau - 1)(\tau - 2)/(\tau - 3)^2$, we obtain

$$\bar{c}(h) \approx \frac{\tau - 2}{3 - \tau} a^{2\tau-4} \ln(b^2) \propto n^{2-\tau} \ln(n), \quad (4.A.14)$$

which proves (4.A.7).

Range II: $h_s^2/h_c < h < h_s$ In this range, we show that

$$\bar{c}(h) \approx h_s^{4-2\tau} \frac{\ln(h_s^2/h^2) + M}{(\tau - 2)(3 - \tau)} \propto n^{2-\tau} \left(\ln(n/h^2) + M \right), \quad (4.A.15)$$

for some positive constant M , which proves (4.3.2).

This range corresponds to $(ab)^{-1} < h < a^{-1}$. For these values of h , we have $ahx, ah_y = 1$ for $x, y = (ah)^{-1} \in (1, b)$ and $xy = 1$ for $y = 1/x \in [a, b]$ when $b^{-1} < x < b$. Then for the denominator of (4.A.6) we compute

$$\begin{aligned} & \int_a^{1/(ah)} ahx^{1-\tau} dx + \int_{1/(ah)}^b x^{-\tau} dx \\ &= \frac{1}{\tau - 2} (a^{3-\tau} h - (ah)^{\tau-1}) + \frac{1}{\tau - 1} ((ah)^{\tau-1} - b^{1-\tau}) \\ &= ah \left(\frac{a^{2-\tau}}{\tau - 2} - \frac{(ah)^{\tau-2}}{(\tau - 1)(\tau - 2)} - \frac{b^{1-\tau}/(ah)}{\tau - 1} \right). \end{aligned} \quad (4.A.16)$$

Splitting up the integral in the numerator results in

$$\begin{aligned} \text{Num}(h) &= \int_a^b \int_a^b (xy)^{-\tau} r(ahx)r(ah_y)r(xy) dx dy \\ &= \int_{1/(ah)}^b \int_{1/(ah)}^b (xy)^{-\tau} dy dx + 2ah \int_{1/(ah)}^b \int_{1/x}^{1/(ah)} (xy)^{-\tau} y dy dx \\ &\quad + 2ah \int_{1/(ah)}^b \int_a^{1/x} (xy)^{1-\tau} y dy dx + a^2 h^2 \int_{ah}^{1/(ah)} \int_a^{1/x} (xy)^{2-\tau} dy dx \\ &\quad + a^2 h^2 \int_{ah}^{1/(ah)} \int_{1/x}^{1/(ah)} (xy)^{1-\tau} dy dx + a^2 h^2 \int_a^{ah} \int_a^{1/(ah)} (xy)^{2-\tau} dy dx \\ &=: I_1 + I_2 + I_3 + I_4 + I_5 + I_6, \end{aligned} \quad (4.A.17)$$

where the factors 2 arise by symmetry of the integrand in x and y . Computing these integrals yields

$$I_1 = a^2 h^2 \left(\frac{(ah)^{\tau-2} - a^{-1} b^{1-\tau} h^{-1}}{\tau-1} \right)^2, \quad (4.A.18)$$

$$I_2 = 2a^2 h^2 \left(\frac{1-1/(abh)}{\tau-2} - \frac{(ah)^{2\tau-4}}{(\tau-1)(\tau-2)} \left(1 - (abh)^{1-\tau} \right) \right), \quad (4.A.19)$$

$$I_3 = 2a^2 h^2 \left(\frac{1-1/(abh)}{3-\tau} - \frac{h^{\tau-3} (1 - (abh)^{2-\tau})}{(3-\tau)(\tau-2)} \right), \quad (4.A.20)$$

$$I_4 = a^2 h^2 \left(\frac{\ln((ah)^{-2})}{3-\tau} + \frac{(a^2 h)^{3-\tau} - h^{\tau-3}}{(3-\tau)^2} \right), \quad (4.A.21)$$

$$I_5 = a^2 h^2 \left(\frac{\ln((ah)^{-2})}{\tau-2} - \frac{1 - (ah)^{2\tau-4}}{(\tau-2)^2} \right), \quad (4.A.22)$$

$$I_6 = a^2 h^2 \left(\frac{1 - h^{\tau-3} + a^{6-2\tau} - (a^2 h)^{3-\tau}}{(3-\tau)^2} \right). \quad (4.A.23)$$

Since $ah < 1 < ahb$, the leading behavior of $\text{Num}(h)$ is determined by the terms involving $\ln((ah)^{-2})$ in I_3 and I_4 , all other terms being bounded. Retaining only these dominant terms, we get

$$\text{Num}(h) = a^2 h^2 \frac{\ln((ah)^{-2})}{(\tau-2)(3-\tau)} (1 + o(1)), \quad (4.A.24)$$

provided that $ah \rightarrow 0$ as $n \rightarrow \infty$. In terms of the variable t in $h = (nE[h])^t$, see (4.5.1), this condition holds when we restrict to $t \in [(\tau-2)/(\tau-1), \frac{1}{2} - \varepsilon]$ for any $\varepsilon > 0$. Furthermore, from (4.A.16),

$$\left(\int_a^b x^{-\tau} r(axx) dx \right)^2 = a^2 h^2 \left(\frac{a^{2-\tau}}{\tau-2} \right)^2 (1 + o(1)). \quad (4.A.25)$$

Hence, when $ah \rightarrow 0$, we have

$$\bar{c}(h) = \frac{\tau-2}{3-\tau} a^{2\tau-4} \ln((ah)^{-2}) (1 + o(1)) \propto n^{2-\tau} \ln(N/h^2). \quad (4.A.26)$$

We compute $\bar{c}(h = 1/a)$ asymptotically by retaining only all constant terms between brackets in (4.A.18)-(4.A.23) since all other terms vanish or tend to 0 as $n \rightarrow \infty$. This gives

$$\begin{aligned} \text{Num}(h = 1/a) &= a^2 h^2 \left(\frac{1}{(\tau-1)^2} + \frac{2}{\tau-2} - \frac{2}{(\tau-1)(\tau-2)} + \frac{2}{3-\tau} + \frac{1}{(3-\tau)^2} \right) (1 + o(1)) \\ &= P a^2 h^2 (1 + o(1)), \end{aligned} \quad (4.A.27)$$

where $P = \frac{1}{(\tau-1)^2} + \frac{1}{(3-\tau)^2} + \frac{2}{\tau-1} + \frac{2}{3-\tau}$. Together with (4.A.25), we find

$$\bar{c}(h = 1/a) = P(\tau-2)^2 a^{2\tau-4} (1 + o(1)) \propto n^{2-\tau}. \quad (4.A.28)$$

In Proposition 3.1, it has been shown that $\bar{c}(h)$ decreases in h , and then (4.A.15) follows from (4.A.26) and (4.A.28).

Range III: $h_s < h < h_c$. We now show that when $h_s < h < h_c$, then

$$\bar{c}(h) \approx \frac{1}{(3-\tau)^2} (h_s/h)^{6-2\tau} h_s^{4-2\tau} \propto n^{5-2\tau} h^{2\tau-6}, \quad (4.A.29)$$

which proves (4.3.3).

This range corresponds to $1/a < h < b/a$. The denominator of (4.A.6) remains the same as in the previous range and is given by (4.A.16). Splitting up the integral in the numerator of (4.A.6) now results in

$$\begin{aligned} \text{Num}(h) &= \int_a^b \int_a^b (xy)^{-\tau} r(ax)r(ahy)r(xy) dx dy \\ &= \int_{1/(ah)}^{ah} \int_{1/x}^b (xy)^{-\tau} dy dx + \int_{ah}^b \int_{1/(ah)}^b (xy)^{-\tau} dy dx \\ &\quad + \int_{1/(ah)}^{ah} \int_{1/(ah)}^{1/x} (xy)^{1-\tau} dy dx + 2ah \int_{ah}^b \int_{1/x}^{1/(ah)} (xy)^{-\tau} y dy dx \\ &\quad + 2ah \int_{1/(ah)}^{ah} \int_a^{1/(ah)} (xy)^{1-\tau} y dy dx + 2ah \int_{ah}^b \int_a^{1/x} (xy)^{1-\tau} y dy dx \\ &\quad + a^2 h^2 \int_a^{1/(ah)} \int_a^{1/(ah)} (xy)^{2-\tau} dy dx \\ &=: I_1 + I_2 + I_3 + I_4 + I_5 + I_6 + I_7. \end{aligned} \quad (4.A.30)$$

Computing these integrals yields

$$I_1 = a^2 h^2 \left((ah)^{-2} \frac{\ln(a^2 h^2)}{\tau-1} + \frac{b^{1-\tau} ((ah)^{-\tau-1} - (ah)^{\tau-3})}{(\tau-1)^2} \right), \quad (4.A.31)$$

$$I_2 = a^2 h^2 \left(\frac{(ah)^{-2} + b^{2-2\tau} (ah)^{-2}}{(\tau-1)^2} - \frac{b^{1-\tau} ((ah)^{\tau-3} + (ah)^{-\tau-1})}{(\tau-1)^2} \right), \quad (4.A.32)$$

$$I_3 = a^2 h^2 \left(-(ah)^{-2} \frac{\ln(a^2 h^2)}{\tau-2} + \frac{(ah)^{2\tau-6} - (ah)^{-2}}{(\tau-2)^2} \right), \quad (4.A.33)$$

$$I_4 = 2a^2 h^{-2} \left(-\frac{(abh)^{-1}}{\tau-2} + \frac{(ah)^{-2}}{\tau-1} + \frac{b^{1-\tau} (ah)^{\tau-3}}{(\tau-1)(\tau-2)} \right), \quad (4.A.34)$$

$$I_5 = 2a^2 h^2 \left(\frac{(ah)^{2\tau-6} + h^{1-\tau} a^{4-2\tau} - h^{\tau-3} - (ah)^{-2}}{(3-\tau)(\tau-2)} \right), \quad (4.A.35)$$

$$I_6 = 2a^2 h^2 \left(\frac{(ab)^{2-\tau} h^{-1} - h^{1-\tau} a^{4-2\tau}}{(3-\tau)(\tau-2)} - \frac{(abh)^{-1} - (ah)^{-2}}{3-\tau} \right), \quad (4.A.36)$$

$$I_7 = a^2 h^2 \left(\frac{a^{6-2\tau} - 2h^{\tau-3} + (ah)^{2\tau-6}}{\tau-3} \right). \quad (4.A.37)$$

A careful inspection of the terms between brackets in (4.A.31) and (4.A.37) shows that the terms involving $(ah)^{2\tau-6}$ are dominant when $ah \rightarrow \infty$. In terms of the

variable t in $h = (n\mathbb{E}[h])^t$, see (4.5.1), we have that $ah \rightarrow \infty$ when we restrict to $t \in [\frac{1}{2} + \varepsilon, 1/(\tau - 1)]$ for any $\varepsilon > 0$. When we retain only these dominant terms, we have, when $ah \rightarrow \infty$,

$$\begin{aligned} \text{Num}(h) &= a^2 h^2 (ah)^{2\tau-6} \left(\frac{1}{(\tau-2)^2} + \frac{2}{(3-\tau)(\tau-2)} + \frac{1}{(3-\tau)^2} \right) (1 + o(1)) \\ &= a^2 h^2 \frac{(ah)^{2\tau-6}}{(\tau-2)^2(3-\tau)^2} (1 + o(1)). \end{aligned} \quad (4.A.38)$$

Using (4.A.25) again, we get, when $ah \rightarrow \infty$,

$$\bar{c}(h) = \frac{1}{(3-\tau)^2} (ah)^{2\tau-6} a^{2\tau-4} (1 + o(1)) \propto n^{5-2\tau} h^{2\tau-6}. \quad (4.A.39)$$

Furthermore, $c(1/a)$ is given by (4.A.28), while $\bar{c}(h)$ decreases in h . This gives (4.A.29).

4.B Exact and asymptotic decay rate of $\bar{c}(h)$ at $h = h_c$ and $h = h_s$

In this appendix, we compute an exact expression for $\sigma'_n(t)$ at $t = \frac{1}{\tau-1}$ and $t = \frac{1}{2}$. We then compute its limit as $n \rightarrow \infty$ and we show that this limit is a lower bound for $\sigma'_n(t)$. More precisely, we show the following result:

Proposition 4.1. *Let a and b be as in (4.A.4). Then,*

$$\sigma'_n \left(\frac{1}{\tau-1} \right) = -2 \left(\frac{A + \frac{3-\tau}{\tau-2} C}{A + \frac{4-\tau}{\tau-2} C} - \frac{D}{E+D} \right), \quad (4.B.1)$$

where

$$A = \frac{1}{b^2} \left(\frac{-\ln(b^2)}{(\tau-1)(\tau-2)} - \frac{1-b^{2(1-\tau)}}{(\tau-1)^2} + \frac{b^{2(\tau-2)}-1}{(\tau-2)^2} \right), \quad (4.B.2)$$

$$C = \left(\frac{b^{\tau-3} - a^{3-\tau}}{3-\tau} \right)^2, \quad (4.B.3)$$

$$D = \frac{1}{b} \frac{b^{\tau-1} - b^{1-\tau}}{\tau-1}, \quad (4.B.4)$$

$$E = \frac{a^{2-\tau} - b^{\tau-2}}{\tau-2}. \quad (4.B.5)$$

Furthermore, for all n ,

$$\sigma'_n \left(\frac{1}{\tau-1} \right) > \lim_{M \rightarrow \infty} \sigma'_M \left(\frac{1}{\tau-1} \right) = -2(3-\tau). \quad (4.B.6)$$

The limit in (4.B.6) is consistent with the limiting value of $\sigma_n(t)$ of (4.5.4).

We let $h_c = (n\mathbb{E}[h])^{1/(\tau-1)}$, where we assume that n is so large that $h_c \leq n$. This requires n to be of the order $(1/\varepsilon)^{1/\varepsilon}$, where $\varepsilon = \tau - 2$. To start the proof of Proposition 4.1, note that in the a, b notation of (4.A.4),

$$\bar{c}(h) = \frac{K(h)}{J(h)}, \quad 0 \leq h \leq h_c, \quad (4.B.7)$$

where

$$K(h) = \int_a^b \int_a^b (xy)^{2-\tau} f(ahx) f(ahy) f(xy) dx dy, \quad (4.B.8)$$

$$J(h) = \left(\int_a^b x^{1-\tau} f(ahx) dx \right)^2, \quad (4.B.9)$$

with $f(u) = \min(1, u^{-1})$. Note that $r(u) = uf(u)$, see (4.A.5). We compute

$$\begin{aligned} \sigma'_n(t) &= \frac{d}{dt} \left(\frac{\ln(c((n\mathbb{E}[h])^t)/\bar{c}(h_c))}{\ln(n\mathbb{E}[h])} \right) \\ &= (n\mathbb{E}[h])^t \ln(n\mathbb{E}[h]) \frac{c'((n\mathbb{E}[h])^t)}{c((n\mathbb{E}[h])^t) \ln(n\mathbb{E}[h])} \\ &= h \frac{c'(h)}{\bar{c}(h)}, \quad h = (n\mathbb{E}[h])^t, \end{aligned} \quad (4.B.10)$$

where the prime on c indicates differentiation with respect to h . With (4.B.7) we get

$$\frac{c'(h)}{\bar{c}(h)} = \frac{K'(h)}{K(h)} - \frac{J'(h)}{J(h)}, \quad (4.B.11)$$

and we have to evaluate $K(h), K'(h), J(h)$ and $J'(h)$ at

$$h = h_c = b/a. \quad (4.B.12)$$

Lemma 4.1.

$$K(h_c) = A + \frac{4-\tau}{2-\tau} C, \quad K'(h_c) = \frac{-2a}{b} \left(A + \frac{3-\tau}{\tau-2} C \right), \quad (4.B.13)$$

$$J(h_c) = (D + E)^2, \quad J'(h_c) = -\frac{2a}{b} (D + E) D, \quad (4.B.14)$$

with A, C, D, E as in (4.B.2)–(4.B.5).

From Lemma 4.1, (4.B.10) and (4.B.12) we get (4.B.1) in Proposition 4.1.

Proof of Lemma 4.1. Since $h_c = b/a$,

$$K(h_c) = \int_a^b \int_a^b (xy)^{2-\tau} f(bx) f(by) f(xy) dx dy. \quad (4.B.15)$$

With $f(u) = \min(1, u^{-1})$ we split up the integration range $[a, b] \times [a, b]$ into the four regions $[a, 1/b] \times [a, 1/b]$, $[1/b, b] \times [1/b, b]$, $[1/b, b] \times [a, 1/b]$ and $[a, 1/b] \times [1/b, b]$, where we observe that $a \leq 1/b \leq 1 \leq b$. We first get

$$\begin{aligned} \int_a^{1/b} \int_a^{1/b} (xy)^{2-\tau} f(bx) f(by) f(xy) dx dy &= \int_a^{1/b} \int_a^{1/b} (xy)^{2-\tau} dx dy \\ &= \left(\frac{b^{\tau-3} - a^{3-\tau}}{3-\tau} \right)^2 = C. \end{aligned} \quad (4.B.16)$$

Next,

$$\begin{aligned} \int_{1/b}^b \int_{1/b}^b (xy)^{2-\tau} f(bx) f(by) f(xy) dx dy &= \int_{1/b}^b \int_{1/b}^b (xy)^{2-\tau} \frac{1}{bx} \frac{1}{by} f(xy) dx dy \\ &= \frac{1}{b^2} \int_{1/b}^b \int_{1/b}^b (xy)^{1-\tau} f(xy) dx dy. \end{aligned} \quad (4.B.17)$$

The remaining double integral with $\tau + 1$ instead of τ has been evaluated in (3.C.3) as

$$-\frac{\ln(b^2)}{(\tau-1)(\tau-2)} - \frac{1-b^{2(1-\tau)}}{(\tau-1)^2} + \frac{b^{2(\tau-2)-1}}{(\tau-2)^2} = b^2 A. \quad (4.B.18)$$

Finally, the two double integrals over $[1/b, b] \times [a, 1/b]$ and $[a, 1/b] \times [1/b, b]$ are by symmetry both equal to

$$\begin{aligned} \int_{1/b}^b \int_a^{1/b} (xy)^{2-\tau} f(bx) f(by) f(xy) dx dy &= \int_{1/b}^b \int_a^{1/b} (xy)^{2-\tau} \frac{1}{bx} \cdot 1 \cdot 1 dx dy \\ &= \frac{1}{b} \frac{b^{\tau-2} - b^{2-\tau}}{\tau-2} \frac{b^{\tau-3} - a^{3-\tau}}{3-\tau} \\ &= \frac{(b^{\tau-3} - a^{3-\tau})^2}{(\tau-2)(3-\tau)} = \frac{3-\tau}{\tau-2} C. \end{aligned} \quad (4.B.19)$$

Here we have used that, see (4.A.4),

$$b^{1-\tau} = a^{3-\tau}. \quad (4.B.20)$$

Now the expression in (4.B.13) for $K(h_c)$ follows.

To evaluate $K'(h_c)$, we observe by symmetry that

$$K'(h) = 2 \int_a^b \int_a^b (xy)^{2-\tau} ax f'(ahx) f(ahy) f(xy) dx dy. \quad (4.B.21)$$

At $h = h_c$, we have $ah = b$, and so

$$K'(h_c) = 2 \frac{a}{b} \int_a^b \int_a^b (xy)^{2-\tau} bx f'(bx) f(by) f(xy) dx dy. \quad (4.B.22)$$

Now $uf'(u) = 0$ for $0 \leq u \leq 1$ and $uf'(u) = -f(u)$ for $u \geq 1$. Hence, splitting up the integration range into the four regions as earlier, we see that those over $[a, 1/b] \times [a, 1/b]$ and $[a, 1/b] \times [1/b, b]$ vanish while those over $[1/b, b] \times [1/b, b]$ and $[1/b, b] \times [a, 1/b]$ give rise to the same double integrals as in (4.B.17) and (4.B.19) respectively. This yields the expression in (4.B.13) for $K'(h_c)$.

The evaluation of $J(h_c)$ and $J'(h_c)$ is straightforward from (4.B.9) with $ah = b$ and a splitting of the integration range $[a, b]$ into $[a, 1/b]$ and $[1/b, b]$. This yields (4.B.14), and the proof of Lemma 4.1 is complete. \square

Proof of (4.B.6). We now turn to the limiting behavior of $\sigma'_n(\frac{1}{\tau-1})$. We write

$$0 < \frac{D}{D+E} = \frac{1-b^{2(1-\tau)}}{\frac{\tau-1}{\tau-2}(ab)^{2-\tau} - \frac{1}{\tau-2} - \frac{1}{\tau-1}b^{2(1-\tau)}}, \quad (4.B.23)$$

in which

$$b^{2(1-\tau)} = (n\mathbb{E}[h])^{\tau-3} \rightarrow 0, \quad (4.B.24)$$

$$(ab)^{2-\tau} = (n\mathbb{E}[h])^{\frac{(\tau-2)^2}{\tau-1}} \rightarrow \infty, \quad (4.B.25)$$

as $n \rightarrow \infty$. Hence, $D/(D+E) \rightarrow 0$ as $n \rightarrow \infty$. Furthermore, we write

$$C = \frac{b^{2(\tau-3)}}{(\tau-3)^2} \left(1 - (ab)^{3-\tau}\right)^2, \quad (4.B.26)$$

and

$$A = \frac{b^{2(\tau-3)}}{(\tau-2)^2} (1-F), \quad (4.B.27)$$

where

$$\begin{aligned} F &= b^{-2(\tau-2)} \left[\frac{\tau-2}{\tau-1} \ln(b^2) + \left(\frac{\tau-2}{\tau-1} \right)^2 (1 - b^{2(1-\tau)}) + 1 \right] \\ &= \frac{1}{\tau-1} b^{-2(\tau-2)} \ln(b^{2(\tau-2)}) \left(1 + O(\ln(b)^{-1}) \right). \end{aligned} \quad (4.B.28)$$

Now, using (4.B.20), we have

$$(ab)^{3-\tau} = b^{-2(\tau-2)} = (n\mathbb{E}[h])^{\frac{(\tau-2)(3-\tau)}{\tau-1}} \rightarrow 0 \quad (4.B.29)$$

as $n \rightarrow \infty$. Thus, we get

$$\lim_{n \rightarrow \infty} \frac{A + \frac{3-\tau}{2-\tau}C}{A + \frac{4-\tau}{2-\tau}C} = \frac{\frac{1}{(\tau-2)^2} + \frac{3-\tau}{\tau-2} \frac{1}{(3-\tau)^2}}{\frac{1}{(\tau-2)^2} + \frac{4-\tau}{\tau-2} \frac{1}{(3-\tau)^2}} = 3 - \tau, \quad (4.B.30)$$

and this yields the equality in (4.B.6).

We finally turn to the inequality in (4.B.6) in Proposition 4.1. Obviously, we have

$$\sigma'_n \left(\frac{1}{\tau-1} \right) > -2 \frac{A + \frac{3-\tau}{\tau-2}C}{A + \frac{4-\tau}{\tau-2}C}. \quad (4.B.31)$$

We shall show that

$$\frac{A + \frac{3-\tau}{\tau-2}C}{A + \frac{4-\tau}{\tau-2}C} \leq \frac{A_{\text{as}} + \frac{3-\tau}{\tau-2}C_{\text{as}}}{A_{\text{as}} + \frac{4-\tau}{\tau-2}C_{\text{as}}} = 3 - \tau, \quad (4.B.32)$$

where

$$A_{\text{as}} = \frac{b^{2(\tau-3)}}{(\tau-2)^2}, \quad C_{\text{as}} = \frac{b^{2(\tau-3)}}{(3-\tau)^2}, \quad (4.B.33)$$

the asymptotic form of A and C as $n \rightarrow \infty$ obtained from (4.B.27) and (4.B.26) by deleting F and $(ab)^{3-\tau}$, respectively. The function

$$x \in [0, \infty) \mapsto \frac{1 + \frac{3-\tau}{\tau-2}x}{1 + \frac{4-\tau}{\tau-2}x} \quad (4.B.34)$$

is decreasing in $x \geq 0$, and so it suffices to show that

$$\frac{C_{\text{as}}}{A_{\text{as}}} \leq \frac{C}{A}, \text{ i.e., that } \frac{C_{\text{as}}}{C} \leq \frac{A_{\text{as}}}{A}. \quad (4.B.35)$$

We have from (4.B.26) that

$$\frac{C_{\text{as}}}{C} = \frac{1}{(1 - (ab)^{3-\tau})^2}, \quad (4.B.36)$$

and from (4.B.27) and (4.B.28) that

$$\begin{aligned} \frac{A}{A_{\text{as}}} &= 1 - F \\ &= 1 - b^{-2(\tau-2)} - b^{-2(\tau-2)} \left[\frac{\tau-2}{\tau-1} \ln(b^2) + \left(\frac{\tau-2}{\tau-1} \right)^2 (1 - b^{2(1-\tau)}) \right]. \end{aligned} \quad (4.B.37)$$

Using that $(ab)^{3-\tau} = b^{-2(\tau-2)}$, see (4.B.29), we see that the inequality $C_{\text{as}}/C \leq A_{\text{as}}/A$ in (4.B.35) is equivalent to

$$\begin{aligned} (1 - b^{-2(\tau-2)})^2 &\geq 1 - b^{-2(\tau-2)} - b^{-2(\tau-2)} \\ &\times \left[\frac{\tau-2}{\tau-1} \ln(b^2) + \left(\frac{\tau-2}{\tau-1} \right)^2 (1 - b^{2(1-\tau)}) \right]. \end{aligned} \quad (4.B.38)$$

Using that $(1-u)^2 - (1-u) = -u(1-u)$ and dividing through by $u = b^{-2(\tau-2)}$, we see that (4.B.38) is equivalent to

$$\frac{\tau-2}{\tau-1} \ln(b^2) + \left(\frac{\tau-2}{\tau-1} \right)^2 (1 - b^{2(1-\tau)}) \geq 1 - b^{-2(\tau-2)}. \quad (4.B.39)$$

With $y = \ln(b^2) \geq 0$, we write (4.B.39) as

$$K(y) := \left(\frac{\tau-2}{\tau-1} \right)^2 (1 - e^{(1-\tau)y}) + \frac{\tau-2}{\tau-1} y - (1 - e^{(2-\tau)y}) \geq 0. \quad (4.B.40)$$

Taylor development of $K(y)$ at $y = 0$ yields

$$K(y) = 0 \cdot y^0 + 0 \cdot y^1 + 0 \cdot y^2 + \frac{1}{6}(\tau-2)^2 y^3 + \dots \quad (4.B.41)$$

Furthermore,

$$K''(y) = (\tau-2)^2 e^{(1-\tau)y} (e^y - 1) > 0, \quad y > 0. \quad (4.B.42)$$

Therefore, $K(0) = K'(0) = 0$, while $K''(y) > 0$ for $y > 0$. This gives $K(y) > 0$ when $y > 0$, as required. \square

Similar to Proposition 4.1, we can derive the following result for $\sigma'_n(\frac{1}{2})$:

Proposition 4.2.

$$\sigma'_n(\frac{1}{2}) = -2 \left(\frac{G + H}{(1 + (\frac{\tau-1}{3-\tau})^2)G + 2H} - \frac{I}{I + J} \right), \quad (4.B.43)$$

where

$$G = \left(\frac{1 - b^{1-\tau}}{\tau - 1} \right)^2, \quad (4.B.44)$$

$$I = \frac{1 - b^{1-\tau}}{\tau - 1}, \quad (4.B.45)$$

$$J = \frac{b^{(\tau-2)(\tau-1)/(3-\tau)} - 1}{\tau - 2}, \quad (4.B.46)$$

$$H = \frac{1 - 1/b - b^{1-\tau}(1 - b^{2-\tau})}{(\tau - 2)(3 - \tau)} - \frac{1 - b^{1-\tau}}{(\tau - 1)(\tau - 2)}. \quad (4.B.47)$$

Furthermore, for all n ,

$$\sigma'_n(\frac{1}{2}) > \lim_{M \rightarrow \infty} \sigma'_M(\frac{1}{2}) = -1 + \frac{2(\tau - 2)}{3 - (\tau - 2)^2}. \quad (4.B.48)$$

4.C From weights to degrees

In this chapter, we focus on computing $\bar{c}(h)$, the local clustering coefficient of a randomly chosen vertex with weight h . However, when studying local clustering in real-world data sets, we can only observe $c(k)$, the local clustering coefficient of a vertex of degree k . In this appendix, we show that for the rank-1 inhomogeneous random graph, the difference between these two methods of computing the clustering coefficient is small and asymptotically negligible. We consider

$$\bar{c}(h) = \frac{\int_1^{h_c} \int_1^{h_c} (h'h'')^{2-\tau} p(h, h') p(h, h'') p(h', h'') dh' dh''}{\left(\int_1^{h_c} x^{1-\tau} p(h, h') dh' \right)^2}. \quad (4.C.1)$$

We define $c(k)$ as the average clustering coefficient over all vertices of degree k . By [201], the probability that a vertex with weight h has degree k equals

$$g(k | h) = \frac{e^{-h} h^k}{k!}. \quad (4.C.2)$$

Then, by [201],

$$c(k) = \begin{cases} \frac{1}{P(k)} \int_1^{h_c} \rho(h) \bar{c}(h) g(k | h) dh, & k \geq 2, \\ 0, & k < 2, \end{cases} \quad (4.C.3)$$

where $c(k) = 0$ for $k < 2$ because a vertex with degree less than 2 cannot be part of a triangle. Here

$$P(k) = \int_1^{h_c} g(k | h) \rho(h) dh \quad (4.C.4)$$

is the probability that a randomly chosen vertex has degree k .

First we consider the case $h > n^{\frac{\tau-2}{\tau-1}}$. The Chernoff bound gives for the tails of the Poisson distribution that

$$\mathbb{P}(\text{Poi}(\lambda) > x) \leq e^{-\lambda} \left(\frac{e\lambda}{x} \right)^x, \quad x > \lambda, \quad (4.C.5)$$

$$\mathbb{P}(\text{Poi}(\lambda) < x) \leq e^{-\lambda} \left(\frac{e\lambda}{x} \right)^x, \quad x < \lambda. \quad (4.C.6)$$

Let $k(h)$ be the degree of a vertex with weight h . Then, for any $M > 1$

$$\sum_{k=Mh}^{\infty} g(k | h) \leq \left(\frac{e^{M-1}}{M^M} \right)^h, \quad (4.C.7)$$

and for any $\delta \in (0, 1)$,

$$\sum_{k=1}^{\delta h} g(k | h) \leq \left(\frac{e^{\delta-1}}{\delta^\delta} \right)^h. \quad (4.C.8)$$

Because $e^{x-1}/x^x < 1$ for $x \neq 1$, (4.C.7) and (4.C.8) tend to zero as $h \rightarrow \infty$. Therefore, for h large,

$$k(h) = h(1 + o(1)) \quad (4.C.9)$$

with high probability. Therefore, when k is large, $c(k) \approx \bar{c}(k)$.

On the other hand, for $h \ll h_s^2/h_c$,

$$\sum_{h_s^2/h_c}^{\infty} g(k | h) \leq e^{-h} \left(\frac{eh}{h_s^2/h_c} \right)^{h_s^2/h_c}, \quad (4.C.10)$$

which is small by the assumption on h . Thus,

$$P(k) \approx \int_1^{h_s^2/h_c} g(k | h) \rho(h) dh. \quad (4.C.11)$$

Furthermore, $\bar{c}(h) = \bar{c}(0)$ in this regime of h . This results in

$$\bar{c}(k) \approx \frac{\bar{c}(0) \int_1^{h_s^2/h_c} \rho(h) g(k | h) dh}{\int_1^{h_s^2/h_c} \rho(h) g(k | h) dh} = \bar{c}(0). \quad (4.C.12)$$

Therefore, $c(h) \approx \bar{c}(h)$ also when h is small.

Figure 4.9 shows that indeed the difference between $c(k)$ and $\bar{c}(k)$ is small. When τ approaches 2, the difference becomes larger. We see that for small values of k , $c(k)$ and $\bar{c}(k)$ are not very close. This is because (4.C.1) does not take into account that a vertex with weight h may have less than 2 neighbors, so that its local clustering is zero. In Chapter 3 we have shown how to adjust (4.A.6) to account for this.

4.D Degree distributions

Figure 4.10 shows the degree distributions of all ten networks of Table 4.1.

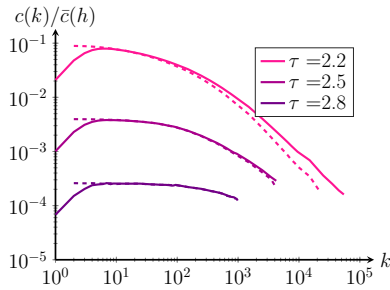


Figure 4.9: $c(k)$ (dashed) and $\bar{c}(h)$ (line) for $n = 10^5$, averaged over 10^4 realizations.

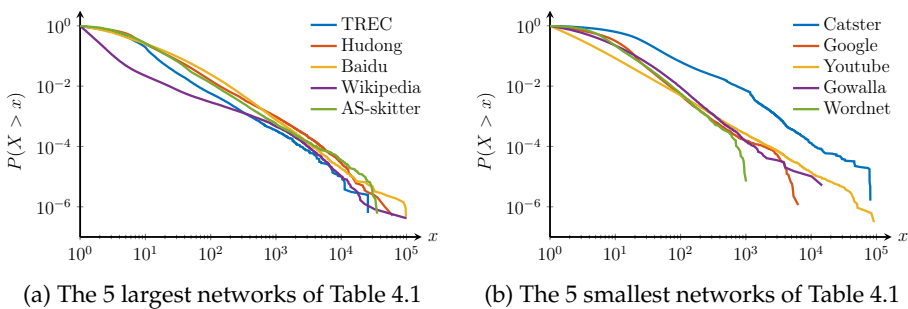


Figure 4.10: The probability that the degree of a vertex exceeds x in the networks of Table 4.1.

5 Local clustering in erased configuration models and uniform random graphs

Based on:

Triadic closure in simple scale-free networks with unbounded degree fluctuations
R. van der Hofstad, J.S.H. van Leeuwen, C. Stegehuis
Journal of Statistical Physics (2018)

and

Triangle counts in power-law uniform random graphs
J. Gao, R. van der Hofstad, A. Southwell, C. Stegehuis
arXiv:1812.04289

We again study the local clustering coefficient $c(k)$, but now for the erased configuration model and the uniform random graph, where the degrees follow a power-law distribution with exponent $\tau \in (2, 3)$. We show that $c(k)$ falls off with k as well as with the graph size n and eventually for $k = \Omega(\sqrt{n})$ settles on a power law $c(k) \sim n^{5-2\tau}k^{-2(3-\tau)}$. We show that, apart from constants, the results for $c(k)$ are similar to those in the rank-1 inhomogeneous random graph of Chapter 4. We show that only triangles consisting of triplets with uniquely specified degrees contribute to the local clustering coefficient. Furthermore, apart from constants, $c(k)$ behaves the same in the uniform random graph and the erased configuration model. Interestingly, despite the fact that the erased configuration model erases edges and the uniform random graph does not, $c(k)$ is higher in the erased configuration model than in the uniform random graph.

5.1 Introduction

Chapter 4 showed that in many real-world networks, as well as in the popular rank-1 inhomogeneous random graph model, the local clustering coefficient $c(k)$ of vertices of degree k decays when k becomes large. In this chapter we again analyze $c(k)$ for networks with a power-law degree distribution with degree exponent $\tau \in (2, 3)$. To analyze $c(k)$, we now consider the configuration model in the large-network limit, and count the number of triangles where at least one of the vertices has degree k . When the degree exponent $\tau > 3$, the total number of triangles in the configuration model converges to a Poisson random variable [106, Chapter 7]. When $\tau \in (2, 3)$, the configuration model contains many self-loops and multiple edges [106]. This creates multiple ways of counting the number of triangles, as we will show below. In this chapter, we count the number of triangles from a *vertex* perspective, which is the

same as counting the number of triangles in the erased configuration model, where all self-loops have been removed and multiple edges have been merged.

Another option to avoid multiple triangles between one set of three vertices, is to analyze uniform random graphs instead: graphs sampled uniformly from the ensemble of all simple graphs with a given degree sequence. Such graphs can be generated by switching algorithms, that start with an initial graph, and rewire some of the edges at each time step [11, 97, 151]. However, the number of necessary steps for these algorithms to result in a uniform sample is unknown for $\tau \in (2, 3)$. We analyze $c(k)$ for uniform random graphs, and show that $c(k)$ behaves similarly as in the erased configuration model, apart from constants. Since erased configuration models are easier to generate than uniform random graphs, this also justifies the approximation of uniform random graphs by erasing edges in the configuration model for $\tau \in (2, 3)$ if one is only interested in the scaling of the clustering coefficient. Taking the precise asymptotics of $c(k)$ into account, our results show that $c(k)$ is higher for erased configuration models than for uniform random graphs.

We show that the local clustering coefficient remains a constant times $n^{2-\tau} \log(n)$ as long as $k = o(\sqrt{n})$. After that, $c(k)$ starts to decay as $c(k) \sim k^{-\gamma} n^{5-2\tau}$. We show that this exponent γ depends on τ and can be larger than one. In particular, when the power-law degree exponent τ is close to two, the exponent γ approaches two, a considerable difference with other random graph models that predict $c(k) \sim k^{-1}$ [75, 132, 187]. Related to this result on the $c(k)$ fall-off, we also show that for every vertex with fixed degree k only pairs of vertices with specific degrees contribute to the triangle count and hence local clustering.

The chapter is structured as follows. Section 5.2 describes the multiple ways of triangle counting in the configuration model. We present our main results in Section 5.3, including Theorems 5.1 and 5.5 that describe the three ranges of $c(k)$ in uniform random graphs. The next sections prove our main results for the erased configuration model, and in particular focus on establishing Propositions 5.1 and 5.2 that are crucial for the proof of Theorem 5.1. Finally, Section 5.7 shows how to adjust these proofs for uniform random graphs.

5.2 Basic notions

In this chapter, we focus on the configuration model, defined in Section 1.1.1 and the uniform random graph, defined in Section 1.1.2. We take the degree sequence as an i.i.d. sample of a random variable D such that

$$\mathbb{P}(D = k) = Ck^{-\tau}, \quad (5.2.1)$$

when $k \rightarrow \infty$, where $\tau \in (2, 3)$ so that $\mathbb{E}[D^2] = \infty$. When this sample constructs a sequence such that the sum of the variables is odd, we add an extra half-edge to the last vertex. This does not affect our computations. In this setting, $D_{\max} = O_{\mathbb{P}}(n^{1/(\tau-1)})$, where $D_{\max} = \max_{v \in [n]} D_v$ denotes the maximal degree of the degree sequence.

Counting triangles. Let $G = (V, E)$ denote a configuration model with vertex set $V = [n] := \{1, \dots, n\}$ and edge set E . We are interested in the number of triangles

in G . There are two ways to count triangles in the configuration model. The first approach is from an *edge perspective*, as illustrated in Figure 5.1. This approach counts the number of triples of edges that together create a triangle. This approach may count multiple triangles between one fixed triple of vertices. Let X_{ij} denote the number of edges between vertex i and j . Then, from an edge perspective, the number of triangles in the configuration model is

$$\sum_{1 \leq i < j < k \leq n} X_{ij} X_{jk} X_{ik}. \quad (5.2.2)$$

A different approach is to count the number of triangles from a *vertex perspective*. This approach counts the number of triples of vertices that are connected. Counting the number of triangles in this way results in

$$\sum_{1 \leq i < j < k \leq n} \mathbb{1}_{\{X_{ij} \geq 1\}} \mathbb{1}_{\{X_{jk} \geq 1\}} \mathbb{1}_{\{X_{ik} \geq 1\}}. \quad (5.2.3)$$

When the configuration model results in a simple graph, these two approaches give the same result. When the configuration model results in a multigraph, these two approaches may give substantially different numbers of triangles. In particular, when the degree distribution follows a power-law with $\tau \in (2, 3)$, the number of triangles is dominated by the number of triangles between the vertices of the highest degrees, even though only few such vertices are present in the graph [162]. When the exponent τ of the degree distribution approaches 2, then the number of triangles between the vertices of the highest degrees will be as high as $\Theta(n^3)$, which is much higher than the number of triangles we would expect in any real-world network of that size. When we count triangles from a vertex perspective, we count only one triangle between these three vertices. Thus, the number of triangles from a vertex perspective will be significantly lower. In this chapter, we focus on the vertex-based approach for counting triangles. This approach is the same as counting triangles in the *erased configuration model* defined in Section 1.1.3, where all multiple edges have been merged, and the self-loops have been removed.

Let Δ_k denote the number of triangles attached to vertices of degree k . When a triangle consists of two vertices of degree k , it is counted twice in Δ_k . Let N_k denote the number of vertices of degree k . Then, the clustering coefficient of vertices with degree k equals

$$c(k) = \frac{1}{N_k} \frac{2\Delta_k}{k(k-1)}. \quad (5.2.4)$$

When we count Δ_k from the vertex perspective, this clustering coefficient can be interpreted as the probability that two random connections of a vertex with degree k are connected. This version of $c(k)$ is the local clustering coefficient of the erased configuration model. It is possible that no vertex of degree k is present in the graph. We therefore analyze

$$c_\varepsilon(k) = \frac{1}{|M_\varepsilon(k)|} \frac{2\Delta_{M_\varepsilon(k)}}{k(k-1)}. \quad (5.2.5)$$

where $M_\varepsilon(k) = \{i \in [n] : D_i \in [k(1-\varepsilon), k(1+\varepsilon)]\}$, and $\Delta_{M_\varepsilon(k)}$ the number of triangles attached to vertices in $M_\varepsilon(k)$. When no vertex in $M_\varepsilon(k)$ exists, we set



Figure 5.1: From the edge perspective in the configuration model, these are two triangles. From the vertex perspective, there is only one triangle.

$c_\varepsilon(k) = 0$. We will show that in the models we analyze, $M_\varepsilon(k)$ is non-empty with high probability, so that $c_\varepsilon(k)$ is well defined.

5.3 Main results

We first present our results for the erased configuration model, and then we present our results for uniform random graphs.

5.3.1 Erased configuration model

The next theorem presents our main result on the behavior of the local clustering coefficient in the erased configuration model:

Theorem 5.1. *Let G be an erased configuration model, where the degrees are an i.i.d. sample from a power-law distribution with exponent $\tau \in (2, 3)$ as in (5.2.1) with $\tau \in (2, 3)$. Take ε_n such that $\lim_{n \rightarrow \infty} \varepsilon_n = 0$ and $\lim_{n \rightarrow \infty} nk^{-(\tau-1)}\varepsilon_n = \infty$. Define $A = -\Gamma(2 - \tau) > 0$ for $\tau \in (2, 3)$, let $\mu = \mathbb{E}[D]$ and C be the constant in (5.2.1). Then, as $n \rightarrow \infty$,*

(Range I) for $1 < k = o(n^{(\tau-2)/(\tau-1)})$,

$$\frac{c_{\varepsilon_n}(k)}{n^{2-\tau} \log(n)} \xrightarrow{\mathbb{P}} \frac{3-\tau}{\tau-1} \mu^{-\tau} C^2 A, \quad (5.3.1)$$

(Range II) for $k = \Omega(n^{(\tau-2)/(\tau-1)})$ and $k = o(\sqrt{n})$,

$$\frac{c_{\varepsilon_n}(k)}{n^{2-\tau} \log(n/k^2)} \xrightarrow{\mathbb{P}} \mu^{-\tau} C^2 A, \quad (5.3.2)$$

(Range III) for $k = \Omega(\sqrt{n})$ and $k \ll n^{1/(\tau-1)}$,

$$\frac{c_{\varepsilon_n}(k)}{n^{5-2\tau} k^{2\tau-6}} \xrightarrow{\mathbb{P}} \mu^{3-2\tau} C^2 A^2. \quad (5.3.3)$$

Theorem 5.1 shows three different ranges for the local clustering coefficient, and is illustrated in Figure 5.2. Let us explain why these three ranges occur. Range I contains small-degree vertices with $k = o(n^{(\tau-2)/(\tau-1)})$. In Section 5.4.2 we show that these vertices are hardly involved in self-loops and multiple edges in the configuration

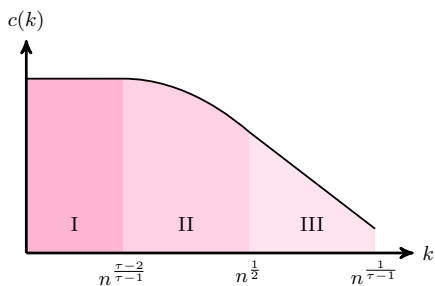


Figure 5.2: The three ranges of $c_{\varepsilon_n}(k)$ defined in Theorem 5.1 on a log-log scale.

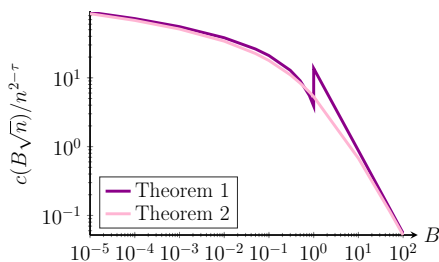


Figure 5.3: The normalized version of $c_{\varepsilon_n}(k)$ for $k = B\sqrt{n}$ obtained from Theorem 5.1 and Theorem 5.2.

model, and hence there is little difference between counting from an edge perspective or from a vertex perspective. It turns out that these vertices barely form triadic closures with hubs, which makes $c_{\varepsilon_n}(k)$ independent of k in Theorem 5.1. Range II contains degrees that are neither small nor large with degrees $k = \Omega(n^{(\tau-2)/(\tau-1)})$ and $k = o(\sqrt{n})$. We can approximate the connection probability between vertices i and j with $1 - e^{-D_i D_j / \mu^n}$, where $\mu = \mathbb{E}[D]$. Therefore, a vertex of degree k connects to vertices of degree at least n/k with positive probability. The vertices in Range II are quite likely to have multiple connections with vertices of degrees at least n/k . Thus, in this degree range, the single-edge constraint of the erased configuration model starts to play a role and causes the slow logarithmic decay of $c_{\varepsilon_n}(k)$ in Theorem 5.1. Range III contains the large-degree vertices with $k = \Omega(\sqrt{n})$. Again we approximate the probability that vertices i and j are connected by $1 - e^{-D_i D_j / \mu^n}$. This shows that vertices in Range III are likely to be connected to one another, possibly through multiple edges. The single-edge constraint on all connections between these core vertices causes the power-law decay of $c_{\varepsilon_n}(k)$ in Theorem 5.1.

Theorem 5.1 shows that the local clustering not only decays in k , it also decays in the graph size n for all values of k . This decay in n is caused by the locally tree-like nature of the configuration model. Chapter 4 showed that in large real-world networks, $c_{\varepsilon_n}(k)$ is typically high for small values of k , which is unlike the behavior in the erased configuration model. The behavior of $c_{\varepsilon_n}(k)$ for more realistic network models is therefore an interesting question for further research. We believe that including small communities to the configuration model such as in Chapter 11 would only change the $k \mapsto c_{\varepsilon_n}(k)$ curve for small values of k with respect to the erased configuration model. Low-degree vertices will then typically be in highly clustered communities and therefore have high local clustering coefficients. Most connections from high-degree vertices will be between different communities, which results in a similar $k \mapsto c_{\varepsilon_n}(k)$ curve for large values of k as in the erased configuration model.

Observe that in Theorem 5.1 the behavior of $c_{\varepsilon_n}(k)$ on the boundary between two different ranges may be different than the behavior inside the ranges. Since $k \mapsto c_{\varepsilon_n}(k)$ is a function on a discrete domain, it is always continuous. However, we can extend the scaling limit of $k \mapsto c_{\varepsilon_n}(k)$ to a continuous domain. Theorem 5.1 then shows that the scaling limit of $k \mapsto c_{\varepsilon_n}(k)$ is a smooth function inside the different

ranges. Furthermore, filling in $k = an^{(\tau-1)/(\tau-2)}$ in Range II of Theorem 5.1 suggests that $k \mapsto c_{\varepsilon_n}(k)$ is also a smooth function on the boundary between Ranges I and II. However, the behavior of $k \mapsto c_{\varepsilon_n}(k)$ on the boundary between Ranges II and III is not clear from Theorem 5.1. We therefore prove the following result in Section 5.6.1:

Theorem 5.2. For $k = B\sqrt{n}$,

$$\frac{c_{\varepsilon_n}(k)}{n^{2-\tau}} \xrightarrow{\mathbb{P}} C^2 \mu^{2-2\tau} B^{-2} \int_0^\infty \int_0^\infty (t_1 t_2)^{-\tau} (1 - e^{-Bt_1})(1 - e^{-Bt_2})(1 - e^{-t_1 t_2 \mu}) dt_1 dt_2. \quad (5.3.4)$$

Figure 5.3 compares $c_{\varepsilon_n}(k)/n^{2-\tau}$ for $k = B\sqrt{n}$ using Theorem 5.2 and Theorem 5.1. The line associated with Theorem 5.1 uses the result for Range II when $B < 1$, and the result for Range III when $B > 1$. We see that there seems to be a discontinuity between these two ranges. Figure 5.3 suggests that the scaling limit of $k \mapsto c_{\varepsilon_n}(k)$ is smooth around $k \approx \sqrt{n}$, because the lines are close for both small and large B -values. Theorem 5.3 shows that indeed the scaling limit of $k \mapsto c_{\varepsilon_n}(k)$ is smooth for k of the order \sqrt{n} :

Theorem 5.3. The scaling limit of $k \mapsto c_{\varepsilon_n}(k)$ is a smooth function.

Most likely configurations. The three different ranges in Theorem 5.1 result from a canonical trade-off caused by the power-law degree distribution. On the one hand, high-degree vertices participate in many triangles. In Section 5.5.1 we show that the probability that a triangle is present between vertices with degrees k, D_u and D_v can be approximated by

$$\left(1 - e^{-kD_u/\mu n}\right) \left(1 - e^{-kD_v/\mu n}\right) \left(1 - e^{-D_u D_v/\mu n}\right). \quad (5.3.5)$$

The probability of this triangle thus increases with D_u and D_v . On the other hand, in power-law distribution high degrees are rare. This creates a trade-off between the occurrence of triangles between $\{k, D_u, D_v\}$ -triplets and the number of them. Surely, large degrees D_u and D_v make a triangle more likely, but larger degrees are less likely to occur. Since (5.3.5) increases only slowly in D_u and D_v as soon as $D_u, D_v = \Omega(\mu n/k)$ or when $D_u D_v = \Omega(\mu n)$, intuitively, triangles with $D_u, D_v = \Omega(\mu n/k)$ or with $D_u D_v = \Omega(\mu n)$ only marginally increase the number of triangles. In fact, we will show that most triangles with a vertex of degree k contain two other vertices of very specific degrees, those degrees that are aligned with the trade-off. The typical degrees of D_u and D_v in a triangle with a vertex of degree k are given by $D_u, D_v \approx \mu n/k$ or by $D_u D_v \approx \mu n$.

Let us now formalize this reasoning. Introduce

$$W_n^k(\delta) = \begin{cases} (u, v) : D_u D_v \in [\delta, 1/\delta] \mu n & \text{for } k \text{ in Range I,} \\ (u, v) : D_u D_v \in [\delta, 1/\delta] \mu n, D_u, D_v < \mu n/(k\delta) & \text{for } k \text{ in Range II,} \\ (u, v) : D_u, D_v \in [\delta, 1/\delta] \mu n/k & \text{for } k \text{ in Range III,} \end{cases} \quad (5.3.6)$$

where the ranges are as in Theorem 5.1. Denote the number of triangles between one vertex of degree k and two other vertices i, j with $(i, j) \in W_n^k(\delta)$ by $\Delta_k(W_n^k(\delta))$. The

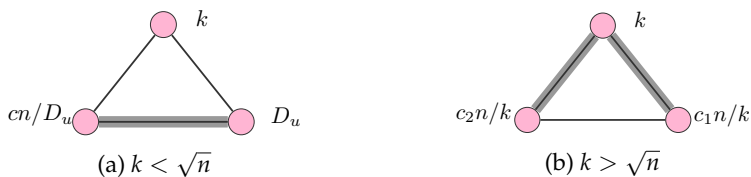


Figure 5.4: The major contributions in the different ranges for k . The highlighted edges are present with asymptotically positive probability.

next theorem shows that these types of triangles dominate all other triangles where one vertex has degree k and formalizes the energy minimization argument for the rank-1 inhomogeneous random graph of Section 4.4:

Theorem 5.4. *Let G be an erased configuration model where the degrees are an i.i.d. sample from a power-law distribution with exponent $\tau \in (2, 3)$. Then, for $\delta_n \rightarrow 0$ sufficiently slowly,*

$$\frac{\Delta_k(W_n^k(\delta_n))}{\Delta_k} \xrightarrow{\mathbb{P}} 1. \quad (5.3.7)$$

For example, when $k = \Omega(\sqrt{n})$, $\Delta_k(W_n^k(\delta_n))$ denotes all triangles between a vertex of degree k and two other vertices with degrees in $[\delta_n, 1/\delta_n]n/k$. Theorem 5.4 then shows that the number of these triangles dominates the number of all other types of triangles where one vertex has degree k . This holds when $\delta_n \rightarrow 0$, so that the degrees of the other two vertices cover the entire $\Theta(n/k)$ range. The convergence of $\delta_n \rightarrow 0$ should be sufficiently slowly, e.g., $\delta_n = 1/\log(n)$, for several combined error terms of δ_n and n to go to zero.

Figure 5.4 illustrates the typical triangles containing a vertex of degree k as given by Theorem 5.4. When k is small (k in Range I or II), a typical triangle containing a vertex of degree k is a triangle with vertices u and v such that $D_u D_v = \Theta(n)$ as shown in in Figure 5.4a. Then, the probability that an edge between u and v exists is asymptotically positive and non-trivial. Since k is small, the probability that an edge exists between a vertex of degree k and u or v is small. On the other hand, when k is larger (in Range III), a typical triangle containing a vertex of degree k is with vertices u and v such that $D_u = \Theta(n/k)$ and $D_v = \Theta(n/k)$. Then, the probability that an edge exists between k and D_u or k and D_v is asymptotically positive whereas the probability that an edge exists between vertices u and v vanishes. Figure 5.4b shows this typical triangle.

Figure 5.5 shows the typical size of the degrees of other vertices in a triangle with a vertex of degree $k = n^\beta$. When $\beta < (\tau - 2)/(\tau - 1)$ (so that k is in Range I), the typical other degrees are independent of the exact value of k . This shows why $c_{\varepsilon_n}(k)$ is independent of k in Range I in Theorem 5.1. When $(\tau - 2)/(\tau - 1) < \beta < \frac{1}{2}$, the range of possible degrees for vertices u and v decreases when k gets larger. Still, the range of possible degrees for D_u and D_v is quite wide. This explains the mild dependence of $c_{\varepsilon_n}(k)$ on k in Theorem 5.1 in Range II. When $\beta > \frac{1}{2}$, k is in Range III. Then the typical values of D_u and D_v are considerably different from those in the previous regime. The values that D_u and D_v can take depend heavily on the value of k . This explains the dependence of $c_{\varepsilon_n}(k)$ on k in Range III.

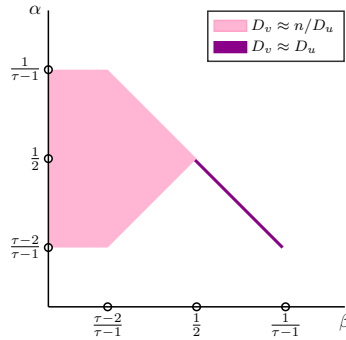


Figure 5.5: Visualization of the contributing degrees when $k = n^\beta$ and $D_u = n^\alpha$. The colored area shows the values of α that contribute to $c(n^\beta)$.

Global and local clustering. The global clustering coefficient divides the total number of triangles by the total number of pairs of neighbors of all vertices. In Chapter 7 we show that the total number of triangles in the configuration model from a vertex perspective is determined by vertices of degree proportional to \sqrt{n} . Thus, only triangles between vertices on the border between Ranges II and III contribute to the global clustering coefficient. The local clustering coefficient counts all triangles where one vertex has degree k and provides a more complete picture of clustering from a vertex perspective, since it covers more types of triangles.

Rank-1 inhomogeneous random graph. Our results for clustering in the erased configuration model are very similar to the results on local clustering in rank-1 inhomogeneous random graphs of Chapter 4. Remember that in the inhomogeneous random graph, every vertex is equipped with a weight h_i , where the weights are sampled from a power-law distribution. Then, vertices i and j are connected with probability $\min(h_i h_j / (\mu n), 1)$ [35, 61]. In the erased configuration model, we use that the probability that a vertex with degree D_i is connected to a vertex with degree D_j can be approximated by

$$1 - e^{-D_i D_j / \mu n}, \quad (5.3.8)$$

which behaves similarly as $\min(D_i D_j / (\mu n), 1)$. Thus, the connection probabilities in the erased configuration model can be interpreted as the connection probabilities in the rank-1 inhomogeneous random graph, where the sampled degrees can be interpreted as the weights. The major difference is that connections in the rank-1 inhomogeneous random graph are *independent* once the weights are sampled, whereas connections in the erased configuration model are *correlated* once the degrees are sampled. Indeed, in the erased configuration model we know that a vertex with degree D_i has at most D_i other vertices as a neighbor, so that the connections from vertex i to other vertices are correlated. Still, our results show that these correlations are small enough for the results for local clustering to be similar as in the rank-1 inhomogeneous random graph.

5.3.2 Uniform random graphs

Since all degrees in uniform random graphs are sampled independently, we may condition on a vertex of degree k being present, and analyze $c(k)$ instead of $c_\varepsilon(k)$. We now present our result on $c(k)$ in the uniform random graph:

Theorem 5.5 (Local clustering.). *Let G be a uniform random graph, where the degrees are an i.i.d. sample from a power-law distribution with exponent $\tau \in (2, 3)$ as in (5.2.1). Define $A_{\text{URG}} = \pi / \sin(\pi\tau) > 0$ for $\tau \in (2, 3)$, let $\mu = \mathbb{E}[D]$ and C be the constant in (5.2.1). Then, as $n \rightarrow \infty$,*

(Range I) for $1 \ll k = o(n^{(\tau-2)/(\tau-1)})$, conditionally on $N_k \geq 1$,

$$\frac{c(k)}{n^{2-\tau} \log(n)} \xrightarrow{\mathbb{P}} \frac{3-\tau}{\tau-1} \mu^{-\tau} C^2 A_{\text{URG}}, \quad (5.3.9)$$

(Range II) for $k = \Omega(n^{(\tau-2)/(\tau-1)})$ and $k = o(\sqrt{n})$, conditionally on $N_k \geq 1$,

$$\frac{c(k)}{n^{2-\tau} \log(n/k^2)} \xrightarrow{\mathbb{P}} \mu^{-\tau} C^2 A_{\text{URG}}, \quad (5.3.10)$$

(Range III) for $k = \Omega(\sqrt{n})$ and $k \ll n^{1/(\tau-1)}$, conditionally on $N_k \geq 1$,

$$\frac{c(k)}{n^{5-2\tau} k^{2\tau-6}} \xrightarrow{\mathbb{P}} \mu^{3-2\tau} C^2 A_{\text{URG}}^2. \quad (5.3.11)$$

Comparison with the erased configuration model. Theorems 5.1 and 5.5 show that $c(k)$ behaves very similarly in the erased configuration model and the uniform random graph. In fact, the only difference is that the constant $A = -\Gamma(2-\tau)$ in the erased configuration model, is replaced by $A_{\text{URG}} = \pi / \sin(\pi\tau)$. Figure 5.6 shows that $A > A_{\text{URG}}$ for $\tau \in (2, 3)$. Thus, the local clustering coefficient of the erased configuration model is higher than the local clustering coefficient of a uniform random graph of the same degree sequence. Interestingly, this implies that the erased configuration model contains *more* triangles with a degree k vertex than a uniform random graph on the same degree sequence, even though edges are removed in the erased configuration model. For the total triangle count, this was also empirically observed in [15]. One possible explanation for this phenomenon is that most hubs in uniform random graphs are forced to connect to many low-degree vertices to still satisfy the simplicity constraint. These low-degree vertices barely participate in triangles. In the erased configuration model, it may be more likely for higher-degree vertices to connect to one another, creating more triangles.

5.3.3 Overview of the proof for the erased configuration model

We now show the outline of the proof of Theorem 5.1 for the erased configuration model. The proof for $c(k)$ in the uniform random graph has a similar structure, and is given later in Section 5.7. To prove Theorem 5.1, we show that there is a major contributing regime for $c_{\varepsilon_n}(k)$, which characterizes the degrees of the other two vertices in a typical triangle with a vertex of degree k . We write this major

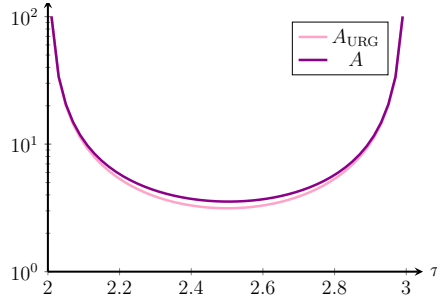


Figure 5.6: The constant A of Theorem 5.1 is bigger than the constant A_{URG} of Theorem 5.5.

contributing regime as $W_n^k(\delta)$ defined in (5.3.6). The number of triangles adjacent to a vertex of degree k is dominated by triangles between the vertex of degree k and other vertices with degrees in a specific regime, depending on k . All three ranges of k have a different spectrum of degrees that contribute to the number of triangles. We write

$$c_{\varepsilon_n}(k) = c_{\varepsilon_n}(W_n^k(\delta)) + c_{\varepsilon_n}(\bar{W}_n^k(\delta)), \quad (5.3.12)$$

where $c_{\varepsilon_n}(W_n^k(\delta))$ denotes the contribution to $c_{\varepsilon_n}(k)$ from triangles where the other two vertices $(u, v) \in W_n^k(\delta)$ and $c_{\varepsilon_n}(\bar{W}_n^k(\delta))$ denotes the contribution to $c_{\varepsilon_n}(k)$ from triangles where the other two vertices $(u, v) \notin W_n^k(\delta)$. Furthermore, we write the order of magnitude of the value of $c_{\varepsilon_n}(k)$ as $f(k, n)$. Theorem 5.1 states that this order should be

$$f(k, n) = \begin{cases} n^{2-\tau} \log(n) & \text{for } k = o(n^{(\tau-2)/(\tau-1)}), \\ n^{2-\tau} \log(n/k^2) & \text{for } k = \Omega(n^{(\tau-2)/(\tau-1)}), k = o(\sqrt{n}), \\ n^{5-2\tau} k^{2\tau-6} & \text{for } k = \Omega(\sqrt{n}). \end{cases} \quad (5.3.13)$$

We let \mathbb{P}_n denote the conditional probability given $(D_i)_{i \in [n]}$, and \mathbb{E}_n the corresponding expectation. The proof of Theorem 5.1 is largely built on the following two propositions:

Proposition 5.1 (Main contribution).

$$\frac{c_{\varepsilon_n}(W_n^k(\delta))}{f(n, k)} \xrightarrow{\mathbb{P}} \begin{cases} C^2 \int_{\delta}^{1/\delta} t^{1-\tau} (1 - e^{-t}) dt & k = o(\sqrt{n}), \\ C^2 \left(\int_{\delta}^{1/\delta} t^{1-\tau} (1 - e^{-t}) dt \right)^2 & k = \Omega(\sqrt{n}). \end{cases} \quad (5.3.14)$$

Proposition 5.2 (Minor contributions). *There exists $\kappa > 0$ such that for all ranges*

$$\limsup_{n \rightarrow \infty} \frac{\mathbb{E}_n \left[c_{\varepsilon_n}(\bar{W}_n^k(\delta)) \right]}{f(n, k)} \xrightarrow{\mathbb{P}} O_{\mathbb{P}}(\delta^{\kappa}). \quad (5.3.15)$$

We now show how these propositions prove Theorem 5.1. Applying Proposition 5.2 together with the Markov inequality yields

$$\mathbb{P} \left(c_{\varepsilon_n}(\bar{W}_n^k(\delta)) > K f(n, k) \delta^{\kappa} \right) = O(K^{-1}). \quad (5.3.16)$$

Therefore,

$$c_{\varepsilon_n}(k) = c_{\varepsilon_n}(W_n^k(\delta)) + O_{\mathbb{P}}(f(k, n)\delta^\kappa). \quad (5.3.17)$$

Replacing δ by δ_n , and letting $\delta_n \rightarrow 0$ slowly enough for all combined error terms of δ_n and $o(f(n, k))$ in the expectation in (5.3.15) to converge to 0 then already proves Theorem 5.4. To prove Theorems 5.1 and 5.2 we use Proposition 5.1, which shows that

$$\frac{c_{\varepsilon_n}(k)}{f(k, n)} \xrightarrow{\mathbb{P}} \begin{cases} C^2 \int_{\delta}^{1/\delta} t^{1-\tau}(1 - e^{-t}) dt + O(\delta^\kappa) & k = o(\sqrt{n}), \\ C^2 \left(\int_{\delta}^{1/\delta} t^{1-\tau}(1 - e^{-t}) dt \right)^2 + O(\delta^\kappa) & k = \Omega(\sqrt{n}). \end{cases} \quad (5.3.18)$$

We take the limit of $\delta \rightarrow 0$ and use that

$$\begin{aligned} \int_0^\infty x^{1-\tau}(1 - e^{-x}) dx &= \int_0^\infty \int_0^x x^{1-\tau} e^{-y} dy dx = \int_0^\infty \int_y^\infty x^{1-\tau} e^{-y} dx dy \\ &= -\frac{1}{2-\tau} \int_0^\infty y^{2-\tau} e^{-y} dy = -\frac{\Gamma(3-\tau)}{2-\tau} = -\Gamma(2-\tau) =: A, \end{aligned} \quad (5.3.19)$$

which proves Theorem 5.1. \square

The rest of the chapter is devoted to proving Propositions 5.1 and 5.2. We prove Proposition 5.1 using a second moment method. We compute the expected value of $c_{\varepsilon_n}(k)$ conditioned on the degrees as

$$\mathbb{E}_n [c_{\varepsilon_n}(k)] = \frac{2 \sum_{w \in M_{\varepsilon_n}(k)} \mathbb{E}_n [\Delta(w)]}{|M_{\varepsilon_n}(k)| k(k-1)}, \quad (5.3.20)$$

where $\Delta(w)$ denotes the number of triangles containing vertex w . Let X_{ij} denote the number of edges between vertex i and j in the configuration model, and \hat{X}_{ij} the number of edges between i and j in the corresponding erased configuration model, so that $\hat{X}_{ij} \in \{0, 1\}$. Now,

$$\mathbb{E}_n [\Delta(w) \mid D_w^{(\text{er})} = k] = \frac{1}{2} \sum_{u, v \neq w} \mathbb{P}_n(\hat{X}_{wu} = \hat{X}_{wv} = \hat{X}_{uv} = 1 \mid D_w^{(\text{er})} = k). \quad (5.3.21)$$

Thus, to find the expected number of triangles, we need to compute the probability that a triangle between vertices u, v and w exists, which we will do in Section 5.5.1. After that, we show that this expectation converges to a constant when taking the randomness of the degrees into account, and that the variance conditioned on the degrees is small in Section 5.5.3. Then, we prove Proposition 5.2 in Section 5.6 using a first moment method. We start in Section 5.4 to state some preliminaries.

5.4 Preliminaries

We now introduce some lemmas that we will use frequently while proving Propositions 5.1 and 5.2. Let D_u denote a uniformly chosen vertex from the degree sequence $(D_i)_{i \in [n]}$ and let $L_n = \sum_{i \in [n]} D_i$ denote the sum of the degrees.

5.4.1 Conditioning on the degrees

In the proof of Proposition 5.1 we first condition on the degree sequence. We compute the clustering coefficient conditional on the degree sequence, and after that we show that this converges to the correct value when taking the random degrees into account. We will use the following lemma several times:

Lemma 5.1. *Let $(G_n)_{n \geq 1}$ be a sequence of erased configuration models on n vertices where the degrees are an i.i.d. sample from a random variable D . Then,*

$$\mathbb{P}_n(\mathcal{D}_u \in [a, b]) = O_{\mathbb{P}}(\mathbb{P}(D \in [a, b])) \quad (5.4.1)$$

$$\mathbb{E}_n[f(\mathcal{D}_u)] = O_{\mathbb{P}}(\mathbb{E}[f(D)]). \quad (5.4.2)$$

Proof. By using the Markov inequality, we obtain for $M > 0$

$$\mathbb{P}(\mathbb{P}_n(\mathcal{D}_u \in [a, b]) \geq M\mathbb{P}(D \in [a, b])) \leq \frac{\mathbb{E}[\mathbb{P}_n(\mathcal{D}_u \in [a, b])]}{M\mathbb{P}(D \in [a, b])} = \frac{1}{M}, \quad (5.4.3)$$

and the second claim can be proven in a similar way. \square

In the proof of Theorem 5.1 we often estimate moments of D , conditional on the degrees. The following lemma shows how to bound these moments, and is a direct consequence of the Stable Law Central Limit Theorem:

Lemma 5.2. *Let \mathcal{D}_u be a uniformly chosen vertex from the degree sequence, where the degrees are an i.i.d. sample from a power-law distribution with exponent $\tau \in (2, 3)$. Then, for $\alpha > \tau - 1$,*

$$\mathbb{E}_n[\mathcal{D}_u^\alpha] = O_{\mathbb{P}}\left(n^{\alpha/(\tau-1)-1}\right). \quad (5.4.4)$$

Proof. We have

$$\mathbb{E}_n[\mathcal{D}_u^\alpha] = \frac{1}{n} \sum_{i=1}^n D_i^\alpha. \quad (5.4.5)$$

Since the D_i are an i.i.d. sample from a power-law distribution with exponent τ ,

$$\mathbb{P}(D_i^\alpha > t) = \mathbb{P}(D_i > t^{1/\alpha}) = Ct^{-\frac{\tau-1}{\alpha}}, \quad (5.4.6)$$

so that D_i^α are distributed as i.i.d. samples from a power-law with exponent $(\tau - 1)/\alpha + 1 < 2$. Then, by the Stable law Central Limit Theorem (see for example [217, Theorem 4.5.1]),

$$\sum_{i=1}^n D_i^\alpha = O_{\mathbb{P}}\left(n^{\frac{\alpha}{\tau-1}}\right), \quad (5.4.7)$$

which proves the lemma. \square

We also need to relate L_n and its expected value μn . Define the events

$$J_n = \left\{ |L_n - \mu n| \leq n^{2/\tau} \right\}, \quad \mathcal{A}_n = \left\{ |M_{\varepsilon_n}(k)| \geq 1 \right\} \quad (5.4.8)$$

By [111], $\mathbb{P}(J_n) \rightarrow 1$ as $n \rightarrow \infty$, and by (2.3.2), $\mathbb{P}(\mathcal{A}_n) \rightarrow 1$. When we condition on the degree sequence, we will assume that the events J_n and \mathcal{A}_n take place.

5.4.2 Erased and non-erased degrees

The degree sequence of the erased configuration model may differ from the original degree sequence of the original configuration model. By Lemma 2.1

$$D_i^{(\text{er})} = D_i(1 - o_{\mathbb{P}}(1)). \quad (5.4.9)$$

Thus, in many proofs, we will exchange D_i and $D_i^{(\text{er})}$ when needed.

5.5 Second moment method on main contribution $W_n^k(\delta)$

We now focus on the triangles that give the main contribution. First, we condition on the degree sequence and compute the expected number of triangles in the main contributing regime. Then, we show that this expectation converges to a constant when taking the i.i.d. degrees into account. After that, we show that the variance of the number of triangles in the main contributing regime is small, and we prove Proposition 5.1.

5.5.1 Conditional expectation inside $W_n^k(\delta)$

In this section, we compute the expected number of triangles in the major contributing ranges of 5.3.6 when we condition on the degree sequence. We define

$$g_n(D_u, D_v, D_w) := (1 - e^{-D_u D_v / L_n})(1 - e^{-D_u D_w / L_n})(1 - e^{-D_v D_w / L_n}). \quad (5.5.1)$$

Then, the following lemma shows that the expectation of $c_{\varepsilon_n}(k)$ conditioned on the degrees is the sum of $g_n(D_u, D_v, D_w)$ over all degrees in the major contributing regime:

Lemma 5.3. *On the event J_n defined in (5.4.8),*

$$\mathbb{E}_n \left[c_{\varepsilon_n}(W_n^k(\delta)) \right] = \frac{\sum_{(u,v) \in W_n^k(\delta)} g_n(k, D_u, D_v)}{k(k-1)} (1 + o_{\mathbb{P}}(1)). \quad (5.5.2)$$

Proof. By (5.3.20) and (5.3.21)

$$\mathbb{E}_n \left[c_{\varepsilon_n}(W_n^k(\delta)) \right] = \frac{1}{2|M_{\varepsilon_n}(k)|} \frac{\sum_{w \in M_{\varepsilon_n}(k)} \sum_{(u,v) \in W_n^k(\delta)} \mathbb{P}_n(\Delta_{u,v,w} = 1)}{k(k-1)/2}, \quad (5.5.3)$$

where $\Delta_{u,v,w}$ denotes the event that a triangle is present on vertices u, v and w . We write the probability that a specific triangle on vertices u, v and w exists as

$$\begin{aligned} \mathbb{P}_n(\Delta_{u,v,w} = 1) &= 1 - \mathbb{P}_n(X_{uw} = 0) - \mathbb{P}_n(X_{vw} = 0) - \mathbb{P}_n(X_{uv} = 0) \\ &\quad + \mathbb{P}_n(X_{uw} = X_{vw} = 0) + \mathbb{P}_n(X_{uv} = X_{vw} = 0) \\ &\quad + \mathbb{P}_n(X_{uv} = X_{uw} = 0) - \mathbb{P}_n(X_{uv} = X_{uw} = X_{vw} = 0). \end{aligned} \quad (5.5.4)$$

In the major contributing ranges, $D_u, D_v, D_w = O_{\mathbb{P}}(n^{1/(\tau-1)})$, and the product of the degrees is $O(n)$. By Lemma 7.1

$$\mathbb{P}_n(X_{uv} = X_{vw} = 0) = e^{-D_u D_v / L_n} e^{-D_v D_w / L_n} (1 + o_{\mathbb{P}}(n^{-(\tau-2)/(\tau-1)})) \quad (5.5.5)$$

and

$$\begin{aligned} \mathbb{P}_n(X_{uv} = X_{vw} = X_{uw} = 0) \\ = e^{-D_u D_v / L_n} e^{-D_v D_w / L_n} e^{-D_u D_w / L_n} (1 + o_{\mathbb{P}}(n^{-(\tau-2)/(\tau-1)})). \end{aligned} \quad (5.5.6)$$

Therefore,

$$\begin{aligned} \mathbb{P}_n(\Delta_{u,v,w} = 1) &= (1 + o_{\mathbb{P}}(1))(1 - e^{-D_u D_v / L_n})(1 - e^{-D_u D_w / L_n})(1 - e^{-D_v D_w / L_n}) \\ &= (1 + o_{\mathbb{P}}(1))g_n(D_u, D_v, D_w), \end{aligned} \quad (5.5.7)$$

where we have used that for $D_u D_v = O(n)$

$$1 - e^{-D_u D_v / L_n} (1 + o_{\mathbb{P}}(n^{-(\tau-2)/(\tau-1)})) = (1 - e^{-D_u D_v / L_n})(1 + o_{\mathbb{P}}(1)). \quad (5.5.8)$$

Lemma 5.1 shows that, given $D_w^{(\text{er})} = k$,

$$g_n(D_w, D_u, D_v) = g_n(k, D_u, D_v)(1 + o_{\mathbb{P}}(1)). \quad (5.5.9)$$

Thus, we obtain

$$\begin{aligned} \mathbb{E}_n \left[c_{\varepsilon_n}(W_n^k(\delta)) \right] &= \frac{\sum_{w \in M_{\varepsilon_n}(k)} \sum_{(u,v) \in W_n^k(\delta)} g_n(D_w, D_u, D_v)}{|M_{\varepsilon_n}(k)|k(k-1)} (1 + o_{\mathbb{P}}(1)) \\ &= \frac{\sum_{(u,v) \in W_n^k(\delta)} g_n(k, D_u, D_v)}{k(k-1)} (1 + o_{\mathbb{P}}(1)), \end{aligned} \quad (5.5.10)$$

which proves the lemma. \square

5.5.2 Analysis of asymptotic formula

In the previous section, we have shown that the expected value of $c_{\varepsilon_n}(k)$ in the major contributing regime is the sum of a function $g_n(k, D_u, D_v)$ over all vertices u and v with degrees in the major contributing regime if we condition on the degrees. That is

$$\mathbb{E}_n \left[c_{\varepsilon_n}(W_n^k(\delta)) \right] = \frac{1 + o_{\mathbb{P}}(1)}{k(k-1)} \sum_{(u,v) \in W_n^k(\delta)} (1 - e^{-kD_v / L_n})(1 - e^{-kD_u / L_n})(1 - e^{-D_u D_v / L_n}). \quad (5.5.11)$$

This expected value does not yet take into account that the degrees are sampled i.i.d. from a power-law distribution. In this section, we prove that this expected value converges to a constant when we take the randomness of the degrees into account. We will make use of the following lemmas:

Lemma 5.4. *Let $A \subset \mathbb{R}^2$ be a bounded set and $f(t_1, t_2)$ be a bounded, continuous function on A . Let $M^{(n)}$ be a random measure such that for all $S \subseteq A$, $M^{(n)}(S) \xrightarrow{\mathbb{P}} \lambda(S) = \int_S d\lambda(t_1, t_2)$ for some deterministic measure λ . Then,*

$$\int_A f(t_1, t_2) dM^{(n)}(t_1, t_2) \xrightarrow{\mathbb{P}} \int_A f(t_1, t_2) d\lambda(t_1, t_2). \quad (5.5.12)$$

Proof. Fix $\eta > 0$. Since f is bounded and continuous on A , for any $\delta > 0$, we can find $m < \infty$, disjoint sets $(B_i)_{i \in [m]}$ and constants $(b_i)_{i \in [m]}$ such that $\cup B_i = A$ and

$$\left| f(t_1, t_2) - \sum_{i=1}^m b_i \mathbb{1}_{\{(t_1, t_2) \in B_i\}} \right| < \delta, \quad (5.5.13)$$

for all $(t_1, t_2) \in A$. Because $M^{(n)}(B_i) \xrightarrow{\mathbb{P}} \lambda(B_i)$ for all i ,

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(|M^{(n)}(B_i) - \lambda(B_i)| > \eta / (mb_i) \right) = 0. \quad (5.5.14)$$

Then,

$$\begin{aligned} & \left| \int_A f(t_1, t_2) dM^{(n)}(t_1, t_2) - \int_A f(t_1, t_2) d\lambda(t_1, t_2) \right| \\ & \leq \left| \int_A f(t_1, t_2) - \sum_{i=1}^m b_i \mathbb{1}_{\{(t_1, t_2) \in B_i\}} dM^{(n)}(t_1, t_2) \right| \\ & \quad + \left| \int_A f(t_1, t_2) - \sum_{i=1}^m b_i \mathbb{1}_{\{(t_1, t_2) \in B_i\}} d\lambda(t_1, t_2) \right| \\ & \quad + \left| \sum_{i=1}^m b_i (M^{(n)}(B_i) - \lambda(B_i)) \right| \\ & \leq \delta M^{(n)}(A) + \delta \lambda(A) + o_{\mathbb{P}}(\eta). \end{aligned} \quad (5.5.15)$$

Now choosing $\delta < \eta / \lambda(A)$ proves the lemma. \square

The following lemma is a straightforward one-dimensional version of Lemma 5.4.

Lemma 5.5. *Let $M^{(n)}[a, b]$ be a random measure such that for all $0 < a < b$, $M^{(n)}[a, b] \xrightarrow{\mathbb{P}} \lambda[a, b] = \int_a^b d\lambda(t)$ for some deterministic measure λ . Let $f(t)$ be a bounded, continuous function on $[\delta, 1/\delta]$. Then,*

$$\int_{\delta}^{1/\delta} f(t) dM^{(n)}(t) \xrightarrow{\mathbb{P}} \int_{\delta}^{1/\delta} f(t) d\lambda(t). \quad (5.5.16)$$

Proof. This proof follows the same lines as the proof of Lemma 5.4. \square

Using these lemmas we investigate the convergence of the expectation of $c_{\varepsilon_n}(k)$ conditioned on the degrees. We treat the three ranges separately, but the proofs follow the same structure. First, we define a random measure $M^{(n)}$ that counts the normalized number of vertices with degrees in the major contributing regime. We then show that this measure converges to a deterministic measure λ , using that the degrees are i.i.d. samples of a power-law distribution. We then write the conditional expectation of the previous section as an integral over $M^{(n)}$. Then, we use Lemmas 5.4 or 5.5 to show that this converges to a deterministic integral.

First, we consider the case where k is in Range I.

Lemma 5.6. (Range I) For $1 < k = o(n^{(\tau-2)/(\tau-1)})$,

$$\frac{\mathbb{E}_n[c_{\varepsilon_n}(W_n^k(\delta))]}{n^{2-\tau} \log(n)} \xrightarrow{\mathbb{P}} \mu^{-\tau} C^2 \frac{3-\tau}{\tau-1} \int_{\delta}^{1/\delta} t^{1-\tau} (1-e^{-t}) dt. \quad (5.5.17)$$

Proof. Since the degrees are i.i.d. samples from a power-law distribution, $D_u = O_{\mathbb{P}}(n^{1/(\tau-1)})$ uniformly in $u \in [n]$. Thus, when $k = o(n^{(\tau-2)/(\tau-1)})$, $kD_u = o_{\mathbb{P}}(n)$ uniformly in $u \in [n]$. Therefore, we can Taylor expand the first two exponentials in (5.5.11), using that $1 - e^{-x} = x + O(x^2)$. By Lemma 5.3, this leads to

$$\mathbb{E}_n \left[c_{\varepsilon_n}(W_n^k(\delta)) \right] = (1 + o_{\mathbb{P}}(1)) \frac{k^2}{k(k-1)} \sum_{(u,v) \in W_n^k(\delta)} \frac{D_u D_v (1 - e^{-D_u D_v / L_n})}{L_n^2}. \quad (5.5.18)$$

Furthermore, since $D_u = O_{\mathbb{P}}(n^{1/(\tau-1)})$ while also $D_u D_v = \Theta(n)$ when $(u, v) \in W_n^k(\delta)$, we can add the indicator that $K_1 n^{(\tau-2)/(\tau-1)} < D_u < K_2 n^{1/(\tau-1)}$ for $0 < K_1, K_2 < \infty$. We then define the random measure

$$M^{(n)}[a, b] = \frac{(\mu n)^{\tau-1}}{\log(n)n^2} \sum_{u \neq v \in [n]} \mathbb{1}_{\{D_u D_v \in n\mu[a, b], K_1 n^{(\tau-2)/(\tau-1)} < D_u < K_2 n^{1/(\tau-1)}\}}. \quad (5.5.19)$$

We write the expected value of this measure as

$$\begin{aligned} \mathbb{E} [M^{(n)}[a, b]] &= \frac{(\mu n)^{\tau-1}}{\log(n)n^2} \mathbb{E} \left[\left| \left\{ u \neq v : D_u D_v \in [a, b] \mu n, D_u \in [K_1 n^{\frac{\tau-2}{\tau-1}}, K_2 n^{\frac{1}{\tau-1}}] \right\} \right| \right] \\ &= \frac{(\mu n)^{\tau-1} (n-1)}{\log(n)n} \mathbb{P} \left(D_1 D_2 \in [a, b] \mu n, D_1 \in [K_1 n^{\frac{\tau-2}{\tau-1}}, K_2 n^{\frac{1}{\tau-1}}] \right) \\ &= \frac{(\mu n)^{\tau-1}}{\log(n)} \int_{K_1 n^{(\tau-2)/(\tau-1)}}^{K_2 n^{\frac{1}{\tau-1}}} \int_{a\mu n/x}^{b\mu n/x} C^2(xy)^{-\tau} dy dx \\ &= C^2 \frac{(\mu n)^{\tau-1} (n-1)}{\log(n)n} \int_{K_1 n^{(\tau-2)/(\tau-1)}}^{K_2 n^{\frac{1}{\tau-1}}} \frac{1}{x} dx \int_{a\mu n}^{b\mu n} u^{-\tau} du \\ &= C^2 \frac{n-1}{n} \int_a^b t^{-\tau} dt \left(\frac{3-\tau}{\tau-1} + \frac{\log(K_2/K_1)}{\log(n)} \right), \end{aligned} \quad (5.5.20)$$

where we used the change of variables $u = xy$ and $t = u/(\mu n)$. Thus,

$$\lim_{n \rightarrow \infty} \mathbb{E} [M^{(n)}[a, b]] = C^2 \frac{3-\tau}{\tau-1} \int_a^b t^{-\tau} dt =: \lambda[a, b]. \quad (5.5.21)$$

Furthermore, the variance of this measure can be written as

$$\begin{aligned} \text{Var} (M^{(n)}[a, b]) &= \frac{(\mu n)^{2\tau-6} \mu^2}{\log^2(n)} \sum_{u, v, w, z} \left(\mathbb{P} \left(D_u D_v, D_w D_z \in \mu n[a, b], D_u, D_w \in [K_1 n^{\frac{\tau-2}{\tau-1}}, K_2 n^{\frac{1}{\tau-1}}] \right) \right. \\ &\quad \left. - \mathbb{P} \left(D_u D_v \in \mu n[a, b], D_u \in [K_1 n^{\frac{\tau-2}{\tau-1}}, K_2 n^{\frac{1}{\tau-1}}] \right) \right)^2 \end{aligned}$$

$$\times \mathbb{P} \left(D_w D_z \in \mu n[a, b], D_w \in [K_1 n^{\frac{\tau-2}{\tau-1}}, K_2 n^{\frac{1}{\tau-1}}] \right). \quad (5.5.22)$$

Since the degrees are an i.i.d. sample from a power-law distribution, the contribution to the variance for $|\{u, v, w, z\}| = 4$ is zero. The contribution from $|\{u, v, w, z\}| = 3$ can be bounded as

$$\begin{aligned} & \frac{(\mu n)^{2\tau-6} \mu^2}{\log^2(n)} \sum_{u,v,w} \mathbb{P} (D_u D_v, D_u D_w \in \mu n[a, b]) \\ &= \frac{\mu^{2\tau-4} n^{2\tau-3}}{\log^2(n)} \mathbb{P} (D_1 D_2, D_1 D_3 \in \mu n[a, b]) \\ &= \frac{\mu^{2\tau-4} n^{2\tau-3}}{\log^2(n)} \int_1^\infty C x^{-\tau} \left(\int_{an/x}^{bn/x} C y^{-\tau} dy \right)^2 dx \\ &\leq K \frac{n^{-1}}{\log^2(n)}, \end{aligned} \quad (5.5.23)$$

for some constant K . Similarly, the contribution for $u = z, v = w$ can be bounded as

$$\begin{aligned} \frac{(\mu n)^{2\tau-6} \mu^2}{\log^2(n)} \sum_{u,v} \mathbb{P} (D_u D_v \in \mu n[a, b]) &= \frac{\mu^{2\tau-6} n^{2\tau-4}}{\log^2(n)} \mathbb{P} (D_1 D_2 \in \mu n[a, b]) \\ &\leq K \frac{n^{2\tau-4}}{\log^2(n)} n^{1-\tau} \log(n) = K \frac{n^{\tau-3}}{\log(n)}, \end{aligned} \quad (5.5.24)$$

for some constant K . Thus, $\text{Var} (M^{(n)}[a, b]) = o(1)$. Therefore, a second moment method yields that for every $a, b > 0$,

$$M^{(n)}[a, b] \xrightarrow{\mathbb{P}} \lambda[a, b]. \quad (5.5.25)$$

Using the definition of $M^{(n)}$ in (5.5.19) and that $L_n^{-1} = (\mu n)^{-1} (1 + o_{\mathbb{P}}(1))$,

$$\begin{aligned} & \sum_{(u,v) \in W_n^k(\delta)} \frac{D_u D_v (1 - e^{-D_u D_v / L_n})}{L_n^2} \\ &= \mu^{1-\tau} n^{3-\tau} \log(n) \int_\delta^{1/\delta} \frac{t}{L_n} (1 - e^{-t}) dM^{(n)}(t) \\ &= \mu^{-\tau} n^{2-\tau} \log(n) \int_\delta^{1/\delta} t (1 - e^{-t}) dM^{(n)}(t) (1 + o_{\mathbb{P}}(1)). \end{aligned} \quad (5.5.26)$$

By Lemma 5.5 and (5.5.25),

$$\begin{aligned} \int_\delta^{1/\delta} t (1 - e^{-t}) dM^{(n)}(t) &\xrightarrow{\mathbb{P}} \int_\delta^{1/\delta} t (1 - e^{-t}) d\lambda(t) \\ &= C^2 \frac{3-\tau}{\tau-1} \int_\delta^{1/\delta} t^{1-\tau} (1 - e^{-t}) dt. \end{aligned} \quad (5.5.27)$$

If we first let $n \rightarrow \infty$, and then $K_1 \rightarrow 0$ and $K_2 \rightarrow \infty$, we obtain from (5.5.18), (5.5.26) and (5.5.27) that

$$\frac{\mathbb{E}_n [c_{\varepsilon_n}(k), W_n^k(\delta)]}{n^{2-\tau} \log(n)} \xrightarrow{\mathbb{P}} C^2 \mu^{-\tau} \frac{3-\tau}{\tau-1} \int_\delta^{1/\delta} t^{1-\tau} (1 - e^{-t}) dt. \quad (5.5.28)$$

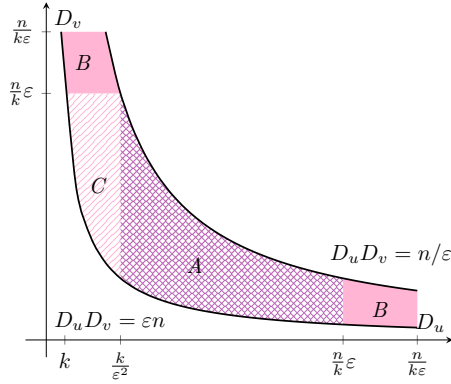


Figure 5.7: Contributing regime for $n^{(\tau-2)/(\tau-1)} < k < \sqrt{n}$.

□

Lemma 5.7. (Range II) When $k = \Omega(n^{(\tau-2)/(\tau-1)})$ and $k = o(\sqrt{n})$,

$$\frac{\mathbb{E}_n[c_{\epsilon n}(W_n^k(\delta))]}{n^{2-\tau} \log(n/k^2)} \xrightarrow{\mathbb{P}} C^2 \mu^{-\tau} \int_{\delta}^{1/\delta} t^{1-\tau} (1 - e^{-t}) dt. \quad (5.5.29)$$

Proof. We split the major contributing regime into three parts, depending on the values of D_u and D_v , as visualized in Figure 5.7. We denote the contribution to the clustering coefficient where $D_u \in [k/\delta^2, \delta n/k]$ (area A of Figure 5.7) by $c_{\epsilon n}^{(1)}(W_n^k(\delta))$, the contribution from D_u or $D_v \in [\delta n/k, n/(\delta k)]$ (area B of Figure 5.7) by $c_{\epsilon n}^{(2)}(W_n^k(\delta))$ and the contribution from $D_u \in [k, k/\delta^2]$ and $D_v \in [\delta^3 n/k, \delta n/k]$ (area C of Figure 5.7) by $c_{\epsilon n}^{(3)}(W_n^k(\delta))$. We first study the contribution of area A. In this situation, $D_u, D_v < \delta n/k$, so that we can Taylor expand the exponentials e^{-kD_u/L_n} and e^{-kD_v/L_n} in (5.5.11). This results in

$$\begin{aligned} \mathbb{E}_n[c_{\epsilon n}^{(1)}(W_n^k(\delta))] &= \frac{1}{k^2} \sum_{\substack{(u,v) \in W_n^k(\delta), \\ D_u \in [k/\delta^2, \delta n/k]}} (1 - e^{-kD_u/L_n})(1 - e^{-kD_v/L_n})(1 - e^{-D_u D_v/L_n}) \\ &= (1 + o_{\mathbb{P}}(1)) \sum_{\substack{(u,v) \in W_n^k(\delta), \\ D_u \in [k/\delta^2, \delta n/k]}} \frac{D_u D_v}{L_n^2} (1 - e^{-D_u D_v/L_n}). \end{aligned} \quad (5.5.30)$$

Now we define the random measure

$$M_1^{(n)}[a, b] = \frac{(\mu n)^{\tau-1}}{\log(\delta^3 n/k^2) n^2} \sum_{u,v \in [n]} \mathbb{1}_{\{D_u D_v \in \mu n[a, b], D_u \in [k/\delta^2, \delta n/k]\}}. \quad (5.5.31)$$

As similar reasoning as in (5.5.25) shows that

$$M_1^{(n)}[a, b] \xrightarrow{\mathbb{P}} C^2 \int_a^b t^{-\tau} dt := \lambda_2[a, b]. \quad (5.5.32)$$

By (5.5.30), we can write the the expected value of $c_{\varepsilon_n}^{(1)}(W_n^k(\delta))$ as

$$\begin{aligned} \mathbb{E}_n [c_{\varepsilon_n}^{(1)}(W_n^k(\delta))] &= (1 + o_{\mathbb{P}}(1)) \sum_{\substack{(u,v) \in W_n^k(\delta), \\ D_u \in [k/\delta^3, \delta n/k]}} \frac{D_u D_v}{L_n^2} (1 - e^{-D_u D_v / L_n}) \\ &= (1 + o_{\mathbb{P}}(1)) \mu^{-\tau} n^{2-\tau} \log(\delta^3 n / k^2) \int_{\delta}^{1/\delta} t(1 - e^{-t}) dM_1^{(n)}(t). \end{aligned} \quad (5.5.33)$$

Thus, by Lemma 5.5

$$\mathbb{E}_n [c_{\varepsilon_n}^{(1)}(W_n^k(\delta))] = (1 + o_{\mathbb{P}}(1)) 2\mu^{-\tau} n^{2-\tau} \log(\delta^3 n / k^2) \int_{\delta}^{1/\delta} t(1 - e^{-t}) d\lambda_2(t). \quad (5.5.34)$$

Then we study the contribution of area B in Figure 5.7. This area consists of two parts, the part where $D_u \in [\delta n/k, n/(k\delta)]$, and the part where $D_v \in [\delta n/k, n/(k\delta)]$. By symmetry, these two contributions are the same and therefore we only consider the case where $D_u \in [\delta n/k, n/(k\delta)]$. Then, we can Taylor expand $e^{-D_v k / L_n}$ in (5.5.11), which yields

$$\mathbb{E}_n [c_{\varepsilon_n}^{(2)}(W_n^k(\delta))] = \frac{2}{k^2} \sum_{\substack{(u,v) \in W_n^k(\delta), \\ D_u > \delta n/k}} (1 - e^{-kD_u / L_n}) \frac{D_v k}{L_n} (1 - e^{-D_u D_v / L_n}). \quad (5.5.35)$$

Define the random measure

$$M_2^{(n)}([a, b], [c, d]) := \frac{(\mu n)^{\tau-1}}{n^2} \sum_{u,v \in [n]} \mathbb{1}_{\{D_u D_v \in \mu n [a,b], D_u \in (\mu n/k) [c,d]\}}. \quad (5.5.36)$$

Then we obtain

$$\begin{aligned} \mathbb{E}_n [c_{\varepsilon_n}^{(2)}(W_n^k(\delta))] &= \frac{2}{kL_n} \sum_{\substack{(u,v) \in W_n^k(\delta), \\ D_u > \delta n/k}} \frac{L_n}{D_u k} (1 - e^{-kD_u / L_n}) \frac{D_u D_v k}{L_n} (1 - e^{-D_u D_v / L_n}) \\ &= 2\mu^{-\tau} n^{2-\tau} \int_{\delta}^{1/\delta} \int_{\delta}^{1/\delta} \frac{t_1}{t_2} (1 - e^{-t_1})(1 - e^{-t_2}) dM_2^{(n)}(t_1, t_2) (1 + o_{\mathbb{P}}(1)). \end{aligned} \quad (5.5.37)$$

Again, using a first moment method and a second moment method, we can show that

$$M_2^{(n)}([a, b], [c, d]) \xrightarrow{\mathbb{P}} C^2 \int_a^b t^{-\tau} dt \int_c^d \frac{1}{v} dv =: \lambda[a, b] \nu[c, d]. \quad (5.5.38)$$

Very similarly to the proof of Lemma 5.4 we can show that

$$\begin{aligned} &\int_{\delta}^{1/\delta} \int_{\delta}^{1/\delta} \frac{t_1}{t_2} (1 - e^{-t_1})(1 - e^{-t_2}) dM_2^{(n)}(t_1, t_2) \\ &\xrightarrow{\mathbb{P}} \int_{\delta}^{1/\delta} \int_{\delta}^{1/\delta} \frac{t_1}{t_2} (1 - e^{-t_1})(1 - e^{-t_2}) d\lambda(t_1) d\nu(t_2). \end{aligned} \quad (5.5.39)$$

The latter integral can be written as

$$\begin{aligned}
& \int_{\delta}^{1/\delta} \int_{\delta}^{1/\delta} \frac{t_1}{t_2} (1 - e^{-t_1})(1 - e^{-t_2}) d\lambda(t_1) d\nu(t_2) \\
&= C^2 \int_{\delta}^{1/\delta} \int_{\delta}^{1/\delta} t_2^{-2} t_1^{1-\tau} (1 - e^{-t_2})(1 - e^{-t_1}) dt_1 dt_2 \\
&= C^2 \int_{\delta}^{1/\delta} \frac{1}{t_2^2} (1 - e^{-t_2}) dt_2 \int_{\delta}^{1/\delta} t_1^{1-\tau} (1 - e^{-t_1}) dt_1. \tag{5.540}
\end{aligned}$$

The left integral results in

$$\begin{aligned}
& \int_{\delta}^{1/\delta} \frac{1}{t_2^2} (1 - e^{-t_2}) dt_2 = \left[\frac{e^{-t_2} - 1}{t_2} + \text{Ei}(t_2) \right]_{t_2=\delta}^{t_2=1/\delta} \\
&= \delta(e^{-1/\delta} - 1) - \frac{e^{-\delta} - 1}{\delta} + \int_{1/\delta}^{\infty} \frac{1}{u} e^{-u} du - \log(\delta) - \sum_{j=1}^{\infty} \frac{\delta^j}{k!j} \\
&= \log\left(\frac{1}{\delta}\right) + \int_{1/\delta}^{\infty} \frac{1}{u} e^{-u} du + \delta(e^{-1/\delta} - 1) - \frac{e^{-\delta} - 1}{\delta} - \sum_{j=1}^{\infty} \frac{\delta^j}{k!j} \\
&= \log\left(\frac{1}{\delta}\right) + f(\delta), \tag{5.541}
\end{aligned}$$

where Ei denotes the exponential integral and we have used the Taylor series for the exponential integral. We can show that $f(\delta) < \infty$ for fixed $\delta \in (0, \infty)$. In fact, $f(\delta) \rightarrow 1$ as $\delta \rightarrow 0$.

Finally, we study the contribution of area C in Figure 5.7, where $D_u \in [k, k/\delta^2]$ and $D_v \in [n/k\delta^3, n/k\delta]$. In this regime, $D_u k = o(n)$ and $D_v k = o(n)$ so that we can Taylor expand the first two exponentials in (5.5.11). This results in

$$\mathbb{E}_n \left[c_{\varepsilon_n}^{(3)}(W_n^k(\delta)) \right] = (1 + o(1)) \sum_{\substack{u, v: D_v \in [\delta^3 n/k, \delta n/k], \\ D_u D_v > \delta n, D_u \in [k, k/\delta^2]}} (1 - e^{-D_u D_v / L_n}) \frac{D_u D_v}{L_n}. \tag{5.542}$$

We define the random measure

$$M_3^{(n)}([a, b], [c, d]) := \frac{(\mu n)^{\tau-1}}{n^2} \sum_{u, v} \mathbb{1}_{\{D_u \in \sqrt{\mu} k [a, b], D_v \in (\sqrt{\mu} n/k) [c, d]\}}. \tag{5.543}$$

Then,

$$\mathbb{E}_n \left[c_{\varepsilon_n}^{(3)}(W_n^k(\delta)) \right] = (1 + o_{\mathbb{P}}(1)) n^{2-\tau} \mu^{-\tau} \int_1^{1/\delta^2} \int_{\delta/t_1}^{\delta} (t_1 t_2) (1 - e^{-t_1 t_2}) dM_3^{(n)}(t_1, t_2). \tag{5.544}$$

Again using a first moment method and a second moment method we can show that

$$M_3^{(n)}([a, b], [c, d]) \xrightarrow{\mathbb{P}} C^2 \int_a^b u^{-\tau} du \int_c^d v^{-\tau} dv. \tag{5.545}$$

In a similar way, we can show that for $B \subseteq [1, 1/\delta^2] \times [\delta^3, \delta]$,

$$M_3^{(n)}(B) \xrightarrow{\mathbb{P}} C^2 \int \int_B (uv)^{-\tau} dudv. \quad (5.5.46)$$

Thus, by Lemma 5.4,

$$\int_1^{1/\delta^2} \int_{\delta/t_1}^{\delta} (t_1 t_2) (1 - e^{-t_1 t_2}) dM_3^{(n)}(t_1, t_2) \xrightarrow{\mathbb{P}} C^2 \int_1^{1/\delta^2} \int_{\delta/x}^{\delta} (xy)^{1-\tau} (1 - e^{-xy}) dy dx. \quad (5.5.47)$$

We evaluate the latter integral as

$$\begin{aligned} \int_1^{1/\delta^2} \int_{\delta/x}^{\delta} (xy)^{1-\tau} (1 - e^{-xy}) dy dx &= \int_1^{1/\delta^2} \int_{\delta}^{\delta v} \frac{1}{v} u^{1-\tau} (1 - e^{-u}) du dv \\ &= \int_{\delta}^{1/\delta} \int_{u/\delta}^{1/\delta^2} \frac{1}{v} u^{1-\tau} (1 - e^{-u}) dv du \\ &= \log\left(\frac{1}{\delta}\right) \int_{\delta}^{1/\delta} u^{1-\tau} (1 - e^{-u}) du \\ &\quad + \int_{\delta}^{1/\delta} \log\left(\frac{1}{u}\right) u^{1-\tau} (1 - e^{-u}) du, \end{aligned} \quad (5.5.48)$$

where we used the change of variables $u = xy$ and $v = x$. Summing all three contributions to the expectation under \mathbb{E}_n of the clustering coefficient yields

$$\begin{aligned} &\mathbb{E}_n[c_{\varepsilon_n}(W_n^k(\delta))] \\ &= \mathbb{E}_n[c_{\varepsilon_n}^{(1)}(W_n^k(\delta))] + \mathbb{E}_n[c_{\varepsilon_n}^{(2)}(W_n^k(\delta))] + \mathbb{E}_n[c_{\varepsilon_n}^{(3)}(W_n^k(\delta))] \\ &= C^2 \mu^{-\tau} n^{2-\tau} (1 + o_{\mathbb{P}}(1)) \left[\int_{\delta}^{1/\delta} t_1^{1-\tau} (1 - e^{-t_1}) dt_1 \right. \\ &\quad \left. \times \left(\log\left(\frac{n\delta^2}{k^2}\right) + 3 \log\left(\frac{1}{\delta}\right) + 2f(\delta) \right) + \int_{\delta}^{1/\delta} \log\left(\frac{1}{u}\right) u^{1-\tau} (1 - e^{-u}) du \right] \\ &= C^2 (1 + o_{\mathbb{P}}(1)) \mu^{-\tau} n^{2-\tau} \left[\int_{\delta}^{1/\delta} t_1^{1-\tau} (1 - e^{-t_1}) dt_1 \left(\log\left(\frac{n}{k^2}\right) + 2f(\delta) \right) \right. \\ &\quad \left. + \int_{\delta}^{1/\delta} \log\left(\frac{1}{u}\right) u^{1-\tau} (1 - e^{-u}) du \right]. \end{aligned} \quad (5.5.49)$$

Dividing by $n^{2-\tau} \log(n/k^2)$ and taking the limit of $n \rightarrow \infty$ then shows that

$$\frac{\mathbb{E}_n[c_{\varepsilon_n}(W_n^k(\delta))]}{n^{2-\tau} \log(n/k^2)} \xrightarrow{\mathbb{P}} C^2 \mu^{-\tau} \int_{\delta}^{1/\delta} x^{1-\tau} (1 - e^{-x}) dx. \quad (5.5.50)$$

□

Lemma 5.8. (Range III) For $k = \Omega(\sqrt{n})$,

$$\frac{\mathbb{E}_n[c_{\varepsilon_n}(W_n^k(\delta))]}{n^{5-2\tau} k^{2\tau-6}} \xrightarrow{\mathbb{P}} C^2 \mu^{3-2\tau} \left(\int_{\delta}^{1/\delta} t^{1-\tau} (1 - e^{-t}) dt \right)^2. \quad (5.5.51)$$

Proof. When $k = \Omega(\sqrt{n})$, the major contribution is from u, v with $D_u, D_v = \Theta(n/k)$, so that $D_u D_v = o(n)$. Therefore, we can Taylor expand the exponential $e^{-D_u D_v / L_n}$ in (5.5.11). Thus, we write the expected value of $c_{\varepsilon_n}(k)$ as

$$\begin{aligned} \mathbb{E}_n \left[c_{\varepsilon_n}(W_n^k(\delta)) \right] &= \frac{1}{k^2} \sum_{(u,v) \in W_n^k(\delta)} (1 - e^{-kD_u/L_n})(1 - e^{-kD_v/L_n})(1 - e^{-D_u D_v / L_n}) \\ &= \frac{1}{k^2} \sum_{(u,v) \in W_n^k(\delta)} (1 - e^{-kD_u/L_n})(1 - e^{-kD_v/L_n}) \frac{D_u D_v}{L_n} (1 + o_{\mathbb{P}}(1)). \end{aligned} \quad (5.5.2)$$

Define the random measure

$$N_1^{(n)}[a, b] = \frac{(\mu n)^{\tau-1}}{n} k^{1-\tau} \sum_{u \in [n]} \mathbb{1}_{\{D_u \in (\mu n/k)[a, b]\}}, \quad (5.5.3)$$

and let $N^{(n)}$ be the product measure $N_1^{(n)} \times N_1^{(n)}$. Since all degrees are i.i.d. samples from a power-law distribution, the number of vertices with degrees in interval $[q_1, q_2]$ is distributed as a $\text{Bin}(n, C(q_1^{1-\tau} - q_2^{1-\tau}))$ random variable. Therefore,

$$\begin{aligned} N_1^{(n)}([a, b]) &= \frac{(\mu n)^{\tau-1} k^{1-\tau}}{n} |\{i : D_i \in (\mu n/k)[a, b]\}| \\ &\xrightarrow{\mathbb{P}} \lim_{n \rightarrow \infty} (\mu n)^{\tau-1} k^{1-\tau} \mathbb{P}(D_i \in (\mu n/k)[a, b]) \\ &= (\mu n)^{\tau-1} k^{1-\tau} \int_{a\mu n/k}^{b\mu n/k} Cx^{-\tau} dx = C \int_a^b t^{-\tau} dt := \lambda([a, b]), \end{aligned} \quad (5.5.4)$$

where we have used the substitution $t = xk/(\mu n)$. Then,

$$\begin{aligned} &\sum_{(u,v) \in W_n^k(\delta)} (1 - e^{-kD_u/L_n})(1 - e^{-kD_v/L_n}) \frac{D_u D_v}{L_n} \\ &= \frac{L_n}{k^2} \sum_{(u,v) \in W_n^k(\delta)} (1 - e^{-kD_u/L_n})(1 - e^{-kD_v/L_n}) \frac{D_u k}{L_n} \frac{D_v k}{L_n} \\ &= (1 + o_{\mathbb{P}}(1)) \mu^{3-2\tau} n^{5-2\tau} k^{2\tau-4} \int_{\delta}^{1/\delta} \int_{\delta}^{1/\delta} t_1 t_2 (1 - e^{-t_1})(1 - e^{-t_2}) dN^{(n)}(t_1, t_2). \end{aligned} \quad (5.5.5)$$

Combining this with (5.5.2) yields

$$\begin{aligned} \frac{\mathbb{E}_n[c_{\varepsilon_n}(W_n^k(\delta))]}{n^{5-2\tau} k^{2\tau-4}} &= (1 + o_{\mathbb{P}}(1)) \mu^{2\tau-3} \int_{\delta}^{1/\delta} \int_{\delta}^{1/\delta} t_1 t_2 (1 - e^{-t_1})(1 - e^{-t_2}) dN^{(n)}(t_1, t_2) \\ &= (1 + o_{\mathbb{P}}(1)) \mu^{2\tau-3} \left(\int_{\delta}^{1/\delta} t_1 (1 - e^{-t_1}) dN_1^{(n)}(t_1) \right)^2. \end{aligned} \quad (5.5.6)$$

We then use Lemma 5.5, which shows that

$$\begin{aligned} \int_{\delta}^{1/\delta} t_1^{1-\tau} (1 - e^{-t_1}) dN_1^{(n)}(t_1) &\xrightarrow{\mathbb{P}} C \int_{\delta}^{1/\delta} t_1 (1 - e^{-t_1}) d\lambda(t_1) \\ &= C \int_{\delta}^{1/\delta} t_1^{1-\tau} (1 - e^{-t_1}) dt_1. \end{aligned} \quad (5.5.7)$$

Then, we can conclude from (5.5.56) and (5.5.57) that

$$\frac{\mathbb{E}_n[c_{\varepsilon_n}(W_n^k(\delta))]}{n^{5-2\tau}k^{2\tau-6}} \xrightarrow{\mathbb{P}} C^2\mu^{3-2\tau} \left(\int_{\delta}^{1/\delta} t_1^{1-\tau}(1-e^{-t_1})dt_1 \right)^2. \quad (5.5.58)$$

□

5.5.3 Variance of the local clustering coefficient

In the following lemma, we give a bound on the variance of $c_{\varepsilon_n}(W_n^k(\delta))$:

Lemma 5.9. *For all ranges, under $J_n \cap \mathcal{A}_n$,*

$$\frac{\text{Var}_n \left(c_{\varepsilon_n}(W_n^k(\delta)) \right)}{\mathbb{E}_n[c_{\varepsilon_n}(W_n^k(\delta))]^2} \xrightarrow{\mathbb{P}} 0. \quad (5.5.59)$$

Proof. We analyze the variance in a similar way as we have analyzed the expected value of $c_{\varepsilon_n}(k)$ conditioned on the degrees in Section 5.5.1. We can write the variance of $c_{\varepsilon_n}(W_n^k(\delta))$ as

$$\begin{aligned} \text{Var}_n \left(c_{\varepsilon_n}(W_n^k(\delta)) \right) &= \frac{1}{k^2(k-1)^2|M_{\varepsilon_n}(k)|^2} \\ &\times \sum_{i,j \in M_{\varepsilon_n}(k)} \sum_{\substack{(u,v) \in W_n^k(\delta), \\ (w,z) \in W_n^k(\delta)}} \mathbb{P}_n(\Delta_{i,u,v}\Delta_{j,w,z}) - \mathbb{P}_n(\Delta_{i,u,v})\mathbb{P}_n(\Delta_{j,w,z}), \end{aligned} \quad (5.5.60)$$

where $\Delta_{i,u,v}$ again denotes the event that vertices i, u and v form a triangle. Equation (5.5.60) splits into various cases, depending on the size of $\{i, j, u, v, w, z\}$. We denote the contribution of $|\{i, j, u, v, w, z\}| = r$ to the variance by $V^{(r)}(k)$. We first consider $V^{(6)}(k)$. By a similar reasoning as (5.5.7)

$$\begin{aligned} &\text{Var}_n \left(c_{\varepsilon_n}(W_n^k(\delta)) \right) \\ &= \frac{1}{|M_{\varepsilon_n}(k)|^2 k^2 (k-1)^2} \sum_{i,j \in M_{\varepsilon_n}(k)} \sum_{\substack{(u,v) \in W_n^k(\delta), \\ (w,z) \in W_n^k(\delta)}} \left(g_n(k, D_u, D_v) g_n(k, D_w, D_z) (1 + o_{\mathbb{P}}(1)) \right. \\ &\quad \left. - g_n(k, D_u, D_v) g_n(k, D_w, D_z) (1 + o_{\mathbb{P}}(1)) \right) \\ &= \sum_{(u,v), (w,z) \in W_n^k(\delta)} o_{\mathbb{P}} \left(\frac{g_n(k, D_u, D_v) g_n(k, D_w, D_z)}{k^2 (k-1)^2} \right) = o_{\mathbb{P}} \left(\mathbb{E}_n[c_{\varepsilon_n}(W_n^k(\delta))]^2 \right), \end{aligned} \quad (5.5.61)$$

where we have again replaced $g_n(D_i, D_u, D_v)$ by $g_n(k, D_u, D_v)$ because of (5.5.9). Since there are no overlapping edges when $|\{i, j, u, v, w, z\}| = 5$, $V^{(5)}(k)$ can be bounded similarly. This already shows that the contribution to the variance from 5 or 6 different vertices involved is small in all three ranges of k .

We then consider $V^{(4)}$, which is the contribution from two triangles where one edge overlaps. We show that these types of overlapping triangles are rare, so that

their contribution to the variance is small. If for example $i = j$ and $u = z$, then one edge from the vertex of degree k overlaps with another triangle. To bound this contribution, we use that $\mathbb{P}_n(\hat{X}_{ij} = 1) \leq \min(1, D_i D_j / L_n)$. Then we can bound the summand in (5.5.60) as

$$\begin{aligned}
& \mathbb{P}_n(\Delta_{i,u,v} \Delta_{i,w,u}) - \mathbb{P}_n(\Delta_{i,u,v}) \mathbb{P}_n(\Delta_{i,w,u}) \\
& \leq \mathbb{P}_n(\Delta_{i,u,v} \Delta_{i,w,u}) \\
& \leq \min\left(1, \frac{kD_u}{L_n}\right) \min\left(1, \frac{kD_v}{L_n - 2}\right) \min\left(1, \frac{D_u D_v}{L_n - 4}\right) \\
& \quad \times \min\left(1, \frac{kD_w}{L_n - 6}\right) \min\left(1, \frac{D_w D_u}{L_n - 8}\right) \\
& = (1 + O(n^{-1})) \min\left(1, \frac{kD_u}{L_n}\right) \min\left(1, \frac{kD_v}{L_n}\right) \\
& \quad \times \min\left(1, \frac{D_u D_v}{L_n}\right) \min\left(1, \frac{kD_w}{L_n}\right) \min\left(1, \frac{D_w D_u}{L_n}\right). \quad (5.5.62)
\end{aligned}$$

We first consider k in Ranges I or II, where $k = o(\sqrt{n})$. We bound the terms involving k in (5.5.62) by taking the second term of the minimum, while we bound $\min(D_u D_v / L_n, 1) \leq 1$, which results in

$$\begin{aligned}
\mathbb{P}_n(\Delta_{i,u,v} \Delta_{i,w,u}) - \mathbb{P}_n(\Delta_{i,u,v}) \mathbb{P}_n(\Delta_{i,w,u}) & \leq (1 + O(n^{-1})) \frac{k^3 D_u D_v D_w}{L_n^3} \\
& \leq O(1) \delta^{-1} \frac{k^3 D_w}{L_n^2}, \quad (5.5.63)
\end{aligned}$$

where we used that $D_u D_v < n/\delta$ when $(u, v) \in W_n^k(\delta)$. Therefore, the contribution to the variance in this situation can be bounded by

$$\begin{aligned}
& \frac{k^3}{k^4 |M_{\varepsilon_n}(k)|^2} \sum_{i \in M_{\varepsilon_n}(k)} \sum_{(u,v),(w,u) \in W_n^k(\delta)} \frac{\delta^{-1} D_w}{L_n^2} = \frac{1}{k |M_{\varepsilon_n}(k)|} \sum_{(u,v),(w,u) \in W_n^k(\delta)} \frac{\delta^{-1} D_w}{L_n^2} \\
& \leq \frac{\delta^{-1} O(n^{-1})}{k |M_{\varepsilon_n}(k)|} \sum_{u \in [n]} \frac{1}{\delta D_u} \left(\sum_{w \in [n]} \mathbb{1}_{\{D_w > \delta n / D_u\}} \right)^2, \quad (5.5.64)
\end{aligned}$$

where we used that $D_w = O(n/D_u)$ in $W_n^k(\delta)$. We then use Lemma 5.1 to further bound this as

$$\begin{aligned}
& \frac{k^3}{k^4 |M_{\varepsilon_n}(k)|^2} \sum_{i \in M_{\varepsilon_n}(k)} \sum_{(u,v),(w,u) \in W_n^k(\delta)} \frac{\delta^{-1} D_w}{L_n^2} \leq K(\delta) O_{\mathbb{P}} \left(\frac{1}{nk^{1-\tau}} \sum_u \left(\frac{n}{D_u} \right)^{3-2\tau} \right) \\
& \leq K(\delta) O_{\mathbb{P}} \left(n^{3-2\tau} k^{\tau-1} n^{(2-\tau)/(\tau-1)} \right). \quad (5.5.65)
\end{aligned}$$

Here $K(\delta)$ is a constant only depending on δ . Since $n^{(2-\tau)/(\tau-1)} k^{\tau-1} = o(n)$ when $k = o(\sqrt{n})$ and $\tau \in (2, 3)$, we have proven that this contribution is smaller than $n^{4-2\tau} \log^2(n)$ and smaller than $n^{4-2\tau} \log^2(n/k^2)$, as required by Lemmas 5.6 and 5.7

respectively. Now we consider the contribution from triangles that share the edge between vertices u and v . Using a similar reasoning as in (5.5.62), the contribution from the case $i \neq j$ and $u = z$ and $v = w$ can be bounded as

$$\begin{aligned} & \frac{1}{k^4 |M_{\varepsilon_n}(k)|^2} \sum_{i,j \in M_{\varepsilon_n}(k)} \sum_{(u,v) \in W_n^k(\delta)} \mathbb{P}_n(\Delta_{i,u,v} \Delta_{j,v,w}) - \mathbb{P}_n(\Delta_{i,u,v}) \mathbb{P}_n(\Delta_{j,v,w}) \\ & \leq \sum_{(u,v) \in W_n^k(\delta)} \frac{k^4 D_u^2 D_v^2}{k^4 L_n^4} \leq \delta^{-2} \mathbb{P}_n\left((D_u, D_v) \in W_n^k(\delta)\right) \\ & = \delta^{-2} O_{\mathbb{P}}\left(n^{1-\tau} \log(n)\right), \end{aligned} \quad (5.5.66)$$

where we used Lemma 5.1 and that $D_u D_v = O(n)$ when $(u, v) \in W_n^k(\delta)$. Since $n^{1-\tau} \log(n) = o(n^{4-2\tau} \log^2(n))$ for $\tau \in (2, 3)$, this shows that this contribution is small enough.

When k is in Range III, we use similar bounds for $V^{(4)}$, now using that $D_u, D_v, D_w < \delta^{-1} n/k$. Under \mathcal{A}_n , $|M_{\varepsilon_n}(k)| \geq 1$. Again, we start by considering the case $i = j$ and $u = z$. We use (5.5.62), where we use that $D_u D_v < n^2 / (k\delta)^2$ and $D_u D_w < n^2 / (k\delta)^2$, and we take 1 for the other minima. This yields

$$\mathbb{P}_n(\Delta_{i,u,v} \Delta_{i,w,u}) - \mathbb{P}_n(\Delta_{i,u,v}) \mathbb{P}_n(\Delta_{i,w,u}) \leq O(n^2) k^{-4} \delta^{-4}. \quad (5.5.67)$$

Thus, the contribution to the variance from this case can be bounded as

$$\begin{aligned} & \frac{1}{k^4 |M_{\varepsilon_n}(k)|} \sum_{(u,v),(u,w) \in W_n^k(\delta)} O(n^2) k^{-4} \delta^{-4} \leq \frac{1}{k^4} O_{\mathbb{P}}\left(n^5 k^{-8} \delta^{-4} \mathbb{P}(D > n / (\delta k))^3\right) \\ & \leq O_{\mathbb{P}}\left(n^5 k^{-8} \delta^{-4} \left(\frac{n}{k\delta}\right)^{3-3\tau}\right) \\ & = O_{\mathbb{P}}\left(k^{3\tau-11} n^{8-3\tau}\right) \delta^{3\tau-7}, \end{aligned} \quad (5.5.68)$$

where we used Lemma 5.1. When $k = \Omega(\sqrt{n})$ and $\tau \in (2, 3)$, this contribution is smaller than $n^{10-4\tau} k^{4\tau-12}$, as required by Lemma 5.8. In the case where $i \neq j$, $u = z$ and $v = w$, we use a similar reasoning as in (5.5.62) to show that

$$\mathbb{P}_n(\Delta_{i,u,v} \Delta_{i,w,u}) - \mathbb{P}_n(\Delta_{i,u,v}) \mathbb{P}_n(\Delta_{i,w,u}) \leq O(n) k^{-2} \delta^{-2}. \quad (5.5.69)$$

Then the contribution of this situation to the variance can be bounded as

$$\frac{1}{k^4} \sum_{(u,v) \in W_n^k(\delta)} O(n) k^{-2} \delta^{-2} \leq O\left(\delta^{-2} n^3 k^{-6} \left(\frac{n}{\delta k}\right)^{2-2\tau}\right) = O\left(n^{5-2\tau} k^{2\tau-8}\right). \quad (5.5.70)$$

Again, this is smaller than $n^{10-4\tau} k^{4\tau-12}$, as required. Thus, the contribution of $V^{(4)}$ is small enough in all three ranges.

Finally, $V^{(3)}$, can be bounded as

$$\begin{aligned} & \frac{1}{k^4 |M_{\varepsilon_n}(k)|^2} \sum_{i \in M_{\varepsilon_n}(k)} \sum_{(u,v) \in W_n^k(\delta)} \mathbb{P}_n(\Delta_{i,u,v}) = \frac{1}{k^4 |M_{\varepsilon_n}(k)|} \mathbb{E}_n \left[c_{\varepsilon_n}(W_n^k(\delta)) \right] \\ & = \frac{1}{k^4 |M_{\varepsilon_n}(k)|} O_{\mathbb{P}}(f(k, n)). \end{aligned} \quad (5.5.71)$$

In Ranges I and II, we use that $|M_{\varepsilon_n}(k)| = O_{\mathbb{P}}(nk^{-\tau})$. Thus, this gives a contribution of

$$V^{(3)}(k) = O_{\mathbb{P}}\left(\frac{n^{2-\tau}\log(n)}{k^{4-\tau}n}\right) = O_{\mathbb{P}}\left(n^{1-\tau}\log(n)k^{\tau-4}\right), \quad (5.5.72)$$

which is small enough since $n^{1-\tau}k^{\tau-4} < n^{4-2\tau}$ for $\tau \in (2, 3)$ and $k = o(\sqrt{n})$. In Range III, again we assume that $|M_{\varepsilon_n}(k)| \geq 1$, since otherwise the variance of $c_{\varepsilon_n}(k)$ would be zero, and therefore small enough. Then (5.5.71) gives the bound

$$V^{(3)}(k) = O_{\mathbb{P}}\left(n^{5-2\tau}k^{2\tau-10}\right), \quad (5.5.73)$$

which is again smaller than $n^{10-4\tau}k^{4\tau-12}$ for $\tau \in (2, 3)$ and $k = \Omega(\sqrt{n})$. Thus, all contributions to the variance are small enough, which proves the claim. \square

Proof of Proposition 5.1. Combining Lemma 5.9 and the fact that $\mathbb{P}(J_n) = 1 - O(n^{-1/\tau})$ shows that

$$\frac{c_{\varepsilon_n}(W_n^k(\delta))}{\mathbb{E}_n[c_{\varepsilon_n}(W_n^k(\delta))]} \xrightarrow{\mathbb{P}} 1. \quad (5.5.74)$$

Then, Lemmas, 5.7 and 5.8 show that

$$\frac{c_{\varepsilon_n}(W_n^k(\delta))}{f(k, n)} \xrightarrow{\mathbb{P}} \begin{cases} C^2 \int_{\delta}^{1/\delta} t^{1-\tau} e^{-t} dt + O(\delta^{\kappa}) & k = o(\sqrt{n}) \\ C^2 \left(\int_{\delta}^{1/\delta} t^{1-\tau} e^{-t} dt \right)^2 + O(\delta^{\kappa}) & k = \Omega(\sqrt{n}), \end{cases} \quad (5.5.75)$$

which proves the proposition. \square

5.6 Contributions outside $W_n^k(\delta)$

In this section, we show that the contribution of triangles with degrees outside of the major contributing ranges as described in (5.3.6) is negligible. The following lemma bounds the contribution from triangles with vertices with degrees outside of $W_n^k(\delta)$:

Lemma 5.10. *There exists $\kappa > 0$ such that*

$$\limsup_{n \rightarrow \infty} \frac{\mathbb{E}_n[c_{\varepsilon_n}(\bar{W}_n^k(\delta))]}{f(n, k)} = O_{\mathbb{P}}(\delta^{\kappa}). \quad (5.6.1)$$

Proof. To compute the expected value of $c_{\varepsilon_n}(k)$, we use that $\mathbb{P}_n(\hat{X}_{ij} = 1) \leq \min(1, \frac{D_i D_j}{L_n})$. This yields

$$\mathbb{E}_n[c_{\varepsilon_n}(k)] \leq \frac{n^2 \mathbb{E}_n \left[\min(1, \frac{kD_u}{L_n}) \min(1, \frac{kD_v}{L_n}) \min(1, \frac{D_u D_v}{L_n}) \right]}{k(k-1)}. \quad (5.6.2)$$

Using Lemma 5.1, we obtain

$$\mathbb{E}_n[c_{\varepsilon_n}(k)] = n^2 k^{-2} O_{\mathbb{P}} \left(\mathbb{E} \left[\min \left(1, \frac{kD_u}{\mu n} \right) \min \left(1, \frac{kD_v}{\mu n} \right) \min \left(1, \frac{D_u D_v}{\mu n} \right) \right] \right), \quad (5.6.3)$$

where D_u and D_v are two independent copies of D . Similarly,

$$\begin{aligned} & \mathbb{E}_n [c_{\varepsilon_n}(\bar{W}_n^k(\delta))] \\ &= n^2 k^{-2} O_{\mathbb{P}} \left(\mathbb{E} \left[\min \left(1, \frac{kD_u}{\mu n} \right) \min \left(1, \frac{kD_v}{\mu n} \right) \min \left(1, \frac{D_u D_v}{\mu n} \right) \mathbb{1}_{\{(D_u, D_v) \in \bar{W}_n^k(\delta)\}} \right] \right), \end{aligned} \quad (5.6.4)$$

where

$$\begin{aligned} & \mathbb{E} \left[\min \left(1, \frac{kD_u}{\mu n} \right) \min \left(1, \frac{kD_v}{\mu n} \right) \min \left(1, \frac{D_u D_v}{\mu n} \right) \mathbb{1}_{\{(D_u, D_v) \in \bar{W}_n^k(\delta)\}} \right] \\ &= \int \int_{(x, y) \in \bar{W}_n^k(\delta)} (xy)^{-\tau} \min \left(1, \frac{kx}{\mu n} \right) \min \left(1, \frac{ky}{\mu n} \right) \min \left(1, \frac{xy}{\mu n} \right) dy dx. \end{aligned} \quad (5.6.5)$$

We analyze this expression separately for all three ranges of k . For ease of notation, we assume that $\mu = 1$ in the rest of this section.

We first consider Range I, where $k = o(n^{(\tau-2)/(\tau-1)})$. Then we have to show that the contribution from vertices u and v such that $D_u D_v < \delta n$ or $D_u D_v > n/\delta$ is small. First, we study the contribution to (5.6.5) for $D_u D_v < \delta n$. We bound this contribution by taking the second term of the minimum in all three cases, which gives

$$\frac{k^2}{n^3} \int_1^n \int_1^{\delta n/x} (xy)^{2-\tau} dy dx = \frac{k^2}{n^3} \int_1^n \frac{1}{x} \int_x^{\delta n} u^{2-\tau} du dx = \frac{k^2 \delta^{3-\tau}}{3-\tau} O(n^{-\tau} \log(n)). \quad (5.6.6)$$

Then, we study the contribution for $D_u D_v > n/\delta$. This contribution can be bounded very similarly by taking kD_u/L_n and $kD_u v/L_n$ and 1 for the minima in (5.6.5) as

$$\frac{nk^2}{n^2} \int_1^n \int_{n/(\delta x)}^n (xy)^{1-\tau} dy dx = \frac{k^2}{n^2} \int_1^n \frac{1}{x} \int_{n/\delta}^{nx} u^{1-\tau} du dx = \frac{k^2 \delta^{\tau-2}}{\tau-2} O(n^{-\tau} \log(n)). \quad (5.6.7)$$

Thus, by (5.6.4),

$$\mathbb{E}_n [c_{\varepsilon_n}(\bar{W}_n^k(\delta))] = O_{\mathbb{P}} \left(n^{2-\tau} \log(n) \delta^k \right). \quad (5.6.8)$$

Dividing by $n^{2-\tau} \log(n)$ and letting $n \rightarrow \infty$ then proves the lemma in Range I.

Now we consider Range II, where $k = \Omega(n^{(\tau-2)/(\tau-1)})$ and $k = o(\sqrt{n})$. We show that the contribution from vertices u and v such that $D_u D_v < \delta n$ or $D_u D_v > n/\delta$ or $D_u, D_v > n/(k\delta)$ is small. We first show that the contribution to (5.6.5) for $D_u > n/(k\delta)$ is small. In this setting, $D_u k > n$, so that the first minimum in (5.6.5) is attained by 1. The contribution can be computed as

$$\begin{aligned} & \int_{n/(k\delta)}^{\infty} \int_1^{\infty} (xy)^{-\tau} \min \left(1, \frac{ky}{n} \right) \min \left(1, \frac{xy}{n} \right) dy dx \\ &= \frac{k}{n^2} \int_{n/(\delta k)}^{\infty} \int_1^{n/x} x^{1-\tau} y^{2-\tau} dy dx + \frac{k}{n} \int_{n/(k\delta)}^{\infty} \int_{n/x}^{n/k} x^{-\tau} y^{1-\tau} dy dx \\ &\quad + \int_{n/(k\delta)}^{\infty} \int_{n/k}^{\infty} x^{-\tau} y^{-\tau} dy dx \\ &= k^2 O(n^{-\tau}) + k^2 O(n^{-\tau}) + \delta^{\tau-1} O(n^{2-2\tau} k^{2\tau-2}). \end{aligned} \quad (5.6.9)$$

By (5.6.4), multiplying by n^2k^{-2} and then dividing by $f(n, k) = n^{2-\tau} \log(n/k^2)$ and letting $n \rightarrow \infty$ shows that this contribution is small. Thus, we may assume that $D_u, D_v < n/(k\delta)$. Now we show that the contribution from $D_u D_v < \delta n$ is negligible. Then, $D_u D_v < n$, so that the third minimum in (5.6.5) is attained for $D_u D_v/n$. The contribution then splits into various cases, depending on D_u .

$$\begin{aligned}
& \frac{1}{n} \int \int_{xy < \delta n} (xy)^{1-\tau} \min\left(1, \frac{kx}{n}\right) \min\left(1, \frac{ky}{n}\right) dy dx \\
&= \int_1^k \int_1^{\delta n/x} (xy)^{-\tau} \frac{kx^2 y}{L_n^2} dy dx + \int_k^{n/k} \int_1^{\delta n/x} (xy)^{-\tau} \frac{k^2 x^2 y^2}{L_n^3} dy dx \\
&\quad + \int_{n/k}^\infty \int_1^{\delta n/x} (xy)^{-\tau} \frac{kxy^2}{L_n^3} dy dx \\
&= k^2 O(n^{-\tau}) \delta^{2-\tau} + k^2 \delta n^{-\tau} O(\log(n/k^2)) + k^2 O(n^{-\tau}) \delta^{3-\tau}. \tag{5.6.10}
\end{aligned}$$

The contribution of $D_u D_v > n/\delta$ can be bounded similarly as

$$\begin{aligned}
& \int \int_{xy > n/\delta} (xy)^{-\tau} \min\left(1, \frac{kx}{n}\right) \min\left(1, \frac{ky}{n}\right) dy dx \\
&= \int_1^k \int_{n/(\delta x)}^\infty (xy)^{-\tau} \frac{kx}{L_n} dy dx + \int_k^{n/k} \int_{n/(\delta x)}^\infty (xy)^{-\tau} \frac{k^2 xy}{L_n^2} dy dx \\
&\quad + \int_{n/k}^\infty \int_{n/(\delta x)}^\infty (xy)^{-\tau} \frac{ky}{L_n} dy dx \\
&= k^2 \delta^{\tau-1} O(n^{-\tau}) + k^2 \delta^{\tau-2} O(n^{-\tau} \log(n/k^2)) + k^2 O(n^{-\tau}) \delta^{\tau-2}. \tag{5.6.11}
\end{aligned}$$

By (5.6.4), multiplying by $k^{-2}n^2$ and then dividing by $f(n, k) = n^{2-\tau} \log(n/k^2)$ proves the lemma in Range II.

Finally, we prove the lemma in Range III, where $k = \Omega(\sqrt{n})$. Here we have to show that the contribution from $D_u, D_v < \delta n/k$ or $D_u, D_v > n/(\delta k)$ is small. We again bound this using (5.6.5). The contribution to (5.6.5) for $D_u > n/(k\delta)$ can be computed as

$$\begin{aligned}
& \int_{n/(k\delta)}^\infty \int_1^\infty (xy)^{-\tau} \min\left(1, \frac{ky}{n}\right) \min\left(1, \frac{xy}{n}\right) dy dx \\
&= \int_{\frac{n}{k\delta}}^k \int_{n/x}^\infty x^{-\tau} y^{-\tau} dy dx + \int_{\frac{n}{k\delta}}^k \int_{n/k}^{n/x} \frac{1}{n} x^{-\tau+1} y^{-\tau+1} dy dx \\
&\quad + \int_{\frac{n}{k\delta}}^k \int_0^{n/k} \frac{k}{n^2} x^{-\tau+1} y^{-\tau+2} dy dx + \int_k^\infty \int_{n/k}^\infty x^{-\tau} y^{-\tau} dy dx \\
&\quad + \int_k^\infty \int_{n/x}^{n/k} \frac{k}{n} x^{-\tau} y^{-\tau+1} dy dx + \int_k^\infty \int_0^{n/x} \frac{k}{n^2} x^{-\tau+1} y^{-\tau+2} dy dx \\
&= O\left(\log\left(\frac{k^2 \delta}{n}\right) n^{1-\tau}\right) + O\left(\delta^{\tau-2} k^{2\tau-4} n^{3-2\tau}\right) + O\left(n^{1-\tau}\right) \\
&\quad + O\left(\delta^{\tau-2} n^{3-2\tau} k^{2\tau-4}\right) + O\left(n^{1-\tau}\right) + O\left(n^{1-\tau}\right) + O\left(n^{1-\tau}\right) \\
&= O\left(\delta^{\tau-2} k^{2\tau-4} n^{3-2\tau}\right). \tag{5.6.12}
\end{aligned}$$

Multiplying this by n^2k^{-2} and then dividing by $f(n, k) = n^{5-2\tau}k^{2\tau-6}$ shows that this contribution is sufficiently small.

Then we study the contribution to (5.6.5) for $D_u < \delta n/k$. This can be computed as

$$\begin{aligned}
& \frac{1}{n} \int_1^{\delta n/k} \int_1^\infty (xy)^{1-\tau} \min\left(1, \frac{ky}{n}\right) \min\left(1, \frac{xy}{n}\right) dy dx \\
&= \int_0^{\frac{n\delta}{k}} \int_0^{n/k} \frac{k^2}{n^3} x^{-\tau+2} y^{-\tau+2} dy dx + \int_0^{\frac{n\delta}{k}} \int_{n/k}^{n/x} \frac{k}{n^2} x^{-\tau+2} y^{-\tau+1} dy dx \\
&\quad + \int_0^{\frac{n\delta}{k}} \int_{n/x}^\infty \frac{k}{n} x^{-\tau+1} y^{-\tau} dy dx \\
&= O\left(\delta^{3-\tau} k^{2\tau-4} n^{3-2\tau}\right) + O\left(\delta^{3-\tau} k^{2\tau-4} n^{3-2\tau}\right) + O\left(\delta n^{1-\tau}\right) \\
&= O\left(\delta^{\tau-2} k^{2\tau-4} n^{3-2\tau}\right). \tag{5.6.13}
\end{aligned}$$

Again, multiplying by n^2k^{-2} and dividing these estimates by $f(n, k) = n^{5-2\tau}k^{2\tau-6}$ completes the proof in Range III. \square

5.6.1 Proof of Theorem 5.2

We now show how we adjust the proof of Theorem 5.1 to prove Theorem 5.2. We use the same major contributing triangles as the ones in Range III in (5.3.6). Then, in fact Lemmas 5.3, 5.9 and Proposition 5.2 still hold. It is easy to derive a similar lemma as Lemma 5.8 for the situation $k = \Theta(\sqrt{n})$. The only difference with the proof of Lemma 5.8 is that we do not Taylor expand the exponentials in (5.5.52). This then proves Theorem 5.2. \square

5.6.2 Proof of Theorem 5.3

We now prove that the scaling limit of $k \mapsto c_{\varepsilon_n}(k)$ is continuous around $k = \sqrt{n}$. When B is large, we rewrite (5.3.4) as

$$\frac{c_{\varepsilon_n}(k)}{n^{2-\tau}} \xrightarrow{\mathbb{P}} C^2 \mu^{2-2\tau} B^{2\tau-4} \int_0^\infty \int_0^\infty (xy)^{-\tau} (1 - e^{-x})(1 - e^{-y})(1 - e^{-xy\mu/B^2}) dx dy. \tag{5.6.14}$$

Taylor expanding the last exponential then yields

$$\begin{aligned}
\frac{c_{\varepsilon_n}(k)}{n^{2-\tau}} &= (1 + o(1)) C^2 \mu^{3-2\tau} B^{2\tau-6} \int_0^\infty \int_0^\infty (xy)^{1-\tau} (1 - e^{-x})(1 - e^{-y}) dx dy \\
&= (1 + o(1)) C^2 \mu^{3-2\tau} B^{2\tau-6} A^2. \tag{5.6.15}
\end{aligned}$$

Substituting $k = B\sqrt{n}$ in Range III of Theorem 5.1 gives

$$\frac{c_{\varepsilon_n}(k)}{n^{2-\tau}} = (1 + o(1)) C^2 \mu^{3-2\tau} B^{2\tau-6} A^2, \tag{5.6.16}$$

which is the same as the result obtained from Theorem 5.2. Therefore, the scaling limit of $k \mapsto c_{\varepsilon_n}(k)$ is smooth for $k > \sqrt{n}$.

For B small, we can Taylor expand the first two exponentials in (5.3.4) as long as x and y are much smaller than $1/B$. The contribution where $x, y < 1/B$ and $B < \mu xy < 1/B$ can be written as

$$\begin{aligned}
& C^2 \mu^{2-2\tau} \left(\int_{B^2}^1 \int_{B/(\mu x)}^{1/B} (xy)^{1-\tau} (1 - e^{-\mu xy}) dy dx \right. \\
& \quad \left. + \int_1^{1/B} \int_{B/(\mu x)}^{1/(Bx)} (xy)^{1-\tau} (1 - e^{-\mu xy}) dy dx \right) \\
& = C^2 \mu^{-\tau} \left(\int_{B^2}^1 \int_B^{v/B} \frac{1}{v} u^{1-\tau} (1 - e^{-u}) du dv + \int_1^{1/B} \int_B^{1/B} \frac{1}{v} u^{1-\tau} (1 - e^{-u}) du dv \right) \\
& = C^2 \mu^{-\tau} \left(\log(B^{-2}) \int_B^{1/B} u^{1-\tau} (1 - e^{-u}) du + \int_B^{1/B} \log(1/u) u^{1-\tau} (1 - e^{-u}) du \right), \tag{5.6.17}
\end{aligned}$$

where we used the change of variables $u = \mu xy$ and $v = x$. The contribution of the second integral becomes small compared to the first part as B gets small, as the second integral is finite for $B > 0$. We can show that the contributions from $x, y > 1/B$, or from $xy > 1/B$ can also be neglected by using that $1 - e^{-x} \leq \min(1, x)$. Thus, as B becomes very small, Theorem 5.2 shows that $c_{\varepsilon_n}(k)$ for $k = B\sqrt{n}$ can be approximated by

$$\frac{c_{\varepsilon_n}(k)}{n^{2-\tau}} \approx C^2 \log(B^{-2}) \int_0^\infty u^{1-\tau} (1 - e^{-u}) du, \tag{5.6.18}$$

which agrees with the value for $k = B\sqrt{n}$ in Range II of Theorem 5.1.

To prove the continuity around $k = n^{(\tau-2)/(\tau-1)}$, we note that the proofs of Lemmas 5.7, 5.9 and 5.10 for Range II still hold if we assume that $k \geq an^{(\tau-2)/(\tau-1)}$ for some $a > 0$ instead of $k = \Omega(n^{(\tau-2)/(\tau-1)})$. Thus, we can also apply the result of Range II in Theorem 5.1 to $k = an^{(\tau-2)/(\tau-1)}$, which yields

$$c_{\varepsilon_n}(k) = n^{2-\tau} \mu^{-\tau} C^2 A \left(\frac{3-\tau}{\tau-1} \log(n) + \log(a^{-2}) \right) (1 + o_p(1)). \tag{5.6.19}$$

This agrees with the $k \mapsto c_{\varepsilon_n}(k)$ curve in Range I when n grows large. \square

5.7 Clustering in uniform random graphs

In this section, we estimate the connection probability in a uniform random graph, which is the key ingredient for proving Theorem 5.1. We say that a sequence of degree sequences $(D_i)_{i \in [n]}$ satisfies the pseudo-power law with parameter $\tau > 1$, if there exists a constant $K > 0$ such that for every $n \geq 1$, and every $i \geq 1$, $|\{v : D_v \geq i\}| \leq Kn i^{1-\tau}$. If $(D_i)_{i \in [n]}$ is an i.i.d. sample of (5.2.1), then with high probability $(D_i)_{i \in [n]}$ satisfies the pseudo-power law with parameter τ' provided $\tau' < \tau$ [88]. Our key lemma is as follows, where the probability space refers to the uniformly random simple graphs with degree sequence $(D_i)_{i \in [n]}$:

Lemma 5.11. *Assume that $2 < \tau < 3$ is fixed and $(D_i)_{i \in [n]}$ satisfies the pseudo-power law with parameter $2 < \tau < 3$. Let C denote a set of unordered pairs of vertices and let E_C denote*

the event that xy is an edge for every $\{x, y\} \in C$. Then, on the event J_n defined in (5.4.8) and assuming that $|C| = O(1)$ and $\{u, v\} \notin C$,

$$\mathbb{P}_n(X_{uv} = 1 \mid E_C) = (1 + o(1)) \frac{(D_u - |C_u|)(D_v - |C_v|)}{L_n + (D_u - |C_u|)(D_v - |C_v|)},$$

where C_x denote the set of pairs in C that contain x .

To prove Theorem 5.5, we use the following corollary of Lemma 5.11:

Corollary 5.1. *Let C denote a set of unordered pairs of vertices, such that $|C| \leq 5$ and $\{u, v\} \notin C$. Let E_C denote the event that xy is an edge for every $\{x, y\} \in C$. Then, on the event J_n , the connection probability of vertices u and v in a uniform random graph on degree sequence $(D_i)_{i \in [n]}$ satisfying the pseudo-power law with $2 < \tau < 3$ satisfies*

(i) For all u, v ,

$$\mathbb{P}_n(X_{uv} = 1 \mid E_C) = O\left(\frac{D_u D_v}{L_n}\right). \quad (5.7.1)$$

(ii) For all u, v such that $D_u, D_v \gg 1$ and $D_u D_v = O(n)$

$$\mathbb{P}_n(X_{uv} = 1 \mid E_C) = (1 + o(1)) \frac{D_u D_v}{L_n + D_u D_v}. \quad (5.7.2)$$

We now proceed to prove Lemma 5.11. As a preparation, we first prove a lemma about the number of 2-paths starting from a specified vertex.

Lemma 5.12. *Assume that $2 < \tau < 3$ is fixed and $(D_i)_{i \in [n]}$ satisfies the pseudo-power law with parameter $2 < \tau < 3$. For any graph G whose degree sequence is $(D_i)_{i \in [n]}$, the number of 2-paths starting from any specified vertex is $o(n)$.*

Proof. W.l.o.g. we assume that $D_1 \geq D_2 \geq \dots \geq D_n$. For every $1 \leq i \leq n$, the number of vertices with degree at least D_i is at least i . By the definition of the pseudo-power law, $KnD_i^{1-\tau} \geq i$ for every $1 \leq i \leq n$. It then follows that $D_i \leq (Kn/i)^{1/(\tau-1)}$. Then the number of 2-paths from any specified vertex is bounded by $\sum_{i=1}^{D_1} D_i$, which is at most

$$\sum_{i=1}^{D_1} \left(\frac{Kn}{i}\right)^{\frac{1}{\tau-1}} = (Kn)^{\frac{1}{\tau-1}} \sum_{i=1}^{D_1} i^{-\frac{1}{\tau-1}} = O(n^{\frac{1}{\tau-1}}) D_1^{\frac{\tau-2}{\tau-1}} = O(n^{\frac{2\tau-3}{(\tau-1)^2}}),$$

since $D_1 \leq (Kn)^{1/(\tau-1)}$. Since $2 < \tau < 3$ the above is $o(n)$. □

Proof of Lemma 5.11. To estimate $\mathbb{P}_n(X_{uv} = 1 \mid E_C)$, we will switch between two classes of graphs: S and \bar{S} . S consists of graphs with degree sequence $(D_i)_{i \in [n]}$ where all edges in $\{u, v\} \cup C$ are present, whereas \bar{S} consists of all graphs with degree sequence $(D_i)_{i \in [n]}$ where all edges in C are present, whereas $\{u, v\}$ is not an edge. Since we sample uniformly from all graphs with degree sequence $(D_i)_{i \in [n]}$:

$$\mathbb{P}_n(X_{uv} = 1 \mid E_C) = \frac{|S|}{|S| + |\bar{S}|} = \frac{1}{1 + |\bar{S}|/|S|}. \quad (5.7.3)$$

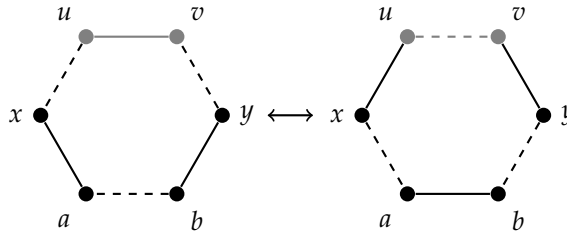


Figure 5.8: The right arrow illustrates a forward switching and the left arrow a backward switching.

In order to estimate the ratio $|\bar{S}|/|S|$, we will define an operation called a *forward switching* which converts a graph in $G \in S$ to a graph $G' \in \bar{S}$. The reverse operation converting G' to G is called a *backward switching*. Then we estimate $|\bar{S}|/|S|$ by counting the number of forward switchings that can be applied to a graph $G \in S$, and the number of backward switchings that can be applied to a graph $G' \in \bar{S}$.

The forward switching is defined by choosing two edges and specifying their ends as $\{x, a\}$ and $\{y, b\}$. The choice must satisfy the following constraints:

1. None of $\{u, x\}$, $\{v, y\}$, or $\{a, b\}$ is an edge;
2. $\{x, a\}, \{y, b\} \notin C$;
3. All of u, v, x, y, a , and b must be distinct except that $x = y$ is permitted.

Given a valid choice, the forward switching replaces the three edges $\{u, v\}$, $\{x, a\}$, and $\{y, b\}$ by $\{u, x\}$, $\{v, y\}$, and $\{a, b\}$. The forward switching preserves the degree sequence, and converts a graph in S to a graph in \bar{S} . The inverse operation of a forward switching is called a backward switching. See Figure 5.8 for an illustration.

Next, we estimate the number of ways to perform a forward switching to a graph G in S , denoted by $f(G)$, and the number of ways to perform a backward switching to a graph G' in \bar{S} , denoted by $b(G')$. Now, the number of total switchings between S and \bar{S} is equal to $|S|\mathbb{E}[f(G)] = |\bar{S}|\mathbb{E}[b(G')]$, where the expectation is over a uniformly random $G \in S$ and $G' \in \bar{S}$ respectively. Consequently,

$$\frac{|\bar{S}|}{|S|} = \frac{\mathbb{E}[f(G)]}{\mathbb{E}[b(G')]} \quad (5.7.4)$$

Given an arbitrary graph $G \in S$, the number of ways of carrying out a forward switching is at most L_n^2 , since there are at most L_n ways to choose $\{x, a\}$, and at most L_n ways to choose $\{y, b\}$. To find a lower bound on the number of ways of performing a forward switching, we subtract from L_n^2 an upper bound on the number of invalid choices for $\{x, a\}$ and $\{y, b\}$. These can be summarized as follows:

- (a) at least one of $\{u, x\}, \{a, b\}, \{v, y\}$ is an edge,
- (b) $\{x, a\} \in C$ or $\{y, b\} \in C$,

- (c) any vertex overlap other than $x = y$ (i.e. if one of a or b is equal to one of x or y , or if $a = b$, or if one of u or v are one of $\{a, b, x, y\}$).

To find an upper bound for (a), note that any choice in case (a) must involve a single edge, and a 2-path starting from a specified vertex. By Lemma 5.12, the number of choices for (a) then is upper bounded by $3 \cdot o(L_n) \cdot L_n = o(L_n^2)$. The number of choices for case (b) is $O(L_n)$ as $|C| = O(1)$, and there are at most L_n ways to choose the other edge which is not restricted to be in C . To bound the number of choices for (c), we investigate each case:

- (C1) a or b is equal to x or y ; or $a = b$. In this case, x, y, a, b forms a 2-path. Thus, there are at most $5 \cdot n \cdot o(L_n) = o(L_n^2)$ choices (noting that $n = O(L_n)$ on the event J_n), where n is the number of ways to choose a vertex, and $o(L_n)$ bounds the number of 2-paths starting from this specified vertex;
- (C2) one of u and v is one of $\{a, b, x, y\}$. In this case, there is one 2-path starting from u or v , and a single edge. Thus, there are at most $8 \cdot L_n D_{\max} = o(L_n^2)$ choices, where D_{\max} bounds the number of ways to choose a vertex adjacent to u or v and L_n bounds the number of ways to choose a single edge.

Thus, the number of invalid choices for $\{x, a\}$ and $\{y, b\}$ is $o(L_n^2)$, so that the number of forward switchings which can be applied to any $G \in S$ is $(1 + o(1))L_n^2$. Thus,

$$\mathbb{E} [f(G)] = L_n^2(1 + o(1)). \quad (5.7.5)$$

Given a graph $G' \in \bar{S}$, consider the backward switchings that can be applied to G' . There are at most $L_n(D_u - |C_u|)(D_v - |C_v|)$ ways to do the backward switching, since we are choosing an edge which is adjacent to u but not in C , an edge which is adjacent to v but not in C , and another "oriented" edge $\{a, b\}$ (oriented in the sense that each edge has two ways to specify its end vertices as a and b). For a lower bound, we consider the following forbidden choices:

- (a') at least one of $\{x, a\}$ or $\{y, b\}$ is an edge,
- (b') $\{a, b\} \in C$,
- (c') any vertices overlap other than $x = y$ (i.e. if $\{a, b\} \cap \{u, v, x, y\} \neq \emptyset$).

For (a'), suppose that $\{x, a\}$ is present. There are at most $(D_u - |C_u|)(D_v - |C_v|)$ ways to choose x and y . Given any choice for x and y , there are at most $o(L_n)$ ways to choose a 2-path starting from x , and hence $o(L_n)$ ways to choose a, b . Thus, the total number of choices is at most $o((D_u - |C_u|)(D_v - |C_v|)L_n)$. The case that $\{y, b\}$ is an edge is symmetric.

For (b'), there are $O(1)$ choices for choosing $\{a, b\}$ since $|C| = O(1)$, and at most $(D_u - |C_u|)(D_v - |C_v|)$ choices x and y . Thus, the number of choices for case (b') is $O((D_u - |C_u|)(D_v - |C_v|)) = o((D_u - |C_u|)(D_v - |C_v|)L_n)$.

For (c'), the case that a or b is equal to x or y corresponds to a 2-path starting from u or v together with a single edge from u or v . Since $o(L_n)$ bounds the number of 2-paths starting from u or v and $D_u - |C_u| + D_v - |C_v|$ bounds the number of ways to choose the single edge, there are $o(L_n(D_v - |C_v|)) + o(L_n(D_u - |C_u|))$ total

choices. If a or b is equal to u or v , there are $(D_u - |C_u|)(D_v - |C_v|)$ ways to choose x and y , and at most $D_u + D_v$ ways to choose the last vertex as a neighbor of u or v . Thus, there are $O((D_u - |C_u|)(D_v - |C_v|)D_{\max}) = o((D_u - |C_u|)(D_v - |C_v|)L_n)$ total choices, since $D_{\max} = O(n^{1/(\tau-1)}) = o(n) = o(L_n)$. This concludes that the number of backward switchings that can be applied to any graph $G' \in S'$ is $(D_u - |C_u|)(D_v - |C_v|)L_n(1 + o(1))$, so that also

$$\mathbb{E} [b(G')] = (D_u - |C_u|)(D_v - |C_v|)L_n(1 + o(1)). \quad (5.7.6)$$

Combining (5.7.4), (5.7.5) and (5.7.6) results in

$$|\bar{S}|/|S| = (1 + o(1)) \frac{L_n^2}{(D_u - |C_u|)(D_v - |C_v|)L_n},$$

and thus (5.7.3) yields

$$\mathbb{P}_n(X_{uv} = 1 \mid E_C) = \frac{1}{1 + |\bar{S}|/|S|} = (1 + o(1)) \frac{(D_u - |C_u|)(D_v - |C_v|)}{L_n + (D_u - |C_u|)(D_v - |C_v|)}.$$

□

We now show how we use Corollary 5.1 to adjust the proof of Theorem 5.1 to prove Theorem 5.5.

Proof of Theorem 5.5. Since by [88] the degree sequence satisfies the pseudo-power law with high probability, we assume that the sampled degree sequence satisfies the pseudo-power law assumption. Define

$$h_n(D_u, D_v, D_w) := \frac{D_u D_v}{D_u D_v + L_n} \frac{D_u D_w}{D_u D_w + L_n} \frac{D_v D_w}{D_v D_w + L_n}. \quad (5.7.7)$$

When $(u, v) \in W_n^k(\delta)$, $D_u, D_v \gg 1$ as well as $D_u D_v = O(n)$, $D_u k = O(n)$ and $D_v k = O(n)$ (see (5.3.6)). Thus, when $(u, v) \in W_n^k(\delta)$, we may use Corollary 5.1(ii) on the probabilities that the edges in a triangle between vertex u, v and a vertex of degree k are present. Then, similarly to the proof of Lemma 5.3, of the event

$$\mathbb{E}_n [c_{\varepsilon_n}(W_n^k(\delta))] = \frac{\sum_{(u,v) \in W_n^k(\delta)} h_n(k, D_u, D_v)}{k(k-1)} (1 + o(1)). \quad (5.7.8)$$

In Range I, when $(u, v) \in W_n^k(\delta)$, $D_u k / L_n = o(1)$ so that $D_u k / (D_u k + L_n) = D_u k / (1 + o(1))$ and similarly $D_v k / (D_v k + L_n) = D_v k / (1 + o(1))$. Thus, in Range I,

$$\mathbb{E}_n [c_{\varepsilon_n}(W_n^k(\delta))] = \sum_{(u,v) \in W_n^k(\delta)} \frac{1}{L_n^2} \frac{D_u^2 D_v^2}{D_u D_v + L_n} (1 + o(1)). \quad (5.7.9)$$

A similar convergence of measure argument as in Lemma 5.6 then shows that in Range I

$$\frac{\mathbb{E}_n [c_{\varepsilon_n}(W_n^k(\delta))]}{n^{2-\tau} \log(n)} \xrightarrow{\mathbb{P}} C^2 \mu^{-\tau} \frac{3-\tau}{\tau-1} \int_{\delta}^{1/\delta} \frac{t^{2-\tau}}{1+t} dt. \quad (5.7.10)$$

Similarly, in Range II we analyze (5.7.8) using Lemma 5.7, again replacing the function $1 - e^{-xy/L_n}$ by $xy/(L_n + xy)$, resulting in

$$\frac{\mathbb{E}_n \left[c_{\varepsilon_n}(W_n^k(\delta)) \right]}{n^{2-\tau} \log(n/k^2)} \xrightarrow{\mathbb{P}} C^2 \mu^{-\tau} \int_{\delta}^{1/\delta} \frac{t^{2-\tau}}{1+t} dt. \quad (5.7.11)$$

Finally, in Range III, when $(u, v) \in W_n^k(\delta)$, $D_u D_v = o(n)$ so that $D_u D_v / (D_u D_v + L_n) = L_n^{-1}(1 + o(1))$. Thus, (5.7.8) becomes

$$\mathbb{E}_n \left[c_{\varepsilon_n}(W_n^k(\delta)) \right] = \frac{1}{k^2 L_n} \sum_{(u,v) \in W_n^k(\delta)} \frac{D_u k}{D_u k + L_n} \frac{D_v k}{D_v k + L_n} (1 + o(1)). \quad (5.7.12)$$

Now applying Lemma 5.8, replacing the function $1 - e^{-xy/L_n}$ by $xy/(L_n + xy)$ shows that for Range III

$$\frac{\mathbb{E}_n \left[c_{\varepsilon_n}(W_n^k(\delta)) \right]}{n^{5-2\tau} k^{2\tau-6}} \xrightarrow{\mathbb{P}} C^2 \mu^{3-2\tau} \left(\int_{\delta}^{1/\delta} \frac{t^{2-\tau}}{1+t} dt \right)^2. \quad (5.7.13)$$

Furthermore, Lemma 5.9 also holds for the uniform random graph, again replacing the edge probability $1 - e^{-D_u D_v / L_n}$ by $D_u D_v / (L_n + D_u D_v) \leq \min(D_u D_v / L_n, 1)$.

Finally, Lemma 5.10 also holds for the uniform random graph. The proof follows the exact same lines as the proof of Lemma 5.10, only instead of using that $\mathbb{P}_n(X_{ij} = 1) \leq \min(D_i D_j / L_n, 1)$, we now use that by Corollary 5.1 $\mathbb{P}_n(X_{ij} = 1) = O(\min(D_i D_j / L_n, 1))$, which does not change the scaling of the terms involved in Lemma 5.10. Combining these lemmas in the same way as in the proof of Theorem 5.1 then proves Theorem 5.5. □

6 Local clustering in dynamic and spatial models

Based on:
Variational principle for random graphs explaining scale-free clustering
C. Stegehuis, R. van der Hofstad, J.S.H. van Leeuwen, arXiv:1812.03002

In this chapter, we introduce a variational principle to explain how vertices tend to cluster as a function of their degrees, generalizing the analysis of $c(k)$ for the rank-1 inhomogeneous random graph, the erased configuration model and the uniform random graph of Chapters 4 and 5. The variational principle reveals the triplets of vertices that dominate the triangle count. We show that this variational principle applies to a wide class of random graph models including the preferential attachment model, the inhomogeneous random graph, the random intersection graph and the hyperbolic random graph, enabling us to find the scaling of $c(k)$ for all these models.

6.1 Introduction

Chapters 4 and 5 show that the rank-1 inhomogeneous random graph, the erased configuration model and the uniform random graph have vanishing clustering levels when the network size grows to infinity, making these models unfit for modeling group formation in the large-network limit, which is often observed in real-world networks. We therefore study the hyperbolic model in this chapter, introduced in Section 1.1.6, which creates a random graph by positioning each vertex at a uniformly chosen location in the hyperbolic space, and then connecting pairs of vertices as a function of their locations. The hyperbolic model is mathematically tractable and capable of matching simultaneously the three key characteristics of real-world networks: sparseness, power-law degrees and clustering.

We again measure clustering in terms of the local clustering coefficient $c(k)$ introduced in Section 1.2.2. It is possible that no vertex of degree k is present, which can be resolved by averaging over degrees close to k , as in Chapter 5, which does not affect the scaling. In this chapter, we therefore analyze $c(k)$ assuming that at least 2 vertices of degree k are present for ease of notation. As Figure 4.2 shows, the $c(k)$ -curves of several observed real-world networks share several similarities. First of all, $c(k)$ decays in k . Furthermore, for small values of k , $c(k)$ is high, indicating the presence of non-trivial clustering. In this paper, we aim to design a mathematical method that, together with a mathematical network model, is able to describe such curves. Taking the hyperbolic model as the network model, we obtain a precise characterization of

clustering in the hyperbolic model by describing how the clustering curve $k \mapsto c(k)$ scales with k and n . We also obtain the scaling behavior for C from the results for $c(k)$.

Studying the local clustering coefficient $c(k)$ is equivalent to studying the number of triangles where at least one of the vertices has degree k . We develop a variational principle that finds the dominant such triangle in terms of the degrees of the other two vertices. This variational principle exploits the trade-off present in power-law networks: on the one hand, high-degree vertices are well connected and therefore participate in many triangles. On the other hand, these vertices are rare because of the power-law degree distribution. Lower-degree vertices typically participate in fewer triangles, but occur more frequently. The variational principle finds the degrees that optimize this trade-off and reveals the structure of the three-point correlations that dictate the degree of clustering.

The variational principle can be applied to a wide range of random graph models. While the focus lies with the hyperbolic model, we also apply the variational principle to the rank-1 inhomogeneous random graph, the preferential attachment model and the random intersection graph. As it turns out, each random graph models comes with a model-specific clustering curve $k \mapsto c(k)$. Our analysis confirms that the preferential attachment model and rank-1 inhomogeneous random graph lead to degenerate clustering curves that vanish in the large-network limit, while the random intersection graph displays a clustering curve that is more similar to the curve of the hyperbolic random graph.

6.2 Variational principle

We now discuss the variational principle in more detail, apply it to characterize clustering in the hyperbolic model, and then apply it to other random graph models.

The variational principle deals with the probability of creating a triangle between a vertex of degree k and two other uniformly chosen vertices, which can be written as

$$\mathbb{P}(\Delta_k) = \sum_{(d_1, d_2)} \mathbb{P}(\Delta \text{ on degrees } k, d_1, d_2) \mathbb{P}(d_1, d_2), \quad (6.2.1)$$

where the sum is over all possible pairs of degrees (d_1, d_2) , and $\mathbb{P}(d_1, d_2)$ denotes the probability that two uniformly chosen vertices have degrees d_1 and d_2 . We then let the degrees d_1 and d_2 scale as n^{α_1} and n^{α_2} and find which degrees give the largest contribution to (6.2.1). Due to the power-law degree distribution, the probability that a vertex has degree proportional to n^α scales as $n^{-(\tau-1)\alpha}$. The maximal summand of (6.2.1) can then be written as

$$\max_{\alpha_1, \alpha_2} \mathbb{P}(\Delta \text{ on degrees } k, n^{\alpha_1}, n^{\alpha_2}) n^{2+(\alpha_1+\alpha_2)(1-\tau)}. \quad (6.2.2)$$

We now study the optimal structure of a triangle using (6.2.2) for several random graph null models. If the optimizer over α_1 and α_2 is unique, and attained by α_1^* and α_2^* , then we can write the probability that a triangle is present between a vertex of degree k and two randomly chosen vertices as

$$\mathbb{P}(\Delta_k) \propto \mathbb{P}(\Delta \text{ on degrees } k, n^{\alpha_1^*}, n^{\alpha_2^*}) n^{(\alpha_1^*+\alpha_2^*)(1-\tau)}. \quad (6.2.3)$$

The local clustering coefficient $c(k)$ is defined as the expected number of triangles containing a uniformly chosen vertex of degree k divided by $\binom{k}{2}$. Therefore,

$$c(k) \propto n^2 k^{-2} \mathbb{P}(\Delta \text{ on degrees } k, n^{\alpha_1^*}, n^{\alpha_2^*}) n^{(\alpha_1^* + \alpha_2^*)(1-\tau)}. \quad (6.2.4)$$

Thus, if we know the probability that a triangle is present between vertices of degrees k, n^{α_1} and n^{α_2} for some random graph null model, the variational principle is able to find the scaling of $c(k)$ in k as well as the graph size n .

Suppose a model assigns to each vertex some parameters that determine the connection probabilities (for example radial and angular coordinates in case of the hyperbolic random graph). The variational principle can then be applied as long as the vertex degree can be expressed as some function of the vertex parameters, so that the probability of triangle formation between three vertices can be viewed as a function of the vertex degrees, and one can search for the optimal contribution to (6.2.2). We show that this is possible for a wide class of random graph models, making the variational principle an important tool to investigate clustering.

6.2.1 Clustering in the hyperbolic random graph

We now use the variational principle to compute $c(k)$ in the hyperbolic random graph, defined in Section 1.1.6, where each vertex is described by a radial coordinate r and an angular coordinate ϕ . As explained in Section 1.1.6, vertices with small radial coordinates are often hubs, whereas vertices with larger radial coordinates usually have small degrees. We will use this relation between radial coordinate and degree to find the most likely triangle in the hyperbolic model in terms of degrees as well as radial coordinates. For a point i with radial coordinate r_i , we define its type t_i as

$$t_i = e^{(R-r_i)/2}. \quad (6.2.5)$$

Then, if D_i denotes the degree of vertex i , by (2.4.16)

$$t_i = \Theta(D_i). \quad (6.2.6)$$

Furthermore, t_i are distributed as a power law with exponent τ [33], so that the degrees have a power-law distribution as well. The t_i 's can be interpreted as the weights in a rank-1 inhomogeneous random graph [33].

We now study $c(k)$ for the hyperbolic random graph, using the constrained variational principle (6.2.2). Thus, we calculate the probability that a triangle is present between vertices of degrees k, n^{α_1} and n^{α_2} . Because the degrees and the types of vertices have the same scaling, we investigate the probability that two neighbors of a vertex of type k connect. We compute the probability that a triangle is formed between a vertex of degree k , a vertex i with $t_i \propto n^{\alpha_1}$ and a vertex j with $t_j \propto n^{\alpha_2}$ with $\alpha_1 \leq \alpha_2$. We can write this probability as

$$\begin{aligned} \mathbb{P}(\Delta \text{ on types } k, n^{\alpha_1}, n^{\alpha_2}) &= \mathbb{P}(k \leftrightarrow n^{\alpha_1}) \mathbb{P}(k \leftrightarrow n^{\alpha_2}) \\ &\quad \times \mathbb{P}(n^{\alpha_1} \text{ and } n^{\alpha_2} \text{ neighbor connect}). \end{aligned} \quad (6.2.7)$$

The probability that two vertices with types t_i and t_j connect satisfies by Lemma 2.5

$$\mathbb{P}(i \leftrightarrow j \mid t_i, t_j) = g(\nu t_i t_j / n) (1 + o_{\mathbb{P}}(1)) \propto \min(2\nu t_i t_j / (\pi n), 1), \quad (6.2.8)$$

with g as in Lemma 2.5. Therefore, the probability that a vertex of type k connects with a randomly chosen vertex of type n^{α_1} can be approximated by

$$\mathbb{P}(k \leftrightarrow n^{\alpha_1}) \propto \min(kn^{\alpha_1-1}, 1). \quad (6.2.9)$$

The third term in (6.2.7) equals the probability that the two neighbors of a vertex of degree k connect to one another, which is more involved. Two neighbors of a vertex are likely to be close to one another, which increases the probability that they connect.

We now compute the order of magnitude of the third term in (6.2.7). Let i and j be neighbors of a vertex with degree k , with $t_i \propto n^{\alpha_1}$ and $t_j \propto n^{\alpha_2}$. Two vertices with types t_i and t_j and angular coordinates ϕ_i and ϕ_j connect if the relative angle between ϕ_i and ϕ_j , $\Delta\theta$, satisfies by (2.4.13)

$$\Delta\theta \leq \Theta(2vt_it_j/n). \quad (6.2.10)$$

W.l.o.g., let the angular coordinate of the vertex with degree k be 0. For i and j to be connected to a vertex with angular coordinate zero, by (6.2.10) ϕ_i and ϕ_j must satisfy

$$\begin{aligned} -\Theta(\min(kn^{\alpha_1-1}, 1)) &\leq \phi_i \leq \Theta(\min(kn^{\alpha_1-1}, 1)), \\ -\Theta(\min(kn^{\alpha_2-1}, 1)) &\leq \phi_j \leq \Theta(\min(kn^{\alpha_2-1}, 1)). \end{aligned} \quad (6.2.11)$$

Because the angular coordinates in the hyperbolic random graph are uniformly distributed, ϕ_i and ϕ_j are uniformly distributed in the above ranges. By (6.2.10), vertices i and j are connected if their relative angle is at most

$$2vn^{\alpha_1+\alpha_2-1}. \quad (6.2.12)$$

Thus, the probability that i and j connect is the probability that two randomly chosen points in the intervals (6.2.11) differ in their angles by at most (6.2.12). Assume that $\alpha_2 \geq \alpha_1$. Then, the probability that i and j are connected is proportional to

$$\begin{aligned} \mathbb{P}(n^{\alpha_1} \text{ and } n^{\alpha_2} \text{ neighbor connect}) &\propto \min\left(\frac{n^{\alpha_1+\alpha_2-1}}{\min(kn^{\alpha_2-1}, 1)}, 1\right) \\ &= \min(n^{\alpha_1} \max(n^{\alpha_2-1}, k^{-1}), 1). \end{aligned} \quad (6.2.13)$$

Thus, (6.2.2) reduces to

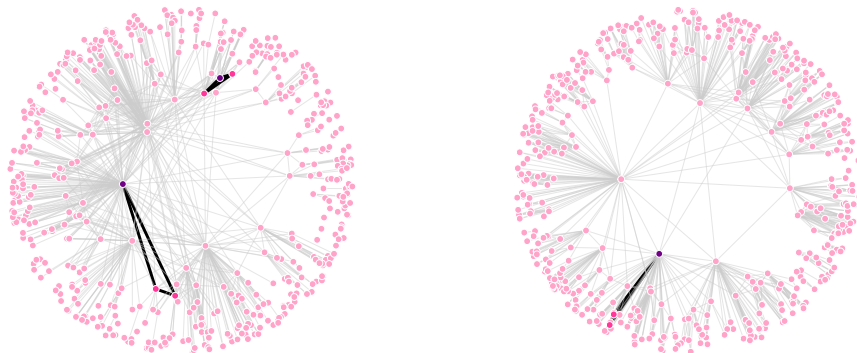
$$\max_{\alpha_1, \alpha_2} n^{(\alpha_1+\alpha_2)(1-\tau)} \min(kn^{\alpha_1-1}, 1) \min(kn^{\alpha_2-1}, 1) \min(n^{\alpha_1} \max(n^{\alpha_2-1}, k^{-1}), 1). \quad (6.2.14)$$

Because of the $\min(kn^{\alpha_2-1}, 1)$ term, it is never optimal to let the max term be attained by n^{α_2-1} . Thus, the equation reduces further to

$$\max_{\alpha_1, \alpha_2} n^{(\alpha_1+\alpha_2)(1-\tau)} \min(kn^{\alpha_1-1}, 1) \min(kn^{\alpha_2-1}, 1) \min(n^{\alpha_1} k^{-1}, 1). \quad (6.2.15)$$

The maximizers over $\alpha_1 \leq \alpha_2$ are given by

$$(n^{\alpha_1}, n^{\alpha_2}) \propto \begin{cases} (n^0, n^0), & \tau > \frac{5}{2}, \\ (k, k) & \tau < \frac{5}{2}, k \ll \sqrt{n}, \\ (n/k, n/k) & \tau < \frac{5}{2}, k \gg \sqrt{n}. \end{cases} \quad (6.2.16)$$



(a) $\tau < 5/2$: For $k \gg \sqrt{n}$ the two other vertices have degree proportional to n/k , whereas for $k \ll \sqrt{n}$ the other two vertices have degree proportional to k .
 (b) $\tau > 5/2$: The other two vertices have constant degree across the entire range of k .

Figure 6.1: Typical triangles containing a vertex of degree k (dark red) in hyperbolic random graphs. A vertex of degree n^α has radial coordinate close to $R - \alpha$, so that the optimal triangle degrees can be translated back to their radial coordinates in the disk.

Then (6.2.4) shows that for each $1 \ll k \ll n^{1/(\tau-1)}$, with high probability

$$c(k) \propto \begin{cases} k^{-1} & \tau > \frac{5}{2}, \\ k^{4-2\tau} & \tau < \frac{5}{2}, k \ll \sqrt{n}, \\ k^{2\tau-6} n^{5-2\tau} & \tau < \frac{5}{2}, k \gg \sqrt{n}. \end{cases} \quad (6.2.17)$$

This result is more detailed than the result in [130], where the scaling $c(k) \sim k^{-1}$ was predicted for fixed k . We find that this scaling only holds for the larger values of τ , while for $\tau < 5/2$ the decay of the curve is significantly different, which was also found in [82]. For $\tau > 5/2$, the $c(k)$ curve does not depend on n . For $\tau < 5/2$ the dependence on n is only present for large values of k . Interestingly, the exponent $\tau = 5/2$ is also the point where the maximal contribution to a bidirectional shortest path in the hyperbolic random graph changes from high-degree to lower-degree vertices [28]. The optimal triangle structures also contain higher vertex degrees for $\tau < 2.5$ than for $\tau > 2.5$ (see Figure 6.1). Figure 6.2a shows simulations of $c(k)$ in the hyperbolic random graph together with the asymptotic slope from (6.2.17).

6.2.2 Locally tree-like random graph models.

We next apply the variational principle to several random graph models that are known to be locally tree-like, so that triangles and clustering disappear in the large-network limit. The rank-1 inhomogeneous random graph is one such model. We analyzed $c(k)$ for the rank-1 inhomogeneous random graph in detail in Chapter 4, but we now show how the variational principle gives another method to obtain the scaling of $c(k)$ in the rank-1 inhomogeneous random graph. We take the connection

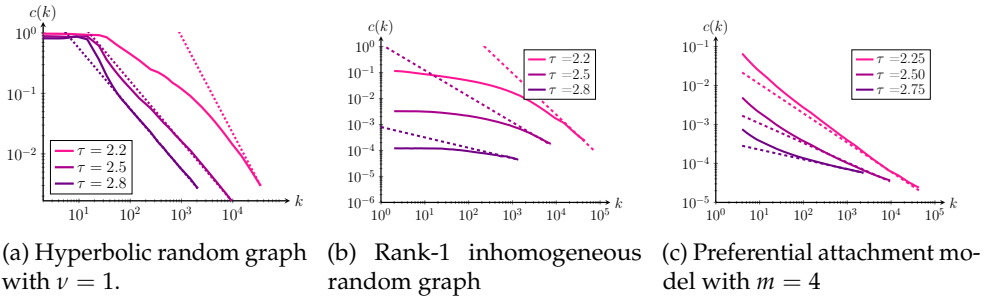


Figure 6.2: Simulations of $c(k)$ for three different models with $n = 10^6$. The solid lines correspond to averages over 10^4 network realizations and the dashed lines indicate the asymptotic slopes of (6.2.17), (6.2.22) and (6.2.28).

probability of vertices i and j with weights h_i and h_j to be given by

$$p(h_i, h_j) = \min(h_i h_j / (\mu n), 1), \quad (6.2.18)$$

where μ denotes the average weight. Thus, the probability that a vertex of degree k forms a triangle together with vertices i and j of degrees n^{α_1} and n^{α_2} , respectively, can be written as

$$\mathbb{P}(\Delta_{i,j,k}) = \Theta(\min(kn^{\alpha_1-1}, 1) \min(kn^{\alpha_2-1}, 1) \min(n^{\alpha_1+\alpha_2-1}, 1)). \quad (6.2.19)$$

Therefore (6.2.2) reduces to

$$\max_{\alpha_1, \alpha_2} n^{(\alpha_1+\alpha_2)(1-\tau)} \min(kn^{\alpha_1-1}, 1) \min(kn^{\alpha_2-1}, 1) \min(n^{\alpha_1+\alpha_2-1}, 1). \quad (6.2.20)$$

Calculating the optimum of (6.2.20) over $\alpha_1, \alpha_2 \in [0, 1/(\tau-1)]$ shows that the maximal contribution to the typical number of constrained triangles is given by

$$\begin{aligned} \alpha_1 + \alpha_2 &= 1, & k &\ll n^{(\tau-2)/(\tau-1)}, \\ \alpha_1 + \alpha_2 &= 1, n^{\alpha_1}, n^{\alpha_2} < n/k, & n^{(\tau-2)/(\tau-1)} &\ll k \ll \sqrt{n}, \\ n^{\alpha_1} &= n/k, n^{\alpha_2} = n/k, & k &\gg \sqrt{n}. \end{aligned} \quad (6.2.21)$$

Thus, for every value of k there exists an optimal constrained triangle, visualized in Figure 6.3. These three ranges of optimal triangle structures result in three ranges in k for $c(k)$ in the rank-1 inhomogeneous random graph. Using these typical constrained subgraphs, for each $1 \ll k \ll n^{1/(\tau-1)}$, with high probability,

$$c(k) \propto \begin{cases} n^{2-\tau} \log(n) & k \ll n^{(\tau-2)/(\tau-1)}, \\ n^{2-\tau} \log(n/k^2) & n^{(\tau-2)/(\tau-1)} \ll k \ll \sqrt{n}, \\ k^{2\tau-6} n^{5-2\tau} & k \gg \sqrt{n}. \end{cases} \quad (6.2.22)$$

Thus, the scaling of $c(k)$ found by the variational principle is indeed the same as found in Chapter 4. Figure 6.3 shows that the three ranges for $c(k)$ in the rank-1 inhomogeneous random graph are also visible in simulations.

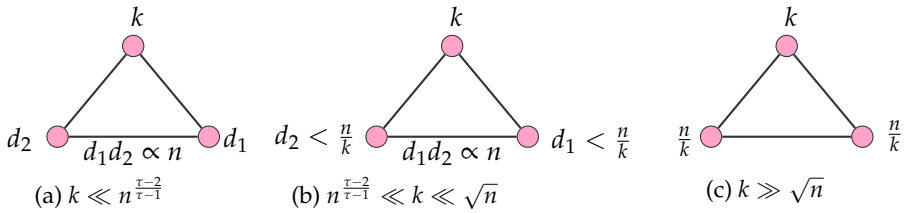


Figure 6.3: Typical triangles where one vertex has degree k in the rank-1 inhomogeneous random graph. When $k < \sqrt{n}$ a typical triangle is with two vertices such that the product of their degrees is proportional to n . When $k > \sqrt{n}$, the other two degrees in a typical triangle are proportional to n/k .

The extended variational principle in Section 6.3 shows that $c(k)$ in the rank-1 inhomogeneous random graph fails to be self-averaging for $k \ll n^{(\tau-2)/(\tau-1)}$, so that $c(k)$ -values for $k \ll n^{(\tau-2)/(\tau-1)}$ heavily fluctuate across various network samples.

Erased configuration models and uniform random graphs. The analysis of the optimal triangle structure in the rank-1 inhomogeneous random graph easily extends to the erased configuration model, described in Section 1.1.3, and the uniform random graph, described in Section 1.1.2. Both models can be approximated by a rank-1 inhomogeneous random graph with specific connection probabilities (see Chapter 5). Therefore, the variational principle shows that the optimal triangle structure as well as the behavior of the local clustering coefficient is the same as in (6.2.22) and Figure 6.3, which is consistent with the results of Chapter 5. The non-self-averaging behavior for $k \ll n^{(\tau-2)/(\tau-1)}$ also extends from the rank-1 inhomogeneous random graph to the erased configuration model and uniform random graphs.

Preferential attachment. Another important network null model is the preferential attachment model described in Section 1.1.5, a dynamic network model that generates scale-free networks when choosing the number of edges attached to each new vertex m and the parameter δ in the connection probability (1.1.4) as $\delta \in (-m, 0)$.

In the preferential attachment model, it is convenient to apply the variational principle to vertices with index of a specific order of magnitude instead of degrees. The vertex with index 1 is the oldest vertex, and the vertex with index n is the youngest vertex in the graph of size n . The probability that vertices with indices $i = n^{\alpha_1}$ and $j = n^{\alpha_2}$ such that $\alpha_1 < \alpha_2$ are connected is proportional to [73]

$$\mathbb{P}(j \rightarrow i) \propto j^{-\chi} i^{1-\chi} \propto n^{\alpha_1(\chi-1) - \alpha_2\chi}, \quad (6.2.23)$$

where $\chi = (\tau - 2)/(\tau - 1)$. Thus, the probability that a vertex with index n^{α_k} creates a triangle with vertices of indices proportional to n^{α_1} and n^{α_2} can be approximated by

$$\mathbb{P}(\Delta \text{ on indices } n^{\alpha_k}, n^{\alpha_1}, n^{\alpha_2}) \propto \begin{cases} n^{2\alpha_1(\chi-1) - \alpha_2 - 2\alpha_k\chi} & \text{if } \alpha_1 \leq \alpha_2 \leq \alpha_k, \\ n^{2\alpha_1(\chi-1) - \alpha_k - 2\alpha_2\chi} & \text{if } \alpha_1 \leq \alpha_k \leq \alpha_2, \\ n^{2\alpha_k(\chi-1) - \alpha_1 - 2\alpha_2\chi} & \text{if } \alpha_k \leq \alpha_1 \leq \alpha_2. \end{cases} \quad (6.2.24)$$

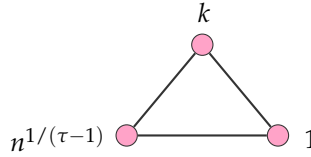


Figure 6.4: The most likely triangle containing a vertex of degree k in the preferential attachment model.

The probability that a randomly chosen vertex has age proportional to n^α is proportional to $n^{\alpha-1}$. Thus, the equivalent optimization problem to (6.2.2) becomes

$$\max_{\alpha_1 \leq \alpha_2} \begin{cases} n^{-2+\alpha_1(2\chi-1)-2\alpha_k\chi} & \text{if } \alpha_1 \leq \alpha_2 \leq \alpha_k, \\ n^{-2+(\alpha_1-\alpha_2)(2\chi-1)-\alpha_k} & \text{if } \alpha_1 \leq \alpha_k \leq \alpha_2, \\ n^{-2+2\alpha_k(\chi-1)-\alpha_2(2\chi-1)} & \text{if } \alpha_k \leq \alpha_1 \leq \alpha_2. \end{cases} \quad (6.2.25)$$

Using that $\chi \in (0, \frac{1}{2})$ when $\tau \in (2, 3)$, we find that for all $0 < \alpha_k < 1$ the unique optimizer is obtained by $\alpha_1^* = 0$ and $\alpha_2^* = 1$. Furthermore, the degree of a vertex of index $i \propto n^{\alpha_i}$ at time n , $d_i(n)$ satisfies with high probability [106, Chapter 8]

$$d_i(n) \propto (n/i)^{1/(\tau-1)} \propto n^{\frac{1-\alpha_i}{\tau-1}}. \quad (6.2.26)$$

Thus, vertices with age proportional to $n^{\alpha_1^*}$ have degrees proportional to $n^{1/(\tau-1)}$, whereas vertices with age proportional to $n^{\alpha_2^*}$ have degrees proportional to a constant. We conclude that for all $1 \ll k \ll n^{1/(\tau-1)}$, in the most likely triangle containing a vertex of degree k one of the other vertices has constant degree and the other has degree proportional to $n^{1/(\tau-1)}$.

Similarly to (6.2.26), a vertex of degree proportional to n^γ has index proportional to $n^{1-\gamma(\tau-1)}$. Thus, when $k \propto n^\gamma$

$$c(n^\gamma) \propto n^{2\gamma} n^{2-2\chi+1-1+\gamma(\tau-1)} = n^{\gamma(\tau-3)-2\chi}. \quad (6.2.27)$$

Thus, for each $1 \ll k \ll n^{1/(\tau-1)}$, with high probability,

$$c(k) \propto k^{\tau-3} n^{-2\chi}. \quad (6.2.28)$$

Figure 6.2c shows that this asymptotic slope in k is a good fit in simulations.

Figure 6.4 shows the most likely triangle containing a vertex of degree k in the preferential attachment model. Interestingly, this dominant triangle remains the same over the entire range of k , which is very different from the three regimes that are present in the rank-1 inhomogeneous random graph.

6.2.3 Random intersection graph

We next consider the random intersection graph [124], a random graph model with overlapping community structures that, like the hyperbolic random graph, generates

non-vanishing clustering levels. The random intersection graph contains n vertices, and m vertex attributes. Every vertex i chooses a random number of X_i vertex attributes, where $(X_i)_{i \in [n]}$ is an i.i.d. sample. These vertex attributes are sampled uniformly without replacement from all m attributes. Two vertices share an edge if they share at least $s \geq 1$ vertex attributes. One can think of the random intersection graph as a model for a social network, vertex attributes model the interests, or the group memberships of a person in the network. Then two vertices connect if their interests or group memberships are sufficiently similar. The overlapping community structures of the random intersection graph make the model highly clustered [30, 31], so that the typical triangles in the random intersection graph should behave considerably different than the typical triangles in the locally tree-like models described above.

To obtain random intersection graphs where vertices have asymptotically constant average degree, we need that $m^s \propto n$ [30], which we assume from now on. We further assume that s is of constant order of magnitude. Then the degree of vertex i with X_i vertex attributes is proportional to X_i^s [30]. Therefore, a vertex of degree k has approximately $k^{1/s}$ vertex attributes. To obtain a power-law degree distribution with exponent τ , the probability of vertex i having X_i vertex attributes scales as

$$\mathbb{P}(X_i = u) \propto u^{-\tau s}. \quad (6.2.29)$$

To apply the variational principle, we calculate the number of triangles between a vertex of degree k , and two vertices of degrees proportional to n^{α_1} and n^{α_2} . These vertices have proportionally to $n^{\alpha_1/s}$, respectively $n^{\alpha_2/s}$, vertex attributes. There are several ways for three vertices to form a triangle. If three vertices share the same set of at least s attributes, then they form a triangle. But if vertex i shares a set of at least s attributes with vertex j , vertex j shares another set of s attributes with vertex k and vertex k shares yet another set of s attributes with vertex i , these vertices also form a triangle. The most likely way for three vertices to form a triangle however, is for all three vertices to share the same set of s attributes [30]. There are $\binom{k^{1/s}}{s}$ ways to choose s attributes from the $k^{1/s}$ attributes of the degree- k vertex. Then, a triangle is formed if the two other vertices also contain these s attributes. Since these vertices have $n^{\alpha_1/s}$ and $n^{\alpha_2/s}$ attributes chosen uniformly without replacement from all m attributes, the probability that the first vertex shares these s attributes is $\binom{m-s}{n^{\alpha_1/s}-s} / \binom{m}{n^{\alpha_1/s}}$. We can then calculate the probability of a triangle being present as

$$\begin{aligned} \mathbb{P}(\triangle \text{ on degrees } k, n^{\alpha_1}, n^{\alpha_2}) &\propto \binom{k^{1/s}}{s} \frac{\binom{m-s}{n^{\alpha_1/s}-s} \binom{m-s}{n^{\alpha_2/s}-s}}{\binom{m}{n^{\alpha_1/s}} \binom{m}{n^{\alpha_2/s}}} \\ &\propto k n^{\alpha_1 + \alpha_2} m^{-2s} \propto k n^{\alpha_1 + \alpha_2 - 2}. \end{aligned} \quad (6.2.30)$$

Combining this with (6.2.4) yields

$$c(k) \propto n^2 k^{-2} \max_{\alpha_1, \alpha_2} k n^{(\alpha_1 + \alpha_2)(2-\tau) - 2} \propto k^{-1}, \quad (6.2.31)$$

where the maximizer is $\alpha_1 = \alpha_2 = 0$. Thus, a most likely triangle in the random intersection graph is a triangle containing one vertex of degree k , where the two

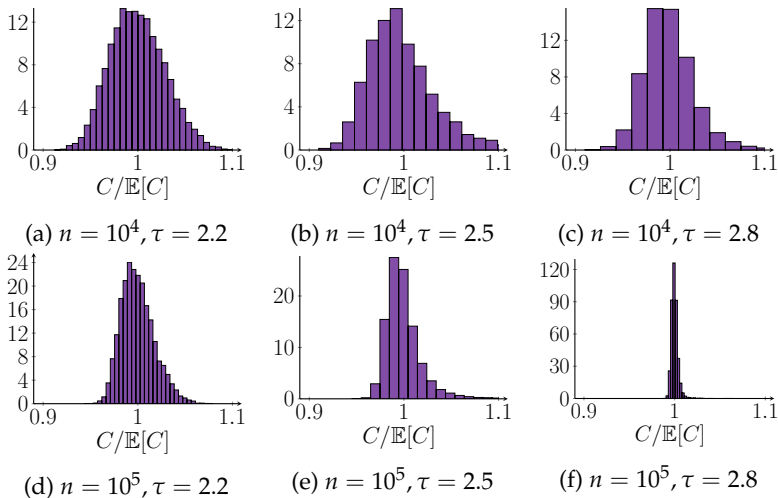


Figure 6.5: The self-averaging behavior of the clustering coefficient in the hyperbolic random graph. The plots show density estimates of the rescaled global clustering coefficient based on 10^4 samples of hyperbolic random graphs with $\nu = 1$.

other vertices have degrees proportional to a constant. The result that $c(k) \propto k^{-1}$ is in agreement with the results obtained in [30]. Moreover, the most likely triangle is a triangle where one vertex has degree k , and the other two vertices have constant degree. Thus, in terms of clustering, the random intersection graph behaves the same as the hyperbolic random graph with $\tau > 5/2$.

6.2.4 Clustering coefficient

The average clustering coefficient of a network is defined as

$$C = \frac{1}{n} \sum_{i=1}^n \frac{N_i^\Delta}{d_i(d_i - 1)} = \sum_k p_k c(k), \quad (6.2.32)$$

where N_i^Δ denotes the number of triangles attached to vertex i and p_k denotes the fraction of vertices of degree k . Because the power-law degree-distribution decays rapidly in k , $C \propto c(k)$ for constant k , since we know that $c(k)$ is approximately constant for constant k (which was shown rigorously for the rank-1 inhomogeneous random graph in Chapter 3). Thus, the self-averaging properties of the average clustering coefficient are determined by the self-averaging properties of $c(k)$ for small values of k . This implies that in the rank-1 inhomogeneous random graph, C is indeed non-self-averaging (see Section 6.3) which supports numerical results in [201]. In the hyperbolic random graph on the other hand, the self-averaging $c(k)$ curve shows that also C is self-averaging. Figure 6.5 shows that indeed the fluctuations in C decrease as n grows. Figure 6.6 shows that in the rank-1 inhomogeneous random graph C is indeed non-self-averaging, since the fluctuations in C persist for large values of n .

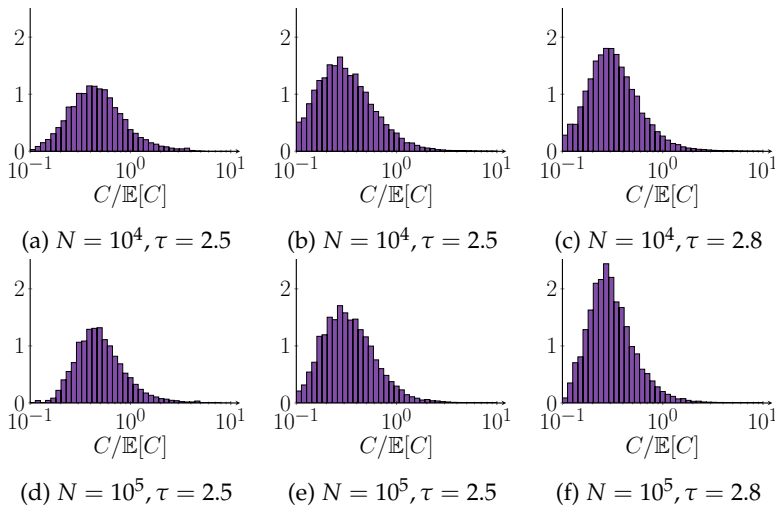


Figure 6.6: The non-self-averaging behavior of the clustering coefficient in the rank-1 inhomogeneous random graph. The plots show density estimates based on 10^4 samples of rank-1 inhomogeneous random graphs.

6.3 Fluctuations

Due to their stochastic nature, whenever a random graph model is employed for modeling real-world networks, one needs to consider the fluctuations, typically expressed in terms of the self-averaging property. We say that $c(k)$ is self-averaging when $\text{Var}(c(k)) / \mathbb{E}[c(k)]^2 \rightarrow 0$ as $n \rightarrow \infty$, so that the fluctuations of $c(k)$ vanish in the large-network limit. When $c(k)$ fails to be self-averaging, the fluctuations persist even in the large-network limit, so that the average of $c(k)$ over many network realizations cannot be viewed as a reliable descriptor of the local clustering. We now show how to apply the variational principle (6.2.4) to constrained subgraphs larger than triangles, which leads to a complete characterization of $\text{Var}(c(k)) / \mathbb{E}[c(k)]^2$ in the large-network limit. The variational principle can hence determine for any value of k whether $c(k)$ is self-averaging or not. In this way we are able to show that for the hyperbolic random graph, $c(k)$ is self-averaging for all values of $\tau \in (2, 3)$ and all k . This implies that for large enough n , one sample of the hyperbolic random graph is sufficient to obtain the characteristic behavior of $c(k)$. With the extended variational principle we also show that the rank-1 inhomogeneous random graph on the other hand is not always self-averaging in terms of $c(k)$.

6.3.1 Fluctuations in the rank-1 inhomogeneous random graph

We first investigate whether $c(k)$ is self-averaging for k small in the rank-1 inhomogeneous random graph. We first study the expected value of $c(k)$. In the variational principle, we obtained the typical number of triangles where one vertex has degree k by putting the hard constraint $\alpha_1, \alpha_2 \leq 1/(\tau - 1)$ on the degrees of the other two

vertices in the triangle. If we relax this constraint, we can compute $\mathbb{E}[c(k)]$. This quantity can be interpreted as the value of $c(k)$ obtained after simulating many rank-1 inhomogeneous random graphs, and taking the average value of $c(k)$ over all these rank-1 inhomogeneous random graphs. As long as $k \gg n^{(\tau-2)/(\tau-1)}$, we see from (6.2.21) that the largest contribution to $c(k)$ is from vertices with degrees strictly smaller than $n^{1/(\tau-1)}$. Thus, removing the constraint on the maximal degree does not influence the major contribution for $c(k)$. When $k \ll n^{(\tau-2)/(\tau-1)}$ however, the major contribution includes vertices of degree $n^{1/(\tau-1)}$. Removing the hard constraint then results in an optimal contribution which is slightly different from (6.2.21):

$$\begin{aligned} \alpha_1 + \alpha_2 &= 1, n^{\alpha_1}, n^{\alpha_2} < n/k & k \ll \sqrt{n}, \\ n^{\alpha_1} &= n/k, n^{\alpha_2} = n/k & k \gg \sqrt{n}. \end{aligned} \quad (6.3.1)$$

Similarly to the computation that leads to (6.2.21), this gives for $\mathbb{E}[c(k)]$ that

$$\mathbb{E}[c(k)] \propto \begin{cases} n^{2-\tau} \log(n/k^2) & k \ll \sqrt{n}, \\ k^{2\tau-6} n^{5-2\tau} & k \gg \sqrt{n}. \end{cases} \quad (6.3.2)$$

Thus, the typical behavior of $c(k)$ is the same as its average behavior for $k \gg n^{(\tau-2)/(\tau-1)}$. For small values of k however, the flat regime disappears and is replaced by a regime that depends on the logarithm of k . The difference between the expected and the typical value of $c(k)$ is important to take into account for the accuracy of numerical experiments, where the median value of $c(k)$ over many network realizations behaves like (6.2.22), but the average value behaves as (6.3.2).

We now proceed to compute the variance of $c(k)$. The variance of $c(k)$ equals

$$\text{Var}(c(k)) = k^{-4} \sum'_{u,v \in [n]} \sum'_{w,z \in [n]} \mathbb{P}(\Delta_{i,u,v} \Delta_{j,w,z}) - \mathbb{P}(\Delta_{i,u,v}) \mathbb{P}(\Delta_{j,w,z}), \quad (6.3.3)$$

where i and j denote two randomly chosen vertices of degree k , so that the weights of i and j satisfy $w_i, w_j = k(1 + o_{\mathbb{P}}(1))$. When i, u, v and j, w, z do not overlap, their weights are independent, so that the event that i, u and v form a triangle and the event that j, w and z form a triangle are independent. Thus, when i, j, u, v, w, z are distinct, $\mathbb{P}(\Delta_{i,u,v} \Delta_{j,w,z}) = \mathbb{P}(\Delta_{i,u,v}) \mathbb{P}(\Delta_{j,w,z})$, so that the contribution from 6 distinct indices to (6.3.3) is zero. Since $w_i = k(1 + o_{\mathbb{P}}(1))$, $\mathbb{P}(\Delta_{i,u,v} \Delta_{i,w,z}) = \mathbb{P}(\Delta_{i,u,v}) \mathbb{P}(\Delta_{i,w,z}) (1 + o_{\mathbb{P}}(1))$ when u, v, w, z are distinct. Thus, the contribution to the variance from $i = j$ and u, v, w, z distinct can be bounded as $o(\mathbb{E}[c(k)]^2)$. When $u = w$ for example, the first term in (6.3.3) denotes the probability that a bow-tie is present with u as middle vertex. Furthermore, since the degrees are i.i.d., for any $i \neq u \neq v$, such that $d_i = k$,

$$\mathbb{P}(\Delta_{i,u,v}) = \frac{\mathbb{E}[\Delta_k]}{2\binom{n}{2}}, \quad (6.3.4)$$

where Δ_k denotes the number of triangles attached to a randomly chosen vertex of degree k . Let $r = \mathbb{P}(i = j)$. Then,

$$\text{Var}(c(k)) = k^{-4} \left(4\mathbb{E} \left[\begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \bullet \end{array} \right] + 4\mathbb{E} \left[\begin{array}{c} \bullet \\ \diagup \quad \bullet \\ \bullet \end{array} \right] + 2\mathbb{E} \left[\begin{array}{c} \bullet \\ \diagup \quad \bullet \\ \bullet \end{array} \right] + 4\mathbb{E} \left[\begin{array}{c} \bullet \\ \diagup \quad \bullet \\ \bullet \end{array} \right] + 8\mathbb{E} \left[\begin{array}{c} \bullet \\ \diagup \quad \bullet \\ \bullet \end{array} \right] \right)$$

$$+ 4\mathbb{E} \left[\begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \circ \end{array} \right] + (1-r) \left(\mathbb{E} \left[\begin{array}{c} \bullet \quad \bullet \\ \diagdown \quad \diagup \\ \bullet \end{array} \right] + 4\mathbb{E} \left[\begin{array}{c} \bullet \quad \bullet \\ \diagdown \quad \diagup \\ \circ \end{array} \right] + 2\mathbb{E} [\Delta_k] \right) + \mathbb{E} [\Delta_k]^2 O(n^{-1}) \quad (6.3.5)$$

where $\begin{array}{c} \bullet \quad \bullet \\ \diagdown \quad \diagup \\ \bullet \end{array}$ denotes the number of bow-ties attached to two randomly chosen (white) vertices of degree k , and $\begin{array}{c} \bullet \quad \bullet \\ \diagdown \quad \diagup \\ \bullet \end{array}$ denotes the number of bow-ties attached to one randomly chosen (white) vertex of degree k . The combinatorial factor 4 arises in the first term because there are 4 ways to construct a bow-tie where two vertices have degree k by letting two triangles containing a degree k vertex overlap. The other combinatorial factors arise similarly. Because $w_i, w_j = k(1 + o_{\mathbb{P}}(1))$, $\mathbb{P}(\Delta_{i,u,v} \Delta_{i,u,z}) = \mathbb{P}(\Delta_{i,u,v} \Delta_{j,u,z}) (1 + o_{\mathbb{P}}(1))$, so that $\mathbb{E} \left[\begin{array}{c} \bullet \quad \bullet \\ \diagdown \quad \diagup \\ \bullet \end{array} \right] = (1 + o_{\mathbb{P}}(1)) \mathbb{E} \left[\begin{array}{c} \bullet \quad \bullet \\ \diagdown \quad \diagup \\ \bullet \end{array} \right]$ and similarly $\mathbb{E} [\Delta_k] = (1 + o_{\mathbb{P}}(1)) \mathbb{E} \left[\begin{array}{c} \bullet \quad \bullet \\ \diagdown \quad \diagup \\ \bullet \end{array} \right]$. Thus, with high probability

$$\begin{aligned} \text{Var}(c(k)) &\propto k^{-4} \left(\mathbb{E} \left[\begin{array}{c} \bullet \quad \bullet \\ \diagdown \quad \diagup \\ \bullet \end{array} \right] + \mathbb{E} \left[\begin{array}{c} \bullet \quad \bullet \\ \diagdown \quad \diagup \\ \bullet \end{array} \right] + \mathbb{E} \left[\begin{array}{c} \bullet \quad \bullet \\ \diagdown \quad \diagup \\ \circ \end{array} \right] + \mathbb{E} \left[\begin{array}{c} \bullet \quad \bullet \\ \diagdown \quad \diagup \\ \bullet \end{array} \right] + \mathbb{E} \left[\begin{array}{c} \bullet \quad \bullet \\ \diagdown \quad \diagup \\ \bullet \end{array} \right] \right. \\ &\quad \left. + \mathbb{E} \left[\begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \circ \end{array} \right] + \mathbb{E} [\Delta_k]^2 O(n^{-1}) \right) + o(\mathbb{E}[c(k)]^2) \end{aligned} \quad (6.3.6)$$

We write the first expectation as

$$\mathbb{E} \left[\begin{array}{c} \bullet \quad \bullet \\ \diagdown \quad \diagup \\ \bullet \end{array} \right] = n^3 \mathbb{P} \left(\begin{array}{c} \bullet \quad \bullet \\ \diagdown \quad \diagup \\ \bullet \end{array} \right), \quad (6.3.7)$$

where $\mathbb{P} \left(\begin{array}{c} \bullet \quad \bullet \\ \diagdown \quad \diagup \\ \bullet \end{array} \right)$ denotes the probability that two randomly chosen vertices of degree k form the constrained bow-tie together with three randomly chosen other vertices. We can compute this probability with a constrained variational principle. By symmetry of the bow-tie subgraph, the optimal degree range of the bottom right vertex and the upper right vertex is the same. Let the degree of the middle vertex scale as n^{α_1} , and the degrees of the other two vertices as n^{α_2} . Then, we write the constrained variational principle, similarly to (6.2.20), as

$$\max_{\alpha_1, \alpha_2} n^{(\alpha_1 + 2\alpha_2)(1-\tau)} \min(kn^{\alpha_1-1}, 1)^2 \min(kn^{\alpha_2-1}, 1)^2 \min(n^{\alpha_1 + \alpha_2 - 1}, 1)^2 \quad (6.3.8)$$

For $k \ll \sqrt{n}$, the unique optimal contribution is from $n^{\alpha_1} = n/k$ and $n^{\alpha_2} = k$, as shown in Figure 6.7a. Thus, the expected number of such bow-ties scales as

$$\mathbb{E} \left[\begin{array}{c} \bullet \quad \bullet \\ \diagdown \quad \diagup \\ \bullet \end{array} \right] \propto n^3 (n/k)^{1-\tau} k^{2(1-\tau)} k^4 n^{-2} = n^{2-\tau} k^{5-\tau}. \quad (6.3.9)$$

Thus,

$$\text{Var}(c(k)) > k^{-4} \mathbb{E} \left[\begin{array}{c} \bullet \quad \bullet \\ \diagdown \quad \diagup \\ \bullet \end{array} \right] \propto n^{2-\tau} k^{1-\tau}, \quad (6.3.10)$$

so that (6.3.2) yields that for k small

$$\frac{\text{Var}(c(k))}{\mathbb{E}[c(k)]^2} > \frac{n^{2-\tau} k^{1-\tau}}{n^{4-2\tau} \log^2(n/k^2)}, \quad (6.3.11)$$

which tends to infinity as long as $k \ll n^{(\tau-2)/(\tau-1)}$. Therefore $c(k)$ is non self-averaging as long as $k \ll n^{(\tau-2)/(\tau-1)}$.

For $n^{(\tau-2)/(\tau-1)} \ll k \ll \sqrt{n}$, we can similarly compute the optimum contributions of all other constrained motifs to the variance as in Figure 6.7. Since all contributions have smaller magnitude than $\mathbb{E}[c(k)]^2$ (obtained from (6.3.2)), $c(k)$ is self-averaging. For $k \gg \sqrt{n}$, a constrained variational principle again provides the contribution of all constrained motifs to the variance of $c(k)$, visualized in Figure 6.8. Comparing this with (6.3.2) shows that $c(k)$ is also self-averaging for $k \gg \sqrt{n}$.

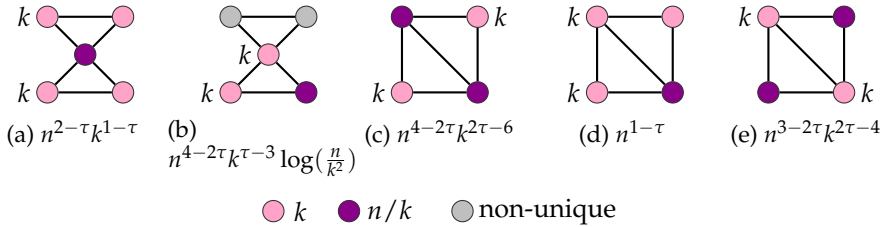


Figure 6.7: Contribution to the variance of $c(k)$ in the rank-1 inhomogeneous random graph (Eq. 6.3.6) from merging two triangles where one vertex has degree $k \ll \sqrt{n}$. The vertex color indicates the optimal vertex degree.

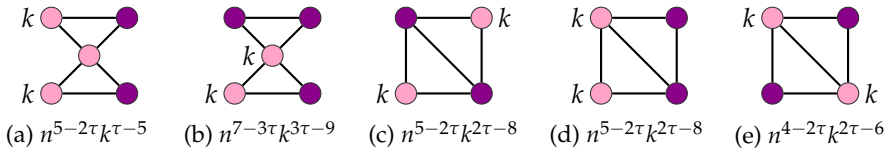


Figure 6.8: Contribution to the variance of $c(k)$ in the rank-1 inhomogeneous random graph (Eq. 6.3.6) from merging two triangles where one vertex has degree $k \gg \sqrt{n}$. The vertex color indicates the optimal vertex degree as in Figure 6.7.

6.3.2 Fluctuations in the hyperbolic random graph

We now analyze the fluctuations in the hyperbolic random graph. Again, we first investigate $\mathbb{E}[c(k)]$. As for the rank-1 inhomogeneous random graph, $\mathbb{E}[c(k)]$ can be obtained by removing the constraint that the maximal degree scales as $n^{1/(\tau-1)}$, which means that we optimize (6.2.15) over $\alpha_1, \alpha_2 \in [0, 1]$ instead of $[0, 1/(\tau-1)]$. Extending the allowed range of α_1 and α_2 does not change the optimizer, so that, similarly to (6.2.17),


$$\mathbb{E}[c(k)] \propto \begin{cases} k^{-1} & \tau > \frac{5}{2}, \\ k^{4-2\tau} & \tau < \frac{5}{2}, k \ll \sqrt{n}, \\ k^{2\tau-6}n^{5-2\tau} & \tau < \frac{5}{2}, k \gg \sqrt{n}. \end{cases} \quad (6.3.12)$$

We then proceed to analyze the variance of $c(k)$. We again use (6.3.5) to analyze the fluctuations of $c(k)$, and we write


$$\text{Var}(c(k)) = k^{-4}O\left(\mathbb{E}\left[\begin{array}{c} \bullet \\ \bullet \\ \bullet \\ \bullet \end{array}\right] + \mathbb{E}\left[\begin{array}{c} \bullet \\ \bullet \\ \bullet \\ \bullet \end{array}\right] + \mathbb{E}\left[\begin{array}{c} \bullet \\ \bullet \\ \bullet \\ \bullet \end{array}\right] + \mathbb{E}\left[\begin{array}{c} \bullet \\ \bullet \\ \bullet \\ \bullet \end{array}\right] + \mathbb{E}\left[\begin{array}{c} \bullet \\ \bullet \\ \bullet \\ \bullet \end{array}\right]\right)$$

$$+ \mathbb{E} \left[\text{triangle} \right] + \mathbb{E} \left[\text{figure 6.9a} \right] + \mathbb{E} \left[\text{figure 6.10a} \right] + \mathbb{E} [\Delta_k] \Big) + \mathbb{E} [\Delta_k]^2 O(n^{-1}) \quad (6.3.13)$$

Thus, as in the rank-1 inhomogeneous random graph, we need to obtain the expected number of merged constrained triangles attached to randomly chosen vertices of degree k . Figures 6.9 and 6.10 show the contribution of all types of merged triangles to the variance of $c(k)$. These contributions are all smaller than $\mathbb{E} [c(k)]^2$ (see (6.3.12)), so that $c(k)$ is self-averaging over its entire spectrum.

For example, the expected number of subgraphs of type  attached to two randomly chosen (white) vertices of degree k , where the top and bottom right vertices have degree proportional to n^{α_1} and the middle vertex has degree proportional to n^{α_2} as


$$n^3 n^{(2\alpha_1 + \alpha_2)(1-\tau)} \min(kn^{\alpha_1-1}, 1)^2 \min(kn^{\alpha_2-1}, 1)^2 \min(n^{\alpha_1} \max(n^{\alpha_2-1}, k^{-1}), 1)^2. \quad (6.3.14)$$

Optimizing this over α_1 and α_2 yields that for $k \ll \sqrt{n}$ the number of  subgraphs is dominated by the type displayed in Figure 6.9a, where $n^{\alpha_1} \propto k$ and $n^{\alpha_2} \propto n/k$. Computing this contribution results in

$$\mathbb{E} \left[\text{figure 6.9a} \right] \propto n^3 k^{2(1-\tau)} \left(\frac{n}{k}\right)^{1-\tau} \left(\frac{k^2}{n}\right)^2 = n^{2-\tau} k^{5-\tau}. \quad (6.3.15)$$

Thus, using (6.3.13) shows that the contribution to the variance is $n^{2-\tau} k^{1-\tau}$, as shown in Figure 6.9a. We obtain using (6.3.12) that for $k \ll \sqrt{n}$,

$$\frac{n^{2-\tau} k^{1-\tau}}{\mathbb{E} [c(k)]^2} \propto \begin{cases} n^{2-\tau} k^{3-\tau} \ll n^{(7-3\tau)/2} & \tau > \frac{5}{2}, \\ n^{2-\tau} k^{3\tau-7} \ll \max(n^{(\tau-3)/2}, n^{2-\tau}) & \tau < \frac{5}{2}, \end{cases} \quad (6.3.16)$$


which tends to zero as $n \rightarrow \infty$. Thus, the contribution to the variance from the  subgraphs tends to zero in the large network limit for $k \ll \sqrt{n}$.

The optimizer of (6.3.14) for $k \gg \sqrt{n}$ is for $n^{\alpha_1} \propto n/k$ and $n^{\alpha_2} \propto n/k$. Thus, similarly to (6.3.15)

$$\mathbb{E} \left[\text{figure 6.10a} \right] \propto n^3 \left(\frac{n}{k}\right)^{3(1-\tau)} \left(\frac{n}{k^2}\right)^2 = n^{8-3\tau} k^{3\tau-7}. \quad (6.3.17)$$

The contribution to 6.3.13 then is $n^{8-3\tau} k^{3\tau-11}$, as Figure 6.10a shows. Thus, for $k \gg \sqrt{n}$,

$$\frac{n^{8-3\tau} k^{3\tau-11}}{\mathbb{E} [c(k)]^2} \propto \begin{cases} n^{8-3\tau} k^{3(\tau-3)} \ll n^{(7-3\tau)/2} & \tau > \frac{5}{2}, \\ n^{\tau-2} k^{1-\tau} \ll n^{(\tau-3)/2} & \tau < \frac{5}{2}, \end{cases} \quad (6.3.18)$$

which tends to zero as $n \rightarrow \infty$, showing that indeed the contribution from the  subgraph to the variance is small when $k \gg \sqrt{n}$. The contributions of other subgraphs can be computed similarly.

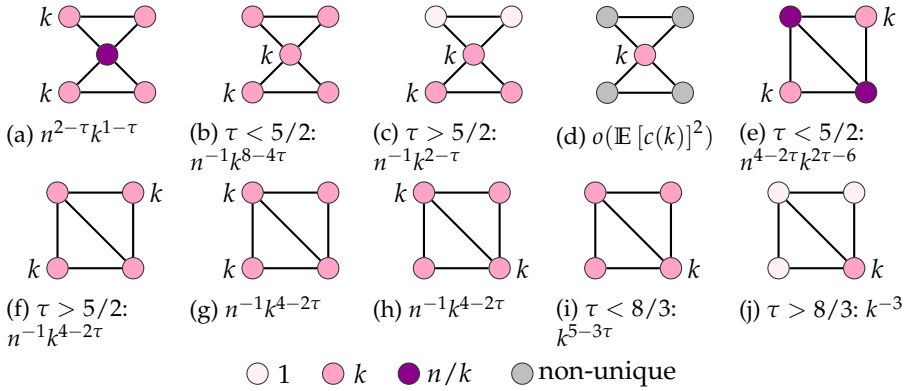


Figure 6.9: Contribution to the variance of $c(k)$ in the hyperbolic model from merging two triangles where one vertex has degree $k \ll \sqrt{n}$ (see (6.3.13)). The vertex color indicates the optimal vertex degree.

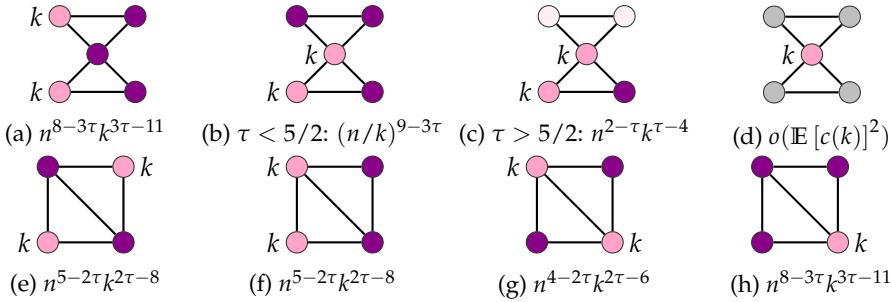


Figure 6.10: Contribution to the variance of $c(k)$ in the hyperbolic model from merging two triangles where one vertex has degree $k \gg \sqrt{n}$ (see (6.3.13)). The vertex color indicates the optimal vertex degree as in Figure 6.9.

6.4 Discussion

Universality of the variational principle. We have introduced a constrained variational principle that finds the optimal triangle structure containing a vertex of degree k , settling the trade-off between the connectedness of high-degree vertices and their rareness. This variational principle can be applied to many random graph null models generating scale-free networks. We have applied the variational principle to find optimal triangle structures in rank-1 inhomogeneous random graphs, the preferential attachment model, random intersection graphs and the hyperbolic random graph, but we believe that the variational principle can easily be applied to other types of random graph null models such as the geometric inhomogeneous random graph [49] or the spatial preferential attachment model [3, 114].

Our method also extend to other types of constrained motifs, which allows for example to investigate higher order clustering [19, 227]. Furthermore, these higher order constrained motifs allow to investigate the self-averaging properties of $c(k)$.

Differences between the models. While the applicability of the variational principle across the different null models is universal, the resulting clustering spectra are significantly different across the various random graph models. For $\tau \in (2, 3)$, in the rank-1 inhomogeneous random graph, the erased configuration model, uniform random graphs (see (6.2.22)) as well as in the preferential attachment model (see (6.2.28)) $c(k)$ decreases with the network size, because of the locally-tree like nature of these networks. This fall-off in n can be explained by the optimal triangle structures. In all optimal triangle structures of Figures 6.3 and 6.4, a vertex whose degree grows in n is present. These vertices are rare, so that the fact that the most likely triangle contains such high-degree vertices is caused by the network being locally tree-like.

In the hyperbolic model and the random intersection graph on the other hand, the optimal triangle structures of Figure 6.1 contain low-degree vertices for small values of k . In models without geometric correlations, the probability of connecting two vertices usually increases in the degree of the vertices involved. Therefore, models without correlations mostly contain triangles with high-degree vertices, causing these networks to be locally tree-like. The geometric correlations in the hyperbolic model on the other hand make it more likely for two low-degree neighbors to connect, causing the most likely triangle to contain lower-degree vertices. These lower-degree vertices are abundant, which explains why, for small k , $c(k)$ does not vanish as n grows large in the hyperbolic model, which is more alike the behavior of $c(k)$ in real-world networks.

Another advantage of the hyperbolic random graph over the locally tree-like networks is that $c(k)$ in the hyperbolic model is self-averaging over the entire range of k . In the rank-1 inhomogeneous random graph, the erased configuration model as well as the uniform random graph, $c(k)$ is non-self-averaging for k small. Thus, it suffices to generate one large hyperbolic random graph to investigate the behavior of $c(k)$. In the other models, the non-self-averaging nature of $c(k)$ makes statistical investigation of the $c(k)$ curve more difficult.

$1/k$ fall-off. The clustering coefficient in the hyperbolic random graph as well as the random intersection graph satisfies $c(k) \sim k^{-1}$ for $\tau > 5/2$. This fall-off has been observed in many other scale-free random graph null models containing non-trivial clustering, such as preferential attachment models with extra triangles [128, 132, 204], and fractal-like random graph models [187]. We now explain heuristically why the k^{-1} fall-off occurs so frequently in network models containing non-trivial clustering. The clustering coefficient can be interpreted as the probability that two randomly chosen neighbors of a vertex of degree k connect. Thus, the clustering coefficient of a vertex of degree k equals the number of triangles containing that vertex divided by $k(k-1)/2 \approx k^2$. Suppose that all neighbors of a vertex of degree k participate in exactly one triangle together with the degree- k vertex. Then, the number of triangles containing the vertex of degree k equals $k/2$, so that $c(k) \propto k^{-1}$. Similarly, when a vertex of degree k typically does not participate in cliques that are larger than some bound M , then the clustering coefficient will scale at most as k^{-1} for k large enough. Therefore, the frequently observed k^{-1} fall-off in real-world networks indicates that in most real-world networks, the number of triangles a degree- k vertex participates

in k grows linearly in k , and they do not form growing cliques.

A network can only have $c(k) \propto k^\gamma$ for some $\gamma > -1$ when a typical vertex of degree k participates in a clique of size growing in k , which was also explained in [196–198]. These growing cliques require the presence of high-degree vertices. However, by the power-law degree distribution only few high-degree vertices are present. Thus, it is not possible for these growing cliques to persist over the entire range of k .

Growing cliques in the hyperbolic model. In the hyperbolic model, for $\tau < 5/2$ and $k \ll \sqrt{n}$, the optimal triangle structure of a vertex of degree k contains two other vertices of degree proportional to k and $c(k) \propto k^{4-2\tau}$. Since $4 - 2\tau > -1$ for $\tau < 5/2$, by the discussion above, a typical vertex of degree k must participate in a clique of size growing in k . Furthermore, since the typical triangle containing a vertex of degree k for $\tau < 5/2$ contains two other degree- k vertices, these growing cliques consist of vertices with degree proportional to k .

In the hyperbolic random graph, a positive fraction of vertices of degree proportional to \sqrt{n} forms a clique [85]. This explains the different regime in $c(k)$ for $k \gg \sqrt{n}$ when $\tau < 5/2$. Indeed, for $k \gg \sqrt{n}$, there are fewer vertices of degree proportional to k than vertices of degree proportional to \sqrt{n} . Most vertices of degree proportional to \sqrt{n} already form a clique. Therefore, from $k \propto \sqrt{n}$ onwards the typical clique that a vertex of degree k participates in does not grow in k anymore, so that, by the $1/k$ explanation for $k \gg \sqrt{n}$ and for $\tau < 5/2$, the fall-off in k scales as k^γ with $\gamma \leq -1$ (see (6.2.17), where the exponent of k equals $2\tau - 6 < -1$ when $k \gg \sqrt{n}$ and $\tau < 5/2$).

Local clustering in real-world networks. Figure 4.2 plots the behavior of $c(k)$ in nine large real-world networks. All networks show non-trivial clustering: their $c(k)$ values are high for small values of k , while most networks consist of several millions of vertices. This shows that the locally tree-like models where $c(k)$ vanishes as n grows large are not suitable to model real-world networks. Several clustering curves are approximately described by $c(k) \propto k^{-1}$ for large values of k (see Supplementary Table 1). Thus, these networks behave similarly in terms of clustering as the hyperbolic model and the random intersection graph. The networks where $c(k) \propto k^\gamma$ with $\gamma < -1$ (such as the three social networks), display similar behavior as the curve of the hyperbolic model in Figure 6.2a for $\tau < 5/2$. First, $c(k)$ decays very slowly in k . Then, when k becomes larger, $c(k)$ suddenly drops faster in k .

Part II

Subgraph structure

7 Subgraphs in erased configuration models

Based on:

Optimal subgraph structures in scale-free configuration models
R. van der Hofstad, J.S.H. van Leeuwen and C. Stegehuis
arXiv:1709.03466

Subgraphs reveal information about the geometry and functionalities of complex networks. In this chapter, we count the number of times a small connected graph occurs as a subgraph (motif counting) or as an induced subgraph (graphlet counting) of the erased configuration model with degree exponent $\tau \in (2, 3)$. We introduce a novel class of optimization problems, where for any given subgraph, the unique optimizer describes the most likely degrees of the vertices that together span the subgraph. We find that every subgraph occurs typically between vertices with specific degree ranges. In this way, we can count and characterize all subgraphs.

7.1 Introduction

In Part I we have investigated the presence of triangles in scale-free networks. The triangle is the most studied network subgraph, because it not only describes the clustering coefficient, but also signals hierarchy and community structure [187]. However, other subgraphs such as larger cliques are equally important for understanding network organization [19, 208]. Indeed, subgraph counts might vary considerably across different networks [152, 153, 223] and any given network will have a set of statistically significant subgraphs. Statistical relevance can be expressed by comparing the real networks to some mathematically tractable null model. This comparison filters out the effect of network properties such as the degree sequence and the network size on the motif count. A popular statistic takes the subgraph count, subtracts the expected number of subgraphs in a null model, and divides by the variance in the null model [87, 153, 169]. Such a standardized test statistic sheds light on whether a subgraph is overrepresented in comparison to the null model that serves as the baseline. This raises the question of what null model to use. To filter out the effect of the network degrees, a natural candidate is the uniform simple graph with the same degrees as the original network.

For $\tau > 3$, when the degree distribution has finite second moment, it is easy to generate such graphs using the configuration model. For $\tau < 3$, however, the configuration model fails to create simple graphs with high probability [116], and null models usually involve rewiring edges of the original graph. Consequently, the counting of subgraphs remains mathematically intractable, and one needs to resort to

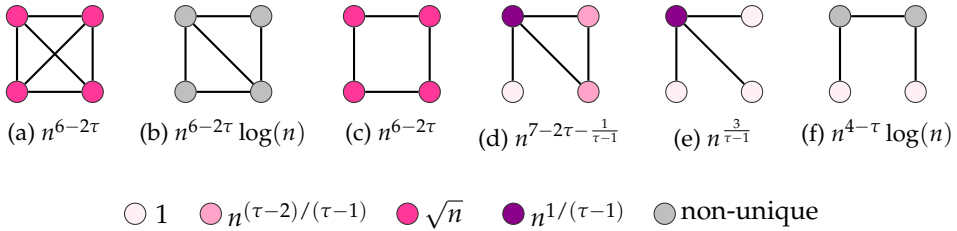


Figure 7.1: Graphlets. Order of magnitude of $N^{(\text{ind})}(H)$ for all connected graphs on 4 vertices (constants ignored). The vertex colors correspond to the typical vertex degrees.

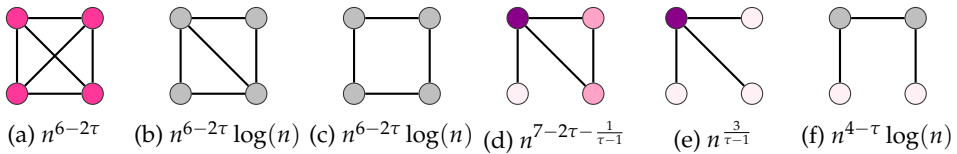


Figure 7.2: Motifs. Order of magnitude of $N^{(\text{sub})}(H)$ for all connected graphs on 4 vertices (constants ignored). The vertex color indicates the optimal vertex degree, as in Figure 7.1.

algorithms for exhaustive counting of motifs [144, 216], or estimations of the number of motifs by sampling [127], which is computationally expensive.

In this chapter, we address this problem by counting subgraphs in the erased configuration model instead of the configuration model. We count the number of times a small connected graph H occurs as a subgraph (motif counting) or as an induced subgraph, where edges not present in H are also not allowed to be present in the subgraph (graphlet counting), in an erased configuration model G with degree exponent $\tau \in (2, 3)$. Let $G = (V, E)$ be a graph, and $H = (V_H, E_H)$ be a small, connected graph. When we count graphlets H , we are interested in $N^{(\text{ind})}(H)$, the number of induced subgraphs of G that are isomorphic to H . We also study motifs, where we count $N^{(\text{sub})}(H)$, the number of occurrences of H as a subgraph of G . When H is a complete graph, $N^{(\text{ind})}(H) = N^{(\text{sub})}(H)$, otherwise $N^{(\text{ind})}(H) \leq N^{(\text{sub})}(H)$. There is thus a subtle difference between graphlets and motifs.

We find that every small graph H , whether it is a graphlet or motif, occurs typically between vertices in G with degrees in specific ranges. An example of these typical degree ranges for subgraphs on 4 vertices is shown in Figures 7.1 and 7.2 (which will be discussed in more detail in Section 7.2.4). We show that many subgraphs consist exclusively of \sqrt{n} -degree vertices, including cliques of all sizes. Hence, in such subgraphs, hubs (of degree close to the maximal value $n^{1/(\tau-1)}$) are unlikely to participate in a typical subgraph. However, hubs can be part of other subgraphs such as stars. We define optimization problems that find these optimal degree ranges for every motif and graphlet.

By studying the erased configuration model, we deal with multiple edges and self-loops of the configuration model by excluding double counting.

The erased configuration model. In this chapter, we study erased configuration models (see Section 1.1.3) where the degree distribution has infinite variance. In particular, we take the degrees to be an i.i.d. sample from a random variable D such that

$$\mathbb{P}(D = k) = ck^{-\tau}(1 + o(1)), \quad \text{as } k \rightarrow \infty, \quad (7.1.1)$$

where $\tau \in (2, 3)$ so that $\mathbb{E}[D^2] = \infty$ and

$$\mathbb{E}[D] = \mu < \infty. \quad (7.1.2)$$

When this sample constructs a degree sequence such that the sum of the degrees is odd, we add an extra half-edge to the last vertex. This does not affect our computations. In this setting, D_{\max} is of order $n^{1/(\tau-1)}$, where D_{\max} denotes the maximal degree of the degree sequence.

Throughout this chapter, we will denote the sampled degree of a vertex in the erased configuration model by D_i . This may not be the same as the actual degree of a vertex in the erased configuration model, since self-loops and multiple edges are removed. Since we study subgraphs H , we sometimes also need to use the degree of a vertex in H inside the subgraph. We will denote the degree of a vertex i of a subgraph H by d_i .

Outline. We present our main results in Section 7.2, including the theorems that characterize all optimal subgraph structures in terms of solutions to optimization problems. We apply these theorems to describe the optimal configurations of all subgraphs with 4 and 5 vertices, and present an outlook for further use of our results. We then prove the first part of the main theorems for motifs in Section 7.3 and for \sqrt{n} -optimal subgraphs in Section 7.4. The proofs of some lemmas introduced along the way are deferred to Section 7.5. Then, the proof of the second part of the main theorem can be found in Section 7.6. We finally show how the proofs for motifs can be adjusted to prove the theorems on graphlets in Section 7.7.

7.2 Main results

The key insight obtained in this chapter is that the creation of subgraphs is crucially affected by the following trade-off, inherently present in power-law networks. On the one hand, hubs contribute substantially to the number of subgraphs, because they are very well connected, and therefore potentially contribute to many graphlets or motifs. On the other hand, hubs are by definition rare. This should be contrasted with lower-degree vertices that occur more frequently, but typically take part in fewer connections and hence fewer subgraphs. Therefore, one may expect every subgraph to consist of a selection of vertices with specific degrees that ‘optimize’ this trade-off and hence maximize the probability that the subgraph occurs.

Write the probability that a motif H of k vertices is created between k uniformly chosen vertices as

$$\mathbb{P}(H \text{ present}) = \sum_D \mathbb{P}(H \text{ motif on degrees } D_1, \dots, D_k) \mathbb{P}(D_1, \dots, D_k), \quad (7.2.1)$$

where the sum is over all possible degrees on k vertices $\mathbf{D} = (D_i)_{i \in k}$, and $\mathbb{P}(D_1, \dots, D_k)$ denotes the probability that a randomly chosen set of k vertices has degrees D_1, \dots, D_k . Because of the power-law degree distribution, the last term decreases as a power of D_1, \dots, D_k . The first term in the sum, however, increases with D_1, \dots, D_k , since higher-degree vertices are part of more subgraphs. We show that for every subgraph, whether it is a motif or a graphlet there is a specific range of D_1, \dots, D_k that gives the maximal contribution to (7.2.1), sufficiently large to ignore all other degree ranges.

We show that there are only four possible ranges of degrees that maximize the term inside the sum in (7.2.1). These ranges are constant degrees, degrees proportional to $n^{(\tau-2)/(\tau-1)}$, degrees proportional to \sqrt{n} or degrees proportional to $n^{1/(\tau-1)}$. Observe that at this stage, these four ranges are merely an ansatz; rigorous underpinning for these choices comes later. For degrees proportional to $n^{1/(\tau-1)}$, the trade-off between the abundance of low-degree vertices and the connectedness of high-degree vertices is won by the high-degree vertices. Thus, intuitively, vertices in subgraphs that have the largest contribution from degrees proportional to $n^{1/(\tau-1)}$ should have more connections inside the subgraph than other vertices in subgraph. On the other hand, for vertices that have constant degree or degree proportional to $n^{(\tau-2)/(\tau-1)}$ in the optimal structure, the trade-off is ‘won’ by the lower-degree vertices. Intuitively, we therefore expect that these vertices are less well-connected inside the subgraph. Vertices with degrees proportional to \sqrt{n} form the middle ground, and are typically connected to vertices with similar degrees. There, the crucial observation is that pairs of vertices of degree of order \sqrt{n} are likely, though not certain, to have an edge between them in the erased configuration model.

7.2.1 An optimization problem

We now present the optimization problems that maximizes the term inside the sum in (7.2.1), first for motifs and later for graphlets. Let $H = (V_H, E_H)$ be a small, connected graph on $k > 2$ vertices. Denote the set of vertices of H that have degree one inside H by V_1 . Let \mathcal{P} be all partitions of $V_H \setminus V_1$ into three disjoint sets S_1, S_2, S_3 . This partition into S_1, S_2 and S_3 corresponds to these orders of magnitude: S_1 denotes the vertices with degree proportional to $n^{(\tau-2)/(\tau-1)}$, S_2 the ones with degrees proportional to $n^{1/(\tau-1)}$, and S_3 the vertices with degrees proportional to \sqrt{n} . The optimization problem finds the partition of the vertices into these three orders of magnitude such that the contribution to the number of motifs or graphlets is the largest. When a vertex in H has degree 1, its degree in the large graph G is typically small, it does not grow with n . Interestingly, vertices with degrees in these orders of magnitude are the only vertices that contribute to the number of motifs or graphlets, as we will prove later.

Given a partition \mathcal{P} , let E_{S_i} denote the number of edges in H between vertices in S_i , E_{S_i, S_j} the number of edges between vertices in S_i and S_j and E_{S_i, V_1} the number of edges between vertices in V_1 and S_i . We now define the optimization problem for motifs that is equivalent to optimizing the term inside the sum in (7.2.1) as

$$B^{(\text{sub})}(H) = \max_{\mathcal{P}} \left[|S_1| - |S_2| - \frac{2E_{S_1} + E_{S_1, S_3} + E_{S_1, V_1} - E_{S_2, V_1}}{\tau - 1} \right]. \quad (7.2.2)$$

The first two terms in the optimization problem give a positive contribution for all vertices in S_1 , vertices with relatively low degree, and a negative contribution for vertices in S_2 having high degrees. Therefore, the first two terms in the optimization problem capture that high-degree vertices are rare, and lower-degree vertices abundant. The last term gives a negative contribution for all edges between vertices with relatively low degrees in the motif. This captures the other part of the trade-off: high-degree vertices are much more likely to form edges with other vertices than low degree vertices. Since putting all vertices in S_3 yields zero, $B^{(\text{sub})}(H) \geq 0$.

For graphlets, we can define a similar optimization problem

$$B^{(\text{ind})}(H) = \max_{\mathcal{P}^{(\text{ind})}} \left[|S_1| - |S_2| - \frac{2E_{S_1} + E_{S_1, S_3} + E_{S_1, V_1} - E_{S_2, V_1}}{\tau - 1} \right],$$

s.t. $(u, v) \in E_H \quad \forall u \in S_2, v \in S_2 \cup S_3,$ (7.2.3)

where again $\mathcal{P}^{(\text{ind})}$ is a partition of $V_H \setminus V_1$ into three sets. The difference with (7.2.2) for motifs is the extra constraint in (7.2.3) which puts a restriction on the partitions into the three sets that are allowed. The extra constraint arises by the constraint in graphlets that H should be present as an induced subgraph, and ensures that edges that are not present in H are indeed not present in the subgraph. Since high-degree vertices are very likely to be connected, the constraint ensures that two vertices that are not connected in the graphlets cannot both have high degrees. Again, $B^{(\text{ind})}(H) \geq 0$ because putting all vertices in S_3 is a valid solution.

We now solve the two above optimization problems, to find the largest contributor to the number of motifs and graphlets. We first show that indeed the optimization problems in (7.2.2) and (7.2.3) find the typical degrees of vertices for any motif and any graphlet, and we show what the relation is between the optimization problems and the scaling of the number of motifs and graphlets. We then present a more detailed result for a special class of subgraphs, where the optimal contribution to (7.2.2) or (7.2.3) comes from $S_3 = V_H$, hence motifs and graphlets where all typical vertices have degrees proportional to \sqrt{n} . For this class, which contains for instance cliques of all sizes, we present sharp asymptotics.

7.2.2 General subgraphs

Let $S_1^{(\text{sub})}, S_2^{(\text{sub})}, S_3^{(\text{sub})}$ be a maximizer of (7.2.2). Furthermore, for any $\alpha = (\alpha_1, \dots, \alpha_k)$ such that $\alpha_i \in [0, 1/(\tau - 1)]$, we define

$$M_n^{(\alpha)}(\varepsilon) = \{(u_1, \dots, u_k) \in [n]^k : D_{u_i} \in [\varepsilon, 1/\varepsilon](\mu n)^{\alpha_i} \forall i \in [k]\}. \quad (7.2.4)$$

These are the sets of degrees such that D_1 is proportional to n^{α_1} and D_2 proportional to n^{α_2} and so on. Then, we denote the number of motifs with vertices in $M_n^{(\alpha)}(\varepsilon)$ by $N^{(\text{sub})}(H, M_n^{(\alpha)}(\varepsilon))$. Define the vector $\alpha^{(\text{sub})}$ as

$$\alpha_i^{(\text{sub})} = \begin{cases} (\tau - 2)/(\tau - 1) & i \in S_1^{(\text{sub})}, \\ 1/(\tau - 1) & i \in S_2^{(\text{sub})}, \\ \frac{1}{2} & i \in S_3^{(\text{sub})}, \\ 0 & i \in V_1. \end{cases} \quad (7.2.5)$$

For graphlets, we let $S_1^{(\text{ind})}, S_2^{(\text{ind})}, S_3^{(\text{ind})}$ be a maximizer of (7.2.3), and define $\alpha^{(\text{ind})}$ as in (7.2.5), replacing $S_i^{(\text{sub})}$ by $S_i^{(\text{ind})}$. By the interpretation of S_1, S_2 and S_3 in the optimization problem (7.2.2) and (7.2.3), sets of vertices in $M_n^{\alpha^{(\text{sub})}}(\varepsilon)$ or $M_n^{\alpha^{(\text{ind})}}$ intuitively contain a large number of subgraphs. The next theorem shows that this is correct, and computes the scaling of the number of motifs and graphlets:

Theorem 7.1 (General motifs and graphlets). *Let H be a motif on k vertices such that the solution to (7.2.2) is unique. Then, the following holds:*

(i) *For any ε_n such that $\lim_{n \rightarrow \infty} \varepsilon_n = 0$,*

$$\frac{N^{(\text{sub})}(H, M_n^{\alpha^{(\text{sub})}}(\varepsilon_n))}{N^{(\text{sub})}(H)} \xrightarrow{\mathbb{P}} 1. \quad (7.2.6)$$

(ii) *Furthermore, for any fixed $0 < \varepsilon < 1$,*

$$\frac{N^{(\text{sub})}(H, M_n^{\alpha^{(\text{sub})}}(\varepsilon))}{n^{\frac{3-\tau}{2}(k_{2+} + B^{(\text{sub})}(H)) + \frac{1}{2}k_1}} = f(\varepsilon)\Theta_{\mathbb{P}}(1) \quad (7.2.7)$$

for some function $f(\varepsilon)$ not depending on n . Here k_{2+} denotes the number of vertices in H of degree at least 2, and k_1 the number of degree one vertices in H .

For graphlets the same statements hold, replacing (sub) by (ind).

Thus, part (i) shows that indeed almost all motifs and graphlets are formed between vertices with the optimal degree structure that follows from (7.2.2), and part (ii) gives the scaling of the number of such subgraphs.

7.2.3 Sharp asymptotics for \sqrt{n} subgraphs

Now we study the special class of motifs for which the unique maximum of (7.2.2) is $S_3 = V_H$. By the above interpretation of S_1, S_2 and S_3 , we study motifs where the maximum contribution to the number of such motifs comes from vertices that have degrees proportional to \sqrt{n} in G . Examples of motifs that fall into this category are all complete graphs. Bipartite graphs on the other hand, do not fall into the \sqrt{n} -class motifs, since we can use the two parts of the bipartite graph as S_1 and S_2 in such a way that (7.2.2) results in a non-negative solution. The next theorem gives asymptotics for the number of such motifs:

Theorem 7.2 (Motifs with \sqrt{n} degrees). *Let H be a connected graph on k vertices with minimal degree 2 such that the solution to (7.2.2) is unique, and $B^{(\text{sub})}(H) = 0$. Then,*

$$\frac{N^{(\text{sub})}(H)}{n^{\frac{k}{2}(3-\tau)}} \xrightarrow{\mathbb{P}} A^{(\text{sub})}(H) < \infty, \quad (7.2.8)$$

with

$$A^{(\text{sub})}(H) = c^k \mu^{-\frac{k}{2}(\tau-1)} \int_0^\infty \cdots \int_0^\infty (x_1 \cdots x_k)^{-\tau} \prod_{(u,v) \in E_H} (1 - e^{-x_u x_v}) dx_1 \cdots dx_k. \quad (7.2.9)$$

We now state a similar theorem for graphlets where the unique optimal solution to (7.2.3) is $S_3 = V_H$. The optimization problem (7.2.3) is the same as (7.2.2) with an extra constraint, which is satisfied when $S_3 = V_H$. Therefore, if for a small graph H , (7.2.2) is optimized for $S_3 = V_H$, then (7.2.3) is also optimized for $S_3 = V_H$. Thus, the graphs H for which Theorem 7.2 can be applied are a subset of the graphs for which Theorem 7.3 can be applied. Therefore, complete graphs fall into the \sqrt{n} class graphlets as well. Section 7.2.4 shows which motifs on 4 and 5 vertices belong to the \sqrt{n} class. If the maximum contribution for H comes from \sqrt{n} vertices for counting motifs as well as graphlets, then $N^{(\text{ind})}(H)$ is of the same order of magnitude as $N^{(\text{sub})}(H)$, as the following theorem shows:

Theorem 7.3 (Graphlets with \sqrt{n} degrees). *Let H be a connected graph on k vertices with minimal degree 2 such that the solution to (7.2.3) is unique, and $B^{(\text{ind})}(H) = 0$. Then,*

$$\frac{N^{(\text{ind})}(H)}{n^{\frac{k}{2}(3-\tau)}} \xrightarrow{\mathbb{P}} A^{(\text{ind})}(H) < \infty, \quad (7.2.10)$$

with

$$\begin{aligned} A^{(\text{ind})}(H) &= c^k \mu^{-\frac{k}{2}(\tau-1)} \int_0^\infty \cdots \int_0^\infty (x_1 \cdots x_k)^{-\tau} \prod_{(u,v) \in E_H} (1 - e^{-x_u x_v}) \\ &\quad \times \prod_{(u,v) \notin E_H} e^{-x_u x_v} dx_1 \cdots dx_k. \end{aligned} \quad (7.2.11)$$

The difference between counting motifs and counting graphlets is visible in (7.2.9) and (7.2.11). In the erased configuration model, the probability that a vertex with degree D_i connects to a vertex with degree D_j can be approximated by $1 - \exp(-D_i D_j / L_n)$, where L_n denotes the sum of all degrees. When rescaling, this results in the factors $1 - e^{-x_u x_v}$ in (7.2.9) for all edges in motif H . When counting graphlets, we count induced subgraphs. Then, we also have to take into account that no other edges than the edges in H are allowed to be present. This gives the extra terms $e^{-x_u x_v}$ in (7.2.11).

7.2.4 Subgraphs on 4 and 5 vertices

We now apply Theorem 7.1 to characterize the optimal subgraph configurations of motifs and graphlets that consist of 4 or 5 vertices. For every partition of the vertices of H into S_1, S_2, S_3 , we compute the contribution to (7.2.2) and (7.2.3). In this way, we can find the partitions that maximize (7.2.2) and (7.2.3), and check whether this maximum is unique. If the maximum is indeed unique, then we can use Theorem 7.1 to calculate the scaling of the number of such motifs or graphlets. Figures 7.1 and 7.3 show the order of magnitude of the number of graphlets on 4 and 5 vertices obtained in this way, together with the optimizing sets of (7.2.3). Figure 7.2 shows the order of magnitude of the number of motifs on 4 vertices together with the optimizing sets of (7.2.2). For example, the optimal values of S_1, S_2 and S_3 for the motif in Figure 7.1d show that

$$B^{(\text{ind})}(H) = 2 - 1 + \frac{2 + 0 + 0 - 1}{\tau - 1} = 1 + \frac{1}{\tau - 1}. \quad (7.2.12)$$

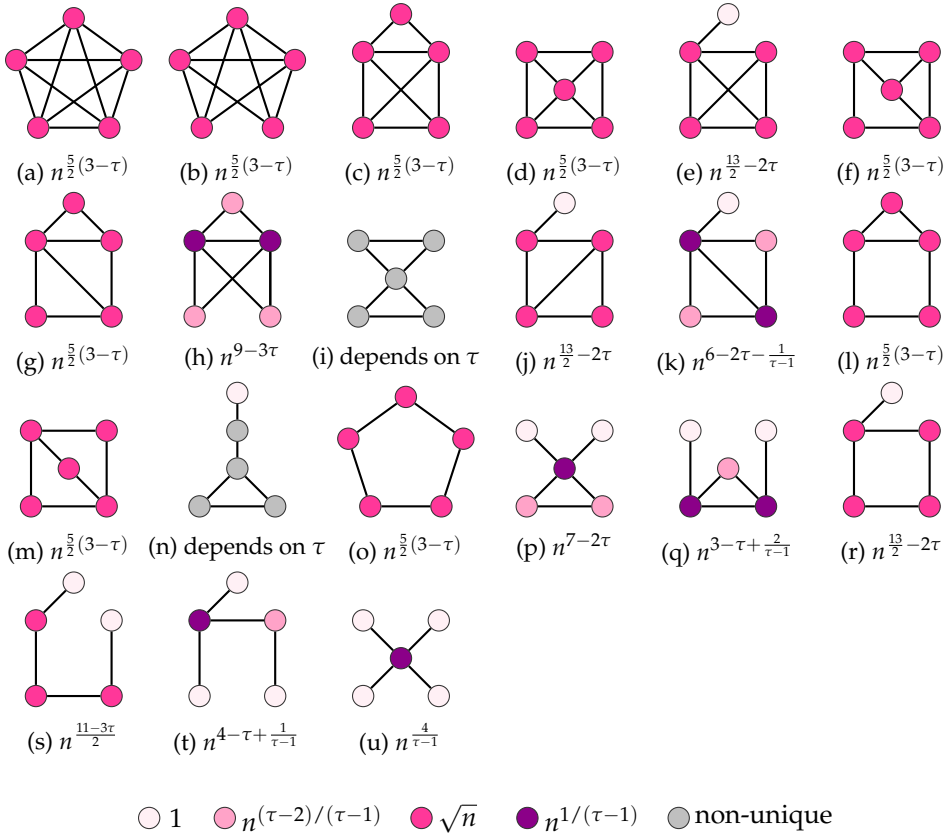


Figure 7.3: Graphlets. Order of magnitude of $N^{(\text{ind})}(H)$ for all connected graphs on 5 vertices (constants ignored). The vertex color indicates the optimal vertex degree.

Theorem 7.1 then shows that the correct scaling of the motif in Figure 7.1d is

$$n^{(3-\tau)/2(4-1/(\tau-1))+1/2} = n^{7-2\tau-\frac{1}{\tau-1}}. \tag{7.2.13}$$

The scaling of the other motifs and graphlets are computed similarly.

Figures 7.1 and 7.2 show the difference between counting motifs or counting graphlets. For example, Figure 7.1c shows that a square occurs $\Theta(n^{6-2\tau})$ times as a graphlet, whereas it occurs $\Theta(n^{6-2\tau} \log(n))$ times as a motif by Figure 7.2c. When we count the number of times the square occurs as a motif, we have to add the contributions from the graphlets in Figures 7.1a, 7.1b and 7.1c, that all contain a square, which also shows that the square occurs $\Theta(n^{6-2\tau} \log(n))$ times as a motif. The major contribution to the number of square motifs is from the graphlet in Figure 7.1b. This graphlet indeed contains a square, and occurs more frequently than the square occurs as a graphlet. In this manner we can infer the order of magnitude of the number of motifs from the number of graphlets. For this reason, Figure 7.3 is not shown for motifs. Using only Figure 7.3, we can argue that the graph in Figure 7.3m

occurs $\Theta(n^{\frac{5}{2}(3-\tau)})$ times as a graphlet, but $\Theta(n^{9-3\tau})$ times as a motif. Indeed, the graph in Figure 7.3h contains Figure 7.3m as a subgraph, and occurs more frequently.

Most motifs and graphlets in Figures 7.1, 7.2 and 7.3 satisfy the constraint in Theorem 7.1 that the solution to the optimization problem (7.2.2) or (7.2.3) should be unique. However, the gray vertices in Figure 7.1 and 7.2 do not have unique optimizers. Still these motifs and graphlets have ranges of degrees that give the major contribution to the number of such graphlets or motifs. The only difference is that these ranges are wider than for the vertices with unique maximizers. For example, for the square motif in Figure 7.2c the major contribution is from vertices where the degrees of vertices at each side of an edge $\{i, j\}$ in the square satisfy $D_i D_j = \Theta(n)$. Having all degrees proportional to \sqrt{n} therefore is one of the main contributors to the square motif. However, contributions where the bottom left vertex and the top right vertex have degrees proportional to n^α and the other two vertices have degrees $n^{1-\alpha}$ give an equal contribution for other values of α . Using that $D_i D_j$ follows a power-law distribution with exponent τ with an extra factor $\log(n)$ [110] then gives the extra factor $\log(n)$ in Figure 7.2b.

Another motif with gray vertices is the bow tie in Figure 7.3i. Unlike the square motif, this graphlet does satisfy the constraint of Theorem 7.1 that the optimal solution to (7.2.3) should be unique. However, the optimal solution depends on τ . For τ small, the maximum of (7.2.3) is uniquely attained at 0, so that for τ small, the major contribution is when all vertices are of degree $\Theta(\sqrt{n})$. On the other hand, when $\tau > 7/3$, (7.2.3) is minimized when S_1 contains all degree 2 vertices, and the middle vertex is in S_1 . This partition gives a contribution to (7.2.3) of

$$4 - 1 - \frac{2 \cdot 2}{\tau - 1} = \frac{3\tau - 7}{\tau - 1}, \quad (7.2.14)$$

which is larger than zero if $\tau > 7/3$. Thus, for τ larger than $7/3$, the major contribution is when the middle vertex has degree $n^{1/(\tau-1)}$, and the other vertices have degrees $n^{(\tau-2)/(\tau-1)}$. When $\tau < 7/3$, the major contribution is from all vertices of degrees \sqrt{n} , so that Theorem 7.3 can be applied. The graphlet of Figure 7.3n also has an optimal structure that depends on τ .

When the maximal contribution to a graphlet comes from vertices with degrees proportional to \sqrt{n} , the number of such graphlets converges to a constant when properly rescaled by Theorem 7.3. When the maximal contribution contains vertices in S_2 and S_1 , this may not be true anymore. For example, counting the claw graphlet of Figure 7.1e is very similar to counting the number of ways to choose three neighbors for every vertex. The only pairs of neighbors that we do not count, are neighbors that are connected themselves. This is only a small fraction of the pairs of neighbors, since Chapter 5 shows that the probability that two randomly chosen neighbors are connected tends to zero in the large-graph limit. Thus the number of claws is approximately equal to

$$\sum_{i \in [n]} \frac{1}{6} D_i (D_i - 1) (D_i - 2) \approx \sum_{i \in [n]} D_i^3. \quad (7.2.15)$$

Since the degrees are an i.i.d. sample from a power-law distribution, the sum of D_i^3 converges to a stable law when normalized properly. Thus, whereas in the \sqrt{n} -degree case, the leading order of the number of motifs or graphlets is constant (see

Theorems 7.2 and 7.2), when vertices of degrees proportional to $n^{1/(\tau-1)}$ contribute, the leading order may contain stable random variables. Therefore, the number of graphlets where the optimal solution to (7.2.3) comes from \sqrt{n} -degree vertices may be less volatile than when the optimal contribution also contains vertices with degrees proportional to $n^{1/(\tau-1)}$. We further investigate the fluctuations in subgraph counts in Chapter 8.

7.2.5 Discussion and outlook

Sampled and resulting degrees. Theorem 7.1 shows that most subgraphs occur between vertices of specific *sampled degrees*. The actual degree of a vertex i with sampled degree D_i may be lower than D_i , because multiple edges and self-loops incident to vertex i are removed in the erased configuration model. However, by Lemma 2.1, the sampled degree and the actual degree of vertex i are close, so that Theorem 7.1 also predicts that in the erased configuration model most subgraphs occur between vertices of degrees of the same order of magnitude as the sampled degrees.

Inhomogeneous random graph. All results in this chapter are proven for the erased configuration model. An interesting question is whether the results on the number of motifs and graphlets of Theorems 7.1-7.3 only hold for the erased configuration model, or whether they also apply to other models that create simple power-law random graphs, such as the rank-1 inhomogeneous random graph, defined in Section 1.1.4. In the rank-1 inhomogeneous random graph, conditionally on the weight sequence, the edge presences are independent. In the erased configuration model however, the edge presences are not independent, even when conditioning on the degree sequence. In this chapter, we show that these dependencies are sufficiently small to be neglected in the motif counts. We prove Theorems 7.1-7.3 for the erased configuration model using the approximation $\mathbb{P}_n(X_{ij} = 1) \approx 1 - \exp(-D_i D_j / L_n)$. Therefore, these theorems remain valid when we study the rank-1 inhomogeneous random graph where the connection probability of vertices with weight w_i and w_j is given by $1 - e^{-w_i w_j / (\mu n)}$.

Hyperbolic random graph. Another random graph model that creates simple power-law random graphs, is the hyperbolic random graph defined in Section 1.1.6. These graphs are very different from the erased configuration model and the rank-1 inhomogeneous random graph, because they contain geometry which creates more clustering. As mentioned before, all complete graphs satisfy the conditions of Theorems 7.2 and 7.3. Thus, a complete graph on k vertices occurs $\Theta(n^{k(3-\tau)/2})$ times as a motif or graphlet in erased configuration models. Interestingly, this is also true for hyperbolic random graphs when k is sufficiently large [27]. From the construction of the proof in [27], we can argue that the largest contribution to the number of cliques in hyperbolic random graphs then comes from vertices at radius $R/2$. These vertices have degrees proportional to \sqrt{n} , which is the same as the largest contribution for erased configuration model. It would be interesting to investigate the presence of other types of motifs in hyperbolic random graphs, and see whether these results are similar to the results for the erased configuration model, or if the geometric

structure in these graphs makes the largest contribution to the number of motifs or graphlets different. In particular, it would be interesting to see whether all other motifs that satisfy the conditions of Theorem 7.2 have the same order of magnitude in the hyperbolic random graph as in the erased configuration model.

Uniqueness of the optimum. Theorem 7.1 only holds when the optimum of (7.2.2), respectively (7.2.3), is unique. Figures 7.2 and 7.3 show that for most graphlets on 4 or 5 vertices, this is indeed the case. In Section 7.3, we show that (7.2.2) and (7.2.3) can both be interpreted as a piecewise linear optimization problem over the optimal degrees of the vertices that together form the subgraph. Thus, if the optimum is not unique, then it is attained by an entire range of degrees. In Section 7.3 we show that in this situation the optimum is attained for degrees such that $D_i D_j = \Theta(n)$ across some edges $\{i, j\}$. One such example is the diamond graphlet of Figure 7.1b discussed in Section 7.2.4, where the product of the degrees of the top left and the bottom right vertices scales as n . We believe that the number of motifs where the optimum is not unique scales as in Theorem 7.1 with some additional multiplicative factors of $\log(n)$. Proving this remains open for further research.

Graphlets on \sqrt{n} degrees. For motifs and graphlets where the most likely degrees are \sqrt{n} , we show in Theorems 7.2 and 7.3 that the rescaled number of motifs or graphlets converges to a constant, $A^{(\text{sub})}(H)$ and $A^{(\text{ind})}(H)$, respectively. Using the optimization problem (7.2.3) we can show that for example all cliques and all cycles are \sqrt{n} graphlets. The constants $A^{(\text{sub})}(H)$ and $A^{(\text{ind})}(H)$ are in general difficult to compute. It would be useful to have good estimates of these constants to be able to see which types of motifs occur more frequently. Furthermore, it would be interesting to investigate the convergence of motifs that do not satisfy the assumptions of Theorems 7.2 or 7.3. In Section 7.2.4, we saw that the normalized number of motifs may converge to a stable distribution for some motifs.

Automorphisms of H . An automorphism of a graph H is a map $V_H \mapsto V_H$ such that the resulting graph is an isomorphism of H . In Theorems 7.1, 7.2 and 7.3 we count automorphisms of H as separate copies of H , so that we may count multiple copies of H on one set of vertices. Since the number of vertices of H is fixed, and Theorem 7.1 only considers the scaling of the number of subgraphs, this does not change Theorems 7.1. Because Theorems 7.2 and 7.3 study the exact scaling of the number of subgraphs, to count the number of subgraphs without automorphisms, one should divide the results of Theorems 7.2 and 7.3 by the number of automorphisms of H .

Self-averaging. Another interesting question relates to the fluctuations of subgraph counts. When the degree distribution follows a power-law with exponent $\tau \in (2, 3)$, the number of motifs may not be self-averaging [171], that is

$$\limsup_{n \rightarrow \infty} \frac{\text{Var}(N^{(\text{sub})}(H))}{\mathbb{E}[N^{(\text{sub})}(H)]^2} \neq 0. \quad (7.2.16)$$

One such example is the triangle. In Chapter 8 we show that the number of triangles in a rank-1 inhomogeneous random graph is not self-averaging when τ is close to 3. Still, the triangle motif satisfies the conditions of Theorem 7.2, so that the number of triangles converges in probability to a constant. This indicates that when we generate configuration models with i.i.d. degrees, most realizations will have a number of triangles that is close to the value that is predicted in Theorem 7.2. Some realizations however, will have a number of triangles that is much larger or smaller than the value predicted in Theorem 7.2, which results in a large variance, making the number of triangles non self-averaging. However, when we first fix the degree sequence, Lemma 7.6 shows that the variance of the motif count is small. Therefore, the fluctuation in the number of motifs arises from the i.i.d. degree sequence, which was also observed in [171]. In particular, when we use the degrees of a real-world model as input in the erased configuration model, the number of motifs is self-averaging. This illustrates the importance of choosing the right null model. A null model with the same degree sequence as the original graph has less variability than a null model where we sample degrees i.i.d. from a power-law distribution with the same exponent as the original degree distribution. We further investigate the self-averaging properties of network motifs in Chapter 8.

Spectral moments. The number of loops of length k in a graph is related to the k -th spectral moment of the graph. For example, $\sum_{i \in [n]} \lambda_i^3 = \Delta/6$, where Δ denotes the number of triangles in the network, and λ_i the eigenvalues of the network [207]. Therefore, Theorem 7.1 could be translated to results on the spectral moments of the erased configuration model. The spectral moments as well as the largest eigenvalue have been investigated for power-law random graphs with a cutoff on the degrees at \sqrt{n} [62, 64, 182], where the largest eigenvalue was found to scale as $\lambda_1 \propto n^{(3-\tau)/2}$ for $2 < \tau < 5/2$. Note that Theorem 7.1 predicts that the number of triangles scales as $n^{3(3-\tau)/2}$, which scales as λ_1^3 . This suggests that the largest eigenvalue for power-law random graphs without cutoff scales similarly as when a cutoff on the degrees is imposed for $2 < \tau < 5/2$.

7.3 Maximum contribution: proof of Theorem 7.1

For every motif, there is a specific range of degrees that gives a major contribution to the number of motifs. We define an optimization problem that identifies these ranges of degrees. In Lemma 7.2 we show that the optimal solutions to these optimization problems have a highly particular structure. We then use these lemmas to prove Theorem 7.1 for motifs. We first investigate the dependence of the presence of the edges in the erased configuration model.

7.3.1 The probability of avoiding a subgraph

We relate $L_n = \sum_i D_i$, the total number of half-edges, to its expected value μn by defining the event

$$J_n = \left\{ |L_n - \mu n| \leq n^{2/\tau} \right\}. \quad (7.3.1)$$

By [112], $\mathbb{P}(J_n) \rightarrow 1$ as $n \rightarrow \infty$. When we condition on the degree sequence, we will condition on the event J_n , so that we can write $L_n = \mu n(1 + o(1))$. Furthermore, we denote by $X_{u,v}$ the indicator that an edge is present between vertices u and v . We let \mathbb{P}_n denote the conditional probability given the degree sequence, and \mathbb{E}_n the corresponding expectation. The presence of the edges that form a motif is not independent. Therefore, we use the following lemma which computes the probability of an edge not being present conditional on other edges not being present:

Lemma 7.1. *Fix $m \in \mathbb{N}$ and $\varepsilon > 0$. Let $(u_i, v_i)_{i \in [m+1]}$ such that $u_i, v_i \in [n]$ for all $i \in [m+1]$ and $(u_{m+1}, v_{m+1}) \neq (u_i, v_i)$ for all $i \in [m]$. Let*

$$\mathcal{E} = \{X_{u_i, v_i} = 0, \forall i \in [m]\}. \quad (7.3.2)$$

If $D_{u_i}, D_{v_i} \leq n^{1/(\tau-1)}/\varepsilon$ for $i \in [m+1]$, then on the event J_n ,

$$\mathbb{P}_n(X_{u_{m+1}, v_{m+1}} = 0 \mid \mathcal{E}) = O_{\mathbb{P}}\left(e^{-D_{u_{m+1}} D_{v_{m+1}} / 2L_n}\right). \quad (7.3.3)$$

Furthermore, when $D_{u_{m+1}} D_{v_{m+1}} \leq n/\varepsilon$,

$$\mathbb{P}_n(X_{u_{m+1}, v_{m+1}} = 0 \mid \mathcal{E}) = e^{-D_{u_{m+1}} D_{v_{m+1}} / L_n} \left(1 + o_{\mathbb{P}}\left(\frac{D_{u_{m+1}} D_{v_{m+1}}}{L_n} n^{-\frac{\tau-2}{\tau-1}}\right)\right). \quad (7.3.4)$$

Proof. For $m = 0$ the claim is proven in [108, Eq (4.6) and (4.9)], which states that for two vertices i and j with $D_i > D_j$,

$$\mathbb{P}_n(X_{i,j} = 0) = e^{-D_i D_j / L_n} + O(D_i^2 D_j / L_n^2), \quad (7.3.5)$$

and that by using [108, Eq. (4.5)]

$$\mathbb{P}_n(X_{i,j} = 0) \leq \prod_{i=1}^{D_j/2} \left(1 - \frac{D_j}{L_n - 2i - 1}\right) \leq e^{-D_i D_j / 2L_n} (1 + o(1)). \quad (7.3.6)$$

Thus we may assume that $m > 0$. $\Omega := \{u_i, v_i\}_{i \in [m]}$ may contain the same vertices multiple times. Let the number of distinct vertices in $\{u_i, v_i\}_{i \in [m]}$ be denoted by r , and let these distinct vertices be denoted by w_1, \dots, w_r . W.l.o.g. we assume that u_{m+1}, v_{m+1} correspond to w_r and w_{r-1} (if they are present in w_1, \dots, w_r at all). We now construct the erased configuration model G conditionally on the edges Ω not being present. We first pair the half-edges of the erased configuration model attached to w_1, \dots, w_r . First we pair all half-edges adjacent to w_1 . Since we condition on the edges Ω not being present, no half-edge from w_1 is allowed to pair to any of its neighbors in Ω . After that, we pair all remaining half-edges from w_2 , conditionally on these half-edges not connecting to one of the neighbors of w_2 in Ω , and so on. We continue until all edges in Ω have at least one incident vertex that has already been paired. Then, if we pair the rest of the half-edges, we know that none of the edges in Ω are present. Let B denote the number of vertices we have to pair before this happens. We never have to pair half-edges adjacent to u_{m+1} or to v_{m+1} (if they

are present in $\{u_i, v_i\}_{i \in [m]}$, since they are last in the ordering, and (u_{m+1}, v_{m+1}) is not present in $\{u_i, v_i\}_{i \in [m]}$. Therefore, all neighbors of u_{m+1} and v_{m+1} in Ω have already been paired before arriving at u_{m+1} or v_{m+1} . Let \hat{X}_{ij} denote the number of half-edges between i and j in the configuration model, so that the edge indicator of the erased configuration model can be written as $X_{ij} = \mathbb{1}\{\hat{X}_{ij} > 0\}$. Furthermore, let $\mathcal{F}_{\leq s} = \sigma(\{\hat{X}_{w_i, j}\}_{i \leq s, j \in [n]})$ be the information about the pairings that have been constructed up to time s .

After B pairings, we denote

$$\tilde{L}_n = L_n - 2 \sum_{i \in [B]} (D_{w_i} - \hat{X}_{w_i, w_i}) \quad (7.3.7)$$

and $D_{\tilde{u}_{m+1}} = D_{u_{m+1}} - \sum_{i \in [B]} \hat{X}_{i, u_{m+1}}$, and we define $D_{\tilde{v}_{m+1}}$ similarly. These quantities are all measurable on $\mathcal{F}_{\leq B}$. Then, the probability that u_{m+1} does not pair to v_{m+1} is the probability that \tilde{u}_{m+1} does not connect to \tilde{v}_{m+1} in a configuration model with \tilde{L}_n half-edges. Thus,

$$\mathbb{P}_n (X_{u_{m+1}, v_{m+1}} = 0 \mid \mathcal{F}_{\leq B}) = e^{-D_{\tilde{u}_{m+1}} D_{\tilde{v}_{m+1}} / \tilde{L}_n} + O\left(D_{\tilde{u}_{m+1}}^2 D_{\tilde{v}_{m+1}} / \tilde{L}_n^2\right), \quad (7.3.8)$$

where we have assumed w.l.o.g. that $D_{\tilde{u}_{m+1}} \geq D_{\tilde{v}_{m+1}}$. Since we are under the event J_n from (7.3.1), $\tilde{L}_n = L_n(1 + o(1))$. Now, we show that $D_{\tilde{u}_{m+1}} = D_{u_{m+1}}(1 + o_{\mathbb{P}}(n^{-(\tau-2)/(\tau-1)}))$. When we pair the half-edges adjacent to w_i , the probability that the j th half-edge pairs to u_{m+1} can be bounded as

$$\begin{aligned} \mathbb{P}_n (\text{jth half-edge pairs to } u_{m+1}) &\leq \frac{D_{u_{m+1}}}{L_n - 2j - 3 - 2 \sum_{s \in [i-1]} D_{w_s}} \\ &\leq K \frac{D_{u_{m+1}}}{L_n}, \end{aligned} \quad (7.3.9)$$

for some $K > 0$. We have to pair at most D_{w_i} half-edges, since some of the half-edges incident to w_i may have been used already in previous pairings. Therefore, we can stochastically dominate $\hat{X}_{w_i, u_{m+1}}$ by Y_i , where $Y_i \sim \text{Bin}(D_{w_i}, KD_{u_{m+1}}/L_n)$. Then,

$$\hat{X}_{u_{m+1}, w_i} = O_{\mathbb{P}}(D_{w_i} D_{u_{m+1}} / L_n). \quad (7.3.10)$$

We conclude that

$$D_{\tilde{u}_{m+1}} = D_{u_{m+1}} \left(1 - \sum_{i \in [B]} O_{\mathbb{P}}(D_{w_i} / L_n)\right) = D_{u_{m+1}}(1 + o_{\mathbb{P}}(n^{-\beta})), \quad (7.3.11)$$

where we let $\beta = (\tau - 2)/(\tau - 1)$. Similarly, $D_{\tilde{v}_{m+1}} = D_{v_{m+1}}(1 + o_{\mathbb{P}}(n^{-\beta}))$. Then, when $D_{u_{m+1}} D_{v_{m+1}} = O(n)$, (7.3.8) becomes

$$\begin{aligned} \mathbb{P}_n (X_{u_{m+1}, v_{m+1}} = 0 \mid \mathcal{F}_{B+1}) &= e^{-D_{u_{m+1}} D_{v_{m+1}} (1 + o_{\mathbb{P}}(n^{-\beta})) / L_n} + O\left(\frac{D_{u_{m+1}}^2 D_{v_{m+1}}}{L_n^2}\right) \\ &= e^{-D_{u_{m+1}} D_{v_{m+1}} / L_n} \left(1 + o_{\mathbb{P}}\left(\frac{D_{u_{m+1}} D_{v_{m+1}}}{L_n} n^{-\beta}\right)\right), \end{aligned} \quad (7.3.12)$$

where we used that $D_{u_{m+1}} = O(n^{1/(\tau-1)})$. By (7.3.6)

$$\mathbb{P}_n(X_{u_{m+1}, v_{m+1}} = 0 \mid \mathcal{F}_{B+1}) \leq e^{-D_{u_{m+1}} D_{v_{m+1}} / 2L_n} = O_{\mathbb{P}}\left(e^{-\frac{D_{u_{m+1}} D_{v_{m+1}}}{2L_n}}\right). \quad (7.3.13)$$

which proves the lemma. \square

7.3.2 An optimization problem

Assume that $D_i = \Theta(n^{\alpha_i})$ for some $\alpha_i \in [0, 1/(\tau-1)]$ for all i . Then, when $\alpha_i + \alpha_j \leq 1$, by (7.3.5)

$$\mathbb{P}_n(X_{ij} = 1) = \left(1 - e^{-\Theta(n^{\alpha_i + \alpha_j - 1})}\right)(1 + o(1)) = \Theta\left(n^{\alpha_i + \alpha_j - 1}\right). \quad (7.3.14)$$

When $\alpha_i + \alpha_j > 1$, by (7.3.6)

$$\mathbb{P}_n(X_{ij} = 1) = 1 - O\left(e^{-n^{\alpha_i + \alpha_j - 1}/2}\right). \quad (7.3.15)$$

For vertices i and j denote $w_{ij} = \min(n^{\alpha_i + \alpha_j - 1 - (\tau-2)/(\tau-1)}, 1)$. By Lemma 7.1, for any set of m edges,

$$\begin{aligned} \mathbb{P}_n(X_{u_1, v_1} = \dots = X_{u_m, v_m} = 0) &= \prod_{\alpha_{u_i} + \alpha_{v_i} < 1} (1 + o_{\mathbb{P}}(w_{u_i, v_i})) \Theta(1 - n^{\alpha_{u_i} + \alpha_{v_i} - 1}) \\ &\times \prod_{\alpha_{u_i} + \alpha_{v_i} > 1} O_{\mathbb{P}}\left(e^{-n^{(\alpha_{v_i} + \alpha_{u_i} - 1)/2}}\right). \end{aligned} \quad (7.3.16)$$

Let H be a motif on k vertices labeled as $1, \dots, k$ and edges $E_H = \{u_1, v_1\}, \dots, \{u_m, v_m\}$. Furthermore, let $G|_i$ be the induced subgraph of the erased configuration model G on vertices $i = (i_1, \dots, i_k)$. Then, we can write the probability that motif H is present on a specified subset of vertices $i = (i_1, \dots, i_k)$ as

$$\begin{aligned} \mathbb{P}_n(G|_i \supseteq E_H) &= 1 - \sum_{l=1}^m \mathbb{P}_n(X_{i_{u_l}, i_{v_l}} = 0) + \sum_{l \neq j} \mathbb{P}_n(X_{i_{u_l}, i_{v_l}} = X_{i_{u_j}, i_{v_j}} = 0) \\ &\quad - \sum_{l \neq j \neq w} \mathbb{P}_n(X_{i_{u_l}, i_{v_l}} = X_{i_{u_j}, i_{v_j}} = X_{i_{u_w}, i_{v_w}} = 0) + \dots \\ &\quad + (-1)^m \mathbb{P}_n(X_{i_{u_1}, i_{v_1}} = \dots = X_{i_{u_m}, i_{v_m}} = 0) \\ &= \Theta_{\mathbb{P}}\left(\prod_{(i,j) \in E_H: \alpha_i + \alpha_j < 1} n^{\alpha_i + \alpha_j - 1}\right), \end{aligned} \quad (7.3.17)$$

where we used that for $\alpha_i + \alpha_j < 1$

$$1 - (1 - n^{\alpha_i + \alpha_j - 1})(1 + o_{\mathbb{P}}(w_{ij})) = \Theta_{\mathbb{P}}(n^{\alpha_i + \alpha_j - 1}), \quad (7.3.18)$$

and that for $\alpha_i + \alpha_j > 1$

$$1 - O_{\mathbb{P}}\left(e^{-n^{(\alpha_i + \alpha_j - 1)/2}}\right) = 1 + o_{\mathbb{P}}(1). \quad (7.3.19)$$

The degrees are an i.i.d. sample from a power-law distribution. Therefore,

$$\mathbb{P}(D_1 \in [\varepsilon, 1/\varepsilon](\mu n)^\alpha) = \int_{\varepsilon(\mu n)^\alpha}^{1/\varepsilon(\mu n)^\alpha} cx^{-\tau} dx = K(\varepsilon)(\mu n)^{\alpha(1-\tau)} \quad (7.3.20)$$

for some constant $K(\varepsilon)$ not depending on n . The number of vertices with degrees in $[\varepsilon, 1/\varepsilon](\mu n)^\alpha$ is $\text{Binomial}(n, K(\varepsilon)(\mu n)^{\alpha(1-\tau)})$, so that the number of vertices with degrees in $[\varepsilon, 1/\varepsilon](\mu n)^\alpha$ is $\Theta_{\mathbb{P}}(n^{(1-\tau)\alpha+1})$. Let $M_n^{(\alpha)}$ be as in (7.2.4). Then,

$$\# \text{ sets of vertices with degrees in } M_n^{(\alpha)} = \Theta_{\mathbb{P}}(n^{k+(1-\tau)\sum_i \alpha_i}). \quad (7.3.21)$$

Combining (7.3.17) and (7.3.21) yields that

$$N^{(\text{sub})}(H, M_n^{(\alpha)}(\varepsilon)) = \Theta_{\mathbb{P}}\left(n^{k+(1-\tau)\sum_i \alpha_i} \prod_{(i,j) \in E_H: \alpha_i + \alpha_j < 1} n^{\alpha_i + \alpha_j - 1}\right). \quad (7.3.22)$$

The maximum contribution is obtained for α_i that maximize

$$\begin{aligned} & \max(1-\tau) \sum_i \alpha_i + \sum_{(i,j) \in E_H: \alpha_i + \alpha_j < 1} \alpha_i + \alpha_j - 1 \\ & \text{s.t. } \alpha_i \in [0, \frac{1}{\tau-1}] \forall i. \end{aligned} \quad (7.3.23)$$

The following lemma shows that this optimization problem attains its maximum for specific values of α :

Lemma 7.2 (Maximum contribution to motifs). *Let H be a connected graph on k vertices. If the solution to (7.3.23) is unique, then the optimal solution satisfies $\alpha_i \in \{0, \frac{\tau-2}{\tau-1}, \frac{1}{2}, \frac{1}{\tau-1}\}$ for all i . If it is not unique, then there exist at least 2 optimal solutions with $\alpha_i \in \{0, \frac{\tau-2}{\tau-1}, \frac{1}{2}, \frac{1}{\tau-1}\}$ for all i . In any optimal solution $\alpha_i = 0$ if and only if vertex i has degree one in H .*

Proof. Defining $\beta_i = \alpha_i - \frac{1}{2}$ yields for (7.3.23)

$$\max \frac{1-\tau}{2}k + (1-\tau) \sum_i \beta_i + \sum_{(i,j) \in E_H: \beta_i + \beta_j < 0} \beta_i + \beta_j, \quad (7.3.24)$$

over all possible values of $\beta_i \in [-\frac{1}{2}, \frac{3-\tau}{2(\tau-1)}]$. Then, we have to prove that $\beta_i \in \{-\frac{1}{2}, \frac{\tau-3}{2(\tau-1)}, 0, \frac{3-\tau}{2(\tau-1)}\}$ for all i in the optimal solution. Since (7.3.24) is a piecewise linear function in β , if (7.3.24) has a unique maximum, it must be attained at the boundary for β_i or at a border of one of the linear sections. Thus, any unique optimal value of β_i satisfies $\beta_i = -\frac{1}{2}$, $\beta_i = \frac{\tau-3}{2(\tau-1)}$ or $\beta_i + \beta_j = 0$ for some j . We ignore the constant factor of $(1-\tau)\frac{k}{2}$ in (7.3.24), since it does not influence the optimal β values. Rewriting (7.3.24) without the constant factor yields

$$\max \sum_i \beta_i (1-\tau + \# \text{ edges to } j \text{ with } \beta_j < -\beta_i). \quad (7.3.25)$$

The proof of the lemma then consists of three steps.

Step 1. Show that $\beta_i = -\frac{1}{2}$ if and only if vertex i has degree 1 in H in any optimal solution.

Step 2. Show that any unique solution does not have vertices i with $|\beta_i| \in (0, \frac{3-\tau}{2(\tau-1)})$.

Step 3. Show that any optimal solution that is not unique can be transformed into two different optimal solutions with $\beta_i \in \{-\frac{1}{2}, \frac{\tau-3}{2(\tau-1)}, 0, \frac{3-\tau}{2(\tau-1)}\}$ for all i .

Step 1. Let i be a vertex of degree 1 in H , and j be the neighbor of i . Let N_j denote the number of edges in H from j to other vertices v not equal to i with $\beta_v < -\beta_j$. The contribution from vertices i and j to (7.3.25) is

$$\beta_j(1 - \tau + N_j) + \beta_i(1 - \tau + \mathbb{1}_{\{\beta_i > -\beta_j\}}) + \beta_j \mathbb{1}_{\{\beta_i < -\beta_j\}}. \quad (7.3.26)$$

For any value of $\beta_j \in [-\frac{1}{2}, \frac{3-\tau}{2(\tau-1)}]$, this contribution is maximized when choosing $\beta_i = -\frac{1}{2}$. Thus, $\beta_i = -\frac{1}{2}$ in the optimal solution if the degree of vertex i is one.

Let i be a vertex with $d_i \geq 2$ in H , and suppose $\beta_i = -\frac{1}{2}$. Because the maximal value of $\beta = \frac{3-\tau}{2(\tau-1)}$, the contribution to (7.3.25) is

$$-\frac{1}{2}(1 - \tau + d_i) < 0. \quad (7.3.27)$$

Increasing β_i to $\frac{\tau-3}{2(\tau-1)}$ then gives a higher contribution. Thus, any vertex i with degree at least 2 in H must have $\beta_i = \frac{3-\tau}{2(\tau-1)}$ or $\beta_i + \beta_j = 0$ for some neighbor j in an optimal solution. Since $\beta_j \leq \frac{3-\tau}{2(\tau-1)}$ for all j this means that $\beta_i \geq \frac{\tau-3}{2(\tau-1)}$ when $d_i \geq 2$.

Step 2. Now we show that when the solution to (7.3.25) is unique, it is never optimal to have $|\beta| \in (0, \frac{3-\tau}{2(\tau-1)})$. Let

$$\tilde{\beta} = \min_{i: |\beta_i| > 0} |\beta_i|. \quad (7.3.28)$$

Let $N_{\tilde{\beta}^-}$ denote the number of vertices with their β value equal to $-\tilde{\beta}$, and $N_{\tilde{\beta}^+}$ the number of vertices with value $\tilde{\beta}$, where $N_{\tilde{\beta}^+} + N_{\tilde{\beta}^-} \geq 1$. Furthermore, let $E_{\tilde{\beta}^-}$ denote the number of edges from vertices with value $-\tilde{\beta}$ to other vertices j such that $\beta_j < \tilde{\beta}$, and $E_{\tilde{\beta}^+}$ the number of edges from vertices with value $\tilde{\beta}$ to other vertices j such that $\beta_j < -\tilde{\beta}$. Then, the contribution from these vertices to (7.3.25) is

$$\tilde{\beta} \left((1 - \tau) (N_{\tilde{\beta}^+} - N_{\tilde{\beta}^-}) + E_{\tilde{\beta}^+} - E_{\tilde{\beta}^-} \right). \quad (7.3.29)$$

Because we assume $\tilde{\beta}$ to be optimal, and the optimum to be unique, the value inside the brackets cannot equal zero. The contribution is linear in $\tilde{\beta}$ and it is the optimal contribution, and therefore $\tilde{\beta} \in \{0, \frac{3-\tau}{2(\tau-1)}\}$. This shows that $\beta_i \in \{\frac{\tau-3}{2(\tau-1)}, 0, \frac{3-\tau}{2(\tau-1)}\}$ for all i such that $d_i \geq 2$.

Step 3. If the solution to (7.3.25) is not unique, then by the same argument that leads to (7.3.29), there exist $\hat{\beta}_1, \dots, \hat{\beta}_s > 0$ for some $s \geq 1$ such that

$$\hat{\beta}_j \left((1 - \tau) (N_{\hat{\beta}_j^+} - N_{\hat{\beta}_j^-}) + E_{\hat{\beta}_j^+} - E_{\hat{\beta}_j^-} \right) = 0 \quad \forall j \in [s]. \quad (7.3.30)$$

Here we use the same notation as in (7.3.29). Setting all $\hat{\beta}_j = 0$ and setting all $\hat{\beta}_j = \frac{3-\tau}{2(\tau-1)}$ are both optimal solutions. Thus, if the solution to (7.3.25) is not unique, at least 2 solutions exist with $\beta_i \in \{\frac{\tau-3}{2(\tau-1)}, 0, \frac{3-\tau}{2(\tau-1)}\}$ for all i . \square

Proof of Theorem 7.1(ii) for motifs. Let $\alpha^{(\text{sub})}$ be the unique optimizer of (7.3.23). By Lemma 7.2, the maximal value of (7.3.23) is attained by partitioning $V_H \setminus V_1$ into the sets S_1, S_2, S_3 such that vertices in S_1 have $\alpha_i^{(\text{sub})} = \frac{\tau-2}{\tau-1}$, vertices in S_2 have $\alpha_i^{(\text{sub})} = \frac{1}{\tau-1}$, vertices in S_3 have $\alpha_i^{(\text{sub})} = \frac{1}{2}$ and vertices in V_1 have $\alpha_i^{(\text{sub})} = 0$. Then, the edges with $\alpha_i^{(\text{sub})} + \alpha_j^{(\text{sub})} < 1$ are edges inside S_1 , edges between S_1 and S_3 and edges from degree 1 vertices. If we denote the number of edges inside S_1 by E_{S_1} , the number of edges between S_1 and S_3 by E_{S_1, S_3} and the number of edges between V_1 and S_i by E_{S_i, V_1} , then we can rewrite (7.3.23) as

$$\begin{aligned} \max_{\mathcal{P}} \left[(1-\tau) \left(\frac{\tau-2}{\tau-1} |S_1| + \frac{1}{\tau-1} |S_2| + \frac{1}{2} |S_3| \right) + \frac{\tau-3}{\tau-1} E_{S_1} \right. \\ \left. + \frac{\tau-3}{2(\tau-1)} E_{S_1, S_3} - \frac{E_{S_1, V_1}}{\tau-1} - \frac{\tau-2}{\tau-1} E_{S_2, V_1} - \frac{1}{2} E_{S_3, V_1} \right], \end{aligned} \quad (7.3.31)$$

over all partitions \mathcal{P} of the vertices of H into S_1, S_2, S_3 . Using that $|S_3| = k - |S_1| - |S_2| - k_1$, $E_{S_3, 1} = k_1 - E_{S_1, 1} - E_{S_2, 1}$, where $k_1 = |V_1|$ and extracting a factor $(3-\tau)/2$ shows that this is equivalent to

$$\frac{1-\tau}{2} k + \max_{\mathcal{P}} \frac{(3-\tau)}{2} \left(|S_1| - |S_2| + \frac{\tau-2}{3-\tau} k_1 - \frac{2E_{S_1} + E_{S_1, S_3}}{\tau-1} - \frac{E_{S_1, V_1} - E_{S_2, V_1}}{\tau-1} \right), \quad (7.3.32)$$

Since k and k_1 are fixed and $3-\tau > 0$, we need to maximize

$$B^{(\text{sub})}(H) = \max_{\mathcal{P}} \left[|S_1| - |S_2| - \frac{2E_{S_1} + E_{S_1, S_3} + E_{S_1, V_1} - E_{S_2, V_1}}{\tau-1} \right]. \quad (7.3.33)$$

By (7.3.22), the contribution of the maximum is then given by

$$n^{\frac{3-\tau}{2}} (k + B^{(\text{sub})}(H)) + \frac{\tau-2}{2} k_1 = n^{\frac{3-\tau}{2}} (k_{2+} + B^{(\text{sub})}(H)) + \frac{1}{2} k_1, \quad (7.3.34)$$

which proves Theorem 7.1(ii) for motifs. \square

7.4 The number of motifs on \sqrt{n} -degrees

We prove Theorem 7.2 using the following lemma. We define

$$W_n^k(\varepsilon) = \{(u_1, \dots, u_k) : D_{u_i} \in [\varepsilon, 1/\varepsilon] \sqrt{\mu n} \quad \forall i \in [k]\}. \quad (7.4.1)$$

Note that $W_n^k(\varepsilon)$ is the special case of $M_n^{(\alpha)}(\varepsilon)$ defined in (7.2.4) where all values of α equal $1/2$. Then, we denote the number of motifs H with all degrees in $W_n^k(\varepsilon)$ by $N^{(\text{sub})}(H, W_n^k(\varepsilon))$.

Lemma 7.3 (Convergence of major contribution to motifs). *Let H be a connected graph on $k > 2$ vertices such that (7.2.2) is uniquely optimized at 0. Then,*

(i) The number of motifs with vertices in $W_n^k(\varepsilon)$ satisfies

$$\begin{aligned} \frac{N^{(\text{sub})}(H, W_n^k(\varepsilon))}{n^{\frac{k}{2}(3-\tau)}} &\xrightarrow{\mathbb{P}} c^k \mu^{-\frac{k}{2}(\tau-1)} \int_{\varepsilon}^{1/\varepsilon} \cdots \int_{\varepsilon}^{1/\varepsilon} (x_1 \cdots x_k)^{-\tau} \\ &\times \prod_{(i,j) \in E_H} (1 - e^{-x_i x_j}) dx_1 \cdots dx_k. \end{aligned} \quad (7.4.2)$$

(ii) $A^{(\text{sub})}(H)$ defined in (7.2.9) satisfies $A^{(\text{sub})}(H) < \infty$.

The proof of Lemma 7.3 can be found in Section 7.5. We now prove Theorem 7.2 using this lemma.

Proof of Theorem 7.2. We first study the expected number of motifs with vertices outside $W_n^k(\varepsilon)$. First, we investigate the expected number of motifs where vertex 1 has degree smaller than $\varepsilon\sqrt{\mu n}$. Because $\mathbb{P}_n(X_{ij} = 1) \leq \min(D_i D_j / L_n, 1)$, this contribution can be bounded as

$$\begin{aligned} &\mathbb{E} \left[N(H) \mathbb{1}_{\{D_1 < \varepsilon\sqrt{\mu n}\}} \right] \\ &\leq n^k \int_1^{\varepsilon\sqrt{\mu n}} \int_1^{\infty} \cdots \int_1^{\infty} (x_1 \cdots x_k)^{-\tau} \prod_{(i,j) \in E_H} \min\left(\frac{x_i x_j}{\mu n}, 1\right) dx_1 \cdots dx_k \\ &= n^k (\mu n)^{\frac{k}{2}(1-\tau)} \int_0^{\varepsilon} \int_0^{\infty} \cdots \int_0^{\infty} (t_1 \cdots t_k)^{-\tau} \prod_{(i,j) \in E_H} \min(t_i t_j, 1) dt_1 \cdots dt_k \\ &\leq K^{|E_H|} n^{\frac{k}{2}(3-\tau)} \mu^{\frac{k}{2}(1-\tau)} \int_0^{\varepsilon} \int_0^{\infty} \cdots \int_0^{\infty} (t_1 \cdots t_k)^{-\tau} \prod_{(i,j) \in E_H} (1 - e^{-t_i t_j}) dt_1 \cdots dt_k \\ &= O(n^{\frac{k}{2}(3-\tau)}) h_1(\varepsilon), \end{aligned} \quad (7.4.3)$$

where we used that $\min(1, x) \leq K(1 - e^{-x})$ for some $K > 0$, and $h_1(\varepsilon)$ is a function of ε . By Lemma 7.3(ii), $h(\varepsilon) \rightarrow 0$ as ε tends to zero. We can bound the situation where one of the other vertices has degree smaller than $\varepsilon\sqrt{n}$, or where one of the vertices has degree larger than \sqrt{n}/ε similarly. This results in

$$\mathbb{E}[N(H, \bar{W}_n^k(\varepsilon))] = O(n^{\frac{k}{2}(3-\tau)}) h(\varepsilon), \quad (7.4.4)$$

for some function $h(\varepsilon)$ not depending on n such that $h(\varepsilon) \rightarrow 0$ when $\varepsilon \rightarrow 0$. Then, by the Markov inequality,

$$N(H, \bar{W}_n^k(\varepsilon)) = h(\varepsilon) O_{\mathbb{P}}(n^{\frac{k}{2}(3-\tau)}). \quad (7.4.5)$$

Combining this with Lemma 7.3(i) gives

$$\begin{aligned} \frac{N(H)}{n^{\frac{k}{2}(3-\tau)}} &\xrightarrow{\mathbb{P}} c^k \mu^{-\frac{k}{2}(\tau-1)} \int_{\varepsilon}^{1/\varepsilon} \cdots \int_{\varepsilon}^{1/\varepsilon} (x_1, \dots, x_k)^{-\tau} \prod_{(u,v) \in E_H} (1 - e^{-x_u x_v}) dx_1 \cdots dx_k \\ &+ O_{\mathbb{P}}(h(\varepsilon)). \end{aligned} \quad (7.4.6)$$

Then letting $\varepsilon \rightarrow 0$ proves the theorem. \square

7.5 Major contribution to \sqrt{n} motifs

We first prove Lemma 7.3(i). We condition on the degree sequence, and compute the expected value and the variance of the number of motifs conditioned on the degrees in Lemmas 7.4 and 7.6. Then we take the i.i.d. degrees into account in Lemma 7.5. Together, these lemmas prove Lemma 7.3(i).

7.5.1 Conditional expectation

In this section, we study the expectation of the number of motifs, conditioned on the degrees. Let H be a motif on k vertices, labeled as $1, \dots, k$, and m edges. We denote the edges by $e_1 = \{u_1, v_1\}, \dots, e_m = \{u_m, v_m\}$.

Lemma 7.4 (Conditional expectation of motifs). *Let H be a motif such that (7.2.2) has a unique maximum, attained at 0. Then, under the event J_n as defined in (7.3.1)*

$$\mathbb{E}_n \left[N^{(\text{sub})}(H, W_n^k(\varepsilon)) \right] = \sum_{(i_1, \dots, i_k) \in W_n^k(\varepsilon)} \prod_{(j,k) \in E_H} (1 - e^{-D_{i_j} D_{i_k} / L_n}) (1 + o_{\mathbb{P}}(1)). \quad (7.5.1)$$

Proof. Let $\mathbf{i} = (i_1, \dots, i_k)$, and let $G|_{\mathbf{i}}$ again denote the induced subgraph of G on vertices \mathbf{i} . We can use (7.3.17) to show that

$$\begin{aligned} \mathbb{E}_n \left[N^{(\text{sub})}(H, W_n^k(\varepsilon)) \right] &= \sum_{\mathbf{i} \in W_n^k(\varepsilon)} \mathbb{P}_n(G|_{\mathbf{i}} \supseteq E_H) \\ &= (1 + o_{\mathbb{P}}(1)) \sum_{\mathbf{i} \in W_n^k(\varepsilon)} \prod_{l=1}^m \left(1 - \mathbb{P}_n(X_{i_{u_l}, i_{v_l}} = 0) \right). \end{aligned} \quad (7.5.2)$$

Because $D_i D_j = O(n)$ and $L_n = \mu n(1 + o(1))$ under J_n , by (7.3.5)

$$\mathbb{P}_n(X_{ij} = 1) = 1 - e^{-D_i D_j / L_n} + O\left(\frac{D_i^2 D_j}{L_n^2}\right) = (1 + o(1)) \left(1 - e^{-D_i D_j / L_n}\right). \quad (7.5.3)$$

This results in

$$\mathbb{E}_n \left[N^{(\text{sub})}(H, W_n^k(\varepsilon)) \right] = (1 + o_{\mathbb{P}}(1)) \sum_{\mathbf{i} \in W_n^k(\varepsilon)} \prod_{(j,k) \in E_H} (1 - e^{-D_{i_j} D_{i_k} / L_n}). \quad (7.5.4)$$

□

7.5.2 Convergence of conditional expectation

We now consider the convergence of the expectation of the number of subgraphs conditioned on the degrees when we take the randomness of the i.i.d. degrees into account.

Lemma 7.5 (Convergence of conditional expectation of \sqrt{n} motifs). *Let H be a motif such that (7.2.2) has a unique maximizer, and the maximum is attained at 0. Then,*

$$\frac{\mathbb{E}_n[N^{(\text{sub})}(H, W_n^k(\varepsilon))]}{n^{\frac{k}{2}(3-\tau)}} \xrightarrow{\mathbb{P}} c^k \mu^{-\frac{k}{2}(\tau-1)} \int_{\varepsilon}^{1/\varepsilon} \cdots \int_{\varepsilon}^{1/\varepsilon} (x_1 \cdots x_k)^{-\tau} \times \prod_{(u,v) \in E_H} (1 - e^{-x_u x_v}) dx_1 \cdots dx_k. \quad (7.5.5)$$

Proof. Let $|E_H| = m$ and denote the edges of H by $(u_1, v_1), \dots, (u_m, v_m)$. Define

$$g(t_1, \dots, t_k) := \prod_{(u,v) \in E_H} (1 - e^{-t_u t_v}). \quad (7.5.6)$$

Taylor expanding $1 - e^{-xy}$ on $[\varepsilon, 1/\varepsilon]$ yields

$$1 - e^{-xy} = \sum_{i=1}^s \frac{(xy)^i}{i!} (-1)^i + O\left(\frac{\varepsilon^{-s}}{(s+1)!}\right). \quad (7.5.7)$$

Since g is a bounded function on $F = [\varepsilon, 1/\varepsilon]^m$, for any $\eta > 0$, we can find s_1, \dots, s_m such that

$$\begin{aligned} g(t_1, \dots, t_k) &= \sum_{i_1=1}^{s_1} \cdots \sum_{i_m=1}^{s_m} \left((-1)^{i_1} \frac{t_{u_1}^{i_1} t_{v_1}^{i_1}}{i_1!} \cdots (-1)^{i_m} \frac{t_{u_m}^{i_m} t_{v_m}^{i_m}}{i_m!} \right) + O(\eta) \\ &= \sum_{i_1=1}^{s_1} \cdots \sum_{i_m=1}^{s_m} \left(\frac{(-1)^{i_1+\dots+i_m}}{i_1! \cdots i_m!} t_1^{\gamma_1} t_2^{\gamma_2} \cdots t_k^{\gamma_k} \right) + O(\eta), \end{aligned} \quad (7.5.8)$$

where

$$\gamma_j := \gamma_j(i_1, \dots, i_k) = \sum_l i_l \mathbb{1}_{\{u_l=j \text{ or } v_l=j\}}. \quad (7.5.9)$$

Let $M^{(n)}$ denote the random measure

$$M^{(n)}([a, b]) = \frac{(\mu n)^{\frac{1}{2}(\tau-1)}}{n} \sum_{i=1}^n \mathbb{1}_{\{D_i \in \sqrt{\mu n}[a, b]\}}. \quad (7.5.10)$$

Because the number of vertices with degrees in $[a, b]$ is binomially distributed,

$$M^{(n)}([a, b]) \xrightarrow{\mathbb{P}} (\mu n)^{\frac{1}{2}(\tau-1)} \int_a^{b/\sqrt{\mu n}} c x^{-\tau} dx = c \int_a^b x^{-\tau} dx := \lambda([a, b]). \quad (7.5.11)$$

Let $N^{(n)}$ denote the product measure $M^{(n)} \times M^{(n)} \times \cdots \times M^{(n)}$ (k times). Then, choosing $\eta = \varepsilon^{k+1}$ in (7.5.8) together with Lemma 7.4 yields

$$\begin{aligned} \frac{\mathbb{E}_n[N^{(\text{sub})}(H, W_n^k(\varepsilon))]}{n^{\frac{k}{2}(3-\tau)} \mu^{\frac{k}{2}(1-\tau)}} &= \int_F g(t_1, \dots, t_k) dN^{(n)}(t_1, \dots, t_k) \\ &= \int_F \sum_{i_1=1}^{s_1} \cdots \sum_{i_m=1}^{s_m} \left(\frac{(-1)^{i_1+\dots+i_m}}{i_1! \cdots i_m!} t_1^{\gamma_1} t_2^{\gamma_2} \cdots t_k^{\gamma_k} \right) + O(\varepsilon^{k+1}) dN^{(n)}(t_1, \dots, t_k) \\ &= \sum_{i_1=1}^{s_1} \cdots \sum_{i_m=1}^{s_m} \frac{(-1)^{i_1+\dots+i_m}}{i_1! \cdots i_m!} \int_{\varepsilon}^{1/\varepsilon} t_1^{\gamma_1} dM^{(n)}(t_1) \cdots \int_{\varepsilon}^{1/\varepsilon} t_k^{\gamma_k} dM^{(n)}(t_k) + O(\varepsilon). \end{aligned} \quad (7.5.12)$$

By Lemma 5.5 for any γ

$$\int_{\varepsilon}^{1/\varepsilon} x^{\gamma} dM^{(n)}(x) \xrightarrow{\mathbb{P}} \int_{\varepsilon}^{1/\varepsilon} x^{\gamma} d\lambda(x). \quad (7.5.13)$$

Combining this with (7.5.12) results in

$$\begin{aligned} & \frac{\mathbb{E}_n[N^{(\text{sub})}(H, W_n^k(\varepsilon))]}{f^{(\text{sub})}(n, H)\mu^{\frac{k}{2}(1-\tau)}} \\ & \xrightarrow{\mathbb{P}} \sum_{i_1=1}^{s_1} \cdots \sum_{i_m=1}^{s_m} \frac{(-1)^{i_1+\cdots+i_m}}{i_1! \cdots i_m!} \int_{\varepsilon}^{1/\varepsilon} t_1^{\alpha_1} d\lambda(t_1) \cdots \int_{\varepsilon}^{1/\varepsilon} t_k^{\alpha_k} d\lambda(t_k) + O(\varepsilon) \\ & = \int_F \sum_{i_1=1}^{s_1} \cdots \sum_{i_m=1}^{s_m} \frac{(-1)^{i_1+\cdots+i_m}}{i_1! \cdots i_m!} t_1^{\alpha_1} \cdots t_k^{\alpha_k} d\lambda(t_1) \cdots d\lambda(t_k) + O(\varepsilon) \\ & = \int_F g(t_1, \dots, t_k) d\lambda(t_1) \cdots d\lambda(t_k) + O(\varepsilon). \end{aligned} \quad (7.5.14)$$

Then, by (7.5.11)

$$\begin{aligned} & \frac{\mathbb{E}_n[N^{(\text{sub})}(H, W_n^k(\varepsilon))]}{f^{(\text{sub})}(n, H)} \\ & \xrightarrow{\mathbb{P}} c^k \mu^{-\frac{k}{2}(\tau-1)} \int_{\varepsilon}^{1/\varepsilon} \cdots \int_{\varepsilon}^{1/\varepsilon} (t_1 \cdots t_k)^{-\tau} g(t_1, \dots, t_k) dt_1 \cdots dt_k, \end{aligned} \quad (7.5.15)$$

which proves the claim. \square

7.5.3 Conditional variance

In this section, we again condition on the degrees. The following lemma shows that the variance of the number of motifs is small compared to the expected value:

Lemma 7.6 (Conditional variance for motifs). *Let H be a motif such that (7.2.2) has a unique maximum attained at 0. Then, under the event J_n defined in (7.3.1)*

$$\frac{\text{Var}_n \left(N^{(\text{sub})}(H, W_n^k(\varepsilon)) \right)}{\mathbb{E}_n \left[N^{(\text{sub})}(H, W_n^k(\varepsilon)) \right]^2} \xrightarrow{\mathbb{P}} 0. \quad (7.5.16)$$

Proof. By Lemma 7.5,

$$\mathbb{E}_n [N^{(\text{sub})}(H, W_n^k(\varepsilon))]^2 = \Theta(n^{(3-\tau)k}). \quad (7.5.17)$$

Thus, we need to prove that the variance is small compared to $n^{(3-\tau)k}$. Denote $\mathbf{i} = (i_1, \dots, i_k)$ and $\mathbf{j} = (j_1, \dots, j_k)$. We write the variance as

$$\begin{aligned} \text{Var}_n \left(N^{(\text{sub})}(H, W_n^k(\varepsilon)) \right) &= \sum_{\mathbf{i} \in W_n^k(\varepsilon)} \sum_{\mathbf{j} \in W_n^k(\varepsilon)} \left(\mathbb{P}_n(G|\mathbf{i}, G|\mathbf{j} \supseteq E_H) \right. \\ & \quad \left. - \mathbb{P}_n(G|\mathbf{i} \supseteq E_H) \mathbb{P}_n(G|\mathbf{j} \supseteq E_H) \right). \end{aligned} \quad (7.5.18)$$

This splits into various cases, depending on the overlap of i and j . When i and j do not overlap, by (7.3.17)

$$\begin{aligned}
 & \sum_{i \in W_n^k(\varepsilon)} \sum_{j \in W_n^k(\varepsilon)} \mathbb{P}_n(G|_i, G|_j \supseteq E_H) - \mathbb{P}_n(G|_i \supseteq E_H) \mathbb{P}_n(G|_j \supseteq E_H) \\
 &= \sum_{i \in W_n^k(\varepsilon)} \sum_{j \in W_n^k(\varepsilon)} (1 + o_{\mathbb{P}}(1)) \prod_{l=1}^m (1 - \mathbb{P}_n(X_{i_{u_l}, i_{v_l}} = 0)) \prod_{l=1}^m (1 - \mathbb{P}_n(X_{j_{u_l}, j_{v_l}} = 0)) \\
 &\quad - (1 + o_{\mathbb{P}}(1)) \prod_{l=1}^m (1 - \mathbb{P}_n(X_{i_{u_l}, i_{v_l}} = 0)) \prod_{l=1}^m (1 - \mathbb{P}_n(X_{j_{u_l}, j_{v_l}} = 0)) \\
 &= \mathbb{E}_n[N^{(\text{sub})}(H, W_n^k(\varepsilon))]^2 o_{\mathbb{P}}(1).
 \end{aligned} \tag{7.5.19}$$

The other contributions are when i and j overlap. In this situation, we use the bound $\mathbb{P}_n(X_{ij} = 1) \leq 1$. When i and j overlap on $s \geq 1$ vertices, we bound the contribution to (7.5.18) as

$$\begin{aligned}
 \sum_{i, j \in W_n^k(\varepsilon): |\{i, j\}| = 2k-s} \mathbb{P}_n(G|_i, G|_j \supseteq E_H) &\leq |\{i : D_i \in \sqrt{\mu n}[\varepsilon, 1/\varepsilon]\}|^{2k-s} \\
 &= O_{\mathbb{P}}\left(n^{\frac{(3-\tau)(2k-s)}{2}}\right),
 \end{aligned} \tag{7.5.20}$$

which is smaller than $n^{(3-\tau)k}$, as required. □

Proof of Lemma 7.3. We start by proving part (i). By Lemma 7.6 and Chebyshev's inequality, conditionally on the degrees

$$N^{(\text{sub})}(H, W_n^k(\varepsilon)) = \mathbb{E}_n[N^{(\text{sub})}(H, W_n^k(\varepsilon))](1 + o_{\mathbb{P}}(1)). \tag{7.5.21}$$

Combining this with Lemma 7.5 proves Lemma 7.3(i). Lemma 7.3(ii) is a direct consequence of Lemma 7.8 where $|S_3^*| = k$. □

7.6 Major contribution to general motifs

In this section we prove Theorem 7.1(i) for motifs. We start by introducing some notation. For any $W \subseteq V_H$, we denote by $d_{i,W}$ the number of edges from vertex i to vertices in W . Let H be a connected subgraph, such that the optimum of (7.2.2) is unique, and let S_1^*, S_2^* and S_3^* be the optimal partition. Then, we define

$$\zeta_i = \begin{cases} 1 & \text{if } d_i = 1, \\ d_{i, V_1} + d_{i, S_1^*} + d_{i, S_3^*} & \text{if } i \in S_1^*, \\ d_{i, V_1} + d_{i, S_1^*} & \text{if } i \in S_3^*, \\ d_{i, V_1} & \text{if } i \in S_2^*. \end{cases} \tag{7.6.1}$$

The following lemma states several properties of the number of edges between vertices in the different optimizing sets:

Lemma 7.7. *Let H be a connected subgraph, such that the optimum of (7.2.2) is unique, and let S_1^*, S_2^* and S_3^* be the optimal partition. Then the following holds:*

- (i) $\zeta_i \leq 1$ for $i \in S_1^*$.
- (ii) $d_{i,S_1^*} + \zeta_i \geq 2$ for $i \in S_2^*$.
- (iii) $\zeta_i \leq 1$ and $d_{i,S_3^*} + \zeta_i \geq 2$ for $i \in S_3^*$.

Proof. Suppose $i \in S_1^*$. Now consider the partition $\hat{S}_1 = S_1^* \setminus i$, $\hat{S}_2 = S_2^*$, $S_3 = S_3^* \cup i$. Then, $E_{\hat{S}_1} = E_{S_1^*} - d_{i,S_1^*}$ and $E_{\hat{S}_1, \hat{S}_3} = E_{S_1^*, S_3^*} + d_{i,S_1^*} - d_{i,S_3^*}$. Furthermore, $E_{\hat{S}_1, 1} = E_{S_1^*, 1} - d_{i,1}$ and $E_{\hat{S}_2, 1} = E_{S_2^*, 1}$. Because the partition into S_1^*, S_2^* and S_3^* achieves the unique optimum of (7.2.2)

$$\begin{aligned} & |S_1^*| - |S_2^*| - \frac{2E_{S_1^*} - E_{S_1^*, S_3^*} + E_{S_2^*, 1} - E_{S_1^*, 1}}{\tau - 1} \\ & > |S_1^*| - 1 - |S_2^*| - \frac{2E_{S_1^*} - E_{S_1^*, S_3^*} - d_{i,S_1^*} - d_{i,S_3^*} + E_{S_2^*, 1} - E_{S_1^*, 1} + d_{i,V_1}}{\tau - 1}, \end{aligned} \quad (7.6.2)$$

which reduces to

$$d_{i,S_3^*} + d_{i,S_1^*} + d_{i,V_1} = \zeta_i < \tau - 1. \quad (7.6.3)$$

Using that $\tau \in (2, 3)$ then yields $d_{i,S_3^*} + d_{i,S_1^*} + d_{i,V_1} \leq 1$.

Similar arguments give the other inequalities. For example, for $i \in S_3^*$, considering the partition where i is moved to S_1^* gives the inequality $d_{i,S_3^*} + d_{i,S_1^*} + d_{i,V_1} \geq 2$, and considering the partition where i is moved to S_2^* results in the inequality $d_{i,S_1^*} + d_{i,V_1} \leq 1$, so that $\zeta_i \leq 1$. \square

We now show that two integrals related to the solution of the optimization problem (7.2.2) are finite, using Lemma 7.7. These integrals are the key ingredient in proving Theorem 7.1(i) for motifs.

Lemma 7.8. *Suppose that the maximum in (7.2.2) is uniquely attained with $|S_3^*| = s > 0$, and say $S_3^* = \{1, \dots, s\}$. Then*

$$\int_0^\infty \cdots \int_0^\infty \prod_{i \in [s]} x_i^{-\tau + \zeta_i} \prod_{(u,v) \in E_{S_3^*}} \min(x_u x_v, 1) dx_s \cdots dx_1 < \infty. \quad (7.6.4)$$

Proof. The integral (7.6.4) consists of multiple regions. One region is $x_1, \dots, x_s \geq 1$. Then, since $-\tau + \zeta_i < -1$ by Lemma 7.7, this integral results in

$$\int_1^\infty \cdots \int_1^\infty \prod_{j \in [s]} x_j^{-\tau + \zeta_j} dx_1 \cdots dx_s < \infty. \quad (7.6.5)$$

Another region is $x_1, \dots, x_s \in [0, 1]$. Since by Lemma 7.7, any vertex in S_3^* has $\zeta_i + d_{i,S_3^*} \geq 2$, this integral can be bounded as

$$\begin{aligned} & \int_0^1 \cdots \int_0^1 \prod_{j \in [s]} x_j^{-\tau + \zeta_j} \prod_{(u,v) \in E_{S_3^*}} x_u x_v dx_1 \cdots dx_s \\ & \leq \int_0^1 \cdots \int_0^1 (x_1 \cdots x_s)^{2-\tau} dx_1 \cdots dx_s < \infty. \end{aligned} \quad (7.6.6)$$

The other regions can be described by the union of all sets $U \subset S_3^*$ such that the integral runs from 1 to ∞ for all $i \in U$, and from 0 to 1 for all $i \in \bar{U} = S_3^* \setminus U$. In such a region, $\min(x_i x_j, 1) = x_i x_j$ when $i, j \notin U$, and $\min(x_i, x_j) = 1$ when $i, j \in U$. W.l.o.g. assume $U = \{1, \dots, t\}$ for some $t \geq 1$. Then, the contribution to (7.6.4) from the region described by U can be written as

$$\int_1^\infty \cdots \int_1^\infty \prod_{i \in [t]} x_i^{-\tau + \zeta_j} \prod_{j=t+1}^s h(j, \mathbf{x}) dx_t \cdots dx_1, \quad (7.6.7)$$

where $\mathbf{x} = (x_i)_{i \in [t]}$ and

$$h(j, \mathbf{x}) = \int_0^1 x_i^{-\tau + \zeta_i + d_{j, \bar{U}}} \prod_{i \in U: (i, j) \in E_H} \min(x_i x_j, 1) dx_j \quad (7.6.8)$$

for $j \in \{t, t+1, \dots, k\}$. The integral in $h(j, \mathbf{x})$ consists of multiple regions, depending on whether $x_i x_j < 1$ or not. Suppose vertex $i \in \bar{U}$ is connected to vertices $1, 2, \dots, l \in U$, and assume that $x_1 < x_2 < \dots < x_l$. Then,

$$\begin{aligned} h(j, \mathbf{x}) &= \int_0^1 x_j^{-\tau + \zeta_j + d_{j, \bar{U}}} \min(x_j x_1, 1) \min(x_j x_2, 1) \cdots \min(x_j x_l, 1) dx_j \\ &= \int_0^{1/x_l} x_j^{-\tau + \zeta_j + l + d_{j, \bar{U}}} x_1 \cdots x_l dx_j + \int_{1/x_l}^{1/x_{l-1}} x_j^{-\tau + \zeta_j + l - 1 + d_{j, \bar{U}}} x_1 \cdots x_{l-1} dx_j \\ &\quad + \cdots + \int_{1/x_1}^1 x_j^{-\tau + \zeta_j + d_{j, \bar{U}}} dx_j \\ &= C_1 x_1 x_2 \cdots x_l^{\tau - \zeta_j - l - d_{j, \bar{U}}} + C_2 (x_1 \cdots x_{l-1}^{\tau - \zeta_j - l - d_{j, \bar{U}} + 1} - x_1 x_2 \cdots x_l^{\tau - \zeta_j - l - d_{j, \bar{U}}}) \\ &\quad + \cdots + C_l (1 - x_1^{\tau - \zeta_j - d_{j, \bar{U}} + 1}) \end{aligned} \quad (7.6.9)$$

for some constants $C_1 > C_2 > \cdots > C_l$. Since $\zeta_j + d_{j, \bar{U}} + l - \tau = \zeta_j + d_{i, S_3^*} - \tau > -1$ by Lemma 7.7, the first integral is finite. Because $1 < x_1 < x_2 < \dots < x_l$, this results in

$$h(j, \mathbf{x}) \leq K \int_0^{1/x_l} x_j^{-\tau + \zeta_j + d_{j, S_3^*}} x_1 \cdots x_l dx_j, \quad \forall j, \quad (7.6.10)$$

for some constant $K > 0$. Assume that $x_1 < x_2 < \dots < x_t$. We let

$$u_j = \max\{i \mid i \in U, (i, j) \in E_H\} \quad (7.6.11)$$

for all $j \in \bar{U}$ such that $d_{j, U} \geq 1$. Furthermore, let

$$f(j) = \begin{cases} 1/x_{u_j} & \text{if } d_{j, U} \geq 1, \\ 1 & \text{else,} \end{cases} \quad (7.6.12)$$

for all $j \in \bar{U}$. Then (7.6.10) results in

$$\int_1^\infty \int_{x_1}^\infty \cdots \int_{x_{t-1}}^\infty \prod_{i=1}^t x_i^{-\tau + \zeta_j} \prod_{j=t+1}^s h(j, \mathbf{x}) dx_t \cdots dx_1$$

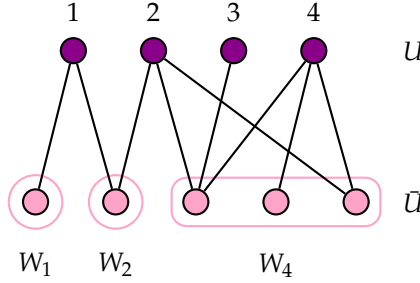


Figure 7.4: Illustration of the sets W_i . For clarity of the picture, edges in U and in \bar{U} are not displayed.

$$\begin{aligned}
 &\leq \tilde{K} \int_1^\infty \int_{x_1}^\infty \cdots \int_{x_{t-1}}^\infty \prod_{i=1}^t x_i^{-\tau + \zeta_i + d_{i,\bar{U}}} \prod_{j=t+1}^s \left(\int_0^{f(j)} x_j^{-\tau + \zeta_j + d_{j,S_3^*}} dx_j \right) dx_t \cdots dx_1 \\
 &= \tilde{K} \int_1^\infty \int_{x_1}^\infty \cdots \int_{x_{t-1}}^\infty \prod_{i=1}^t x_i^{-\tau + \zeta_i + d_{i,\bar{U}}} \prod_{j=t+1}^s f(j)^{-\tau + 1 + \zeta_j + d_{j,S_3^*}} dx_t \cdots dx_1, \quad (7.6.13)
 \end{aligned}$$

for some constant $\tilde{K} > 0$. Let $W_i = \{j \in \bar{U} : u_j = i\}$ for $i \in [t]$ (for an illustration, see Figure 7.4) and let $\hat{W}_i = V_1 \cup S_1^* \cup S_3^* \setminus W_i$. Then, $\sum_{j \in W_i} \zeta_j + d_{j,S_3^*} = \sum_{j \in W_i} d_{j,V_1} + d_{j,S_1^*} + d_{j,S_3^*} = 2E_{W_i} + E_{W_i, \hat{W}_i}$, where E_{W_i} denotes the number of edges inside W_i and E_{W_i, \hat{W}_i} denotes the number of edges between W_i and \hat{W}_i . Thus, (7.6.13) results in

$$\tilde{K} \int_1^\infty \int_{x_1}^\infty \cdots \int_{x_{t-1}}^\infty \prod_{i=1}^t x_i^{-\tau + \zeta_i + d_{i,\bar{U}} + (\tau-1)|W_i| - 2E_{W_i} - E_{W_i, \hat{W}_i}} dx_t \cdots dx_1. \quad (7.6.14)$$

We now show that

$$-\tau + \zeta_t + d_{t,\bar{U}} + (\tau-1)|W_t| - 2E_{W_t} - E_{W_t, \hat{W}_t} < -1, \quad (7.6.15)$$

so that the integral in (7.6.14) over x_t is finite. By definition of (7.6.11) and W_t (see also Figure 7.4) $d_{t,\bar{U}} = d_{t,W_t}$. Setting $\hat{S}_2 = \hat{S}_2^* \cup \{t\}$, $\hat{S}_1 = \hat{S}_1^* \cup W_t$ and $\hat{S}_3 = S_3^* \setminus (W_t \cup \{t\})$, gives

$$E_{\hat{S}_1} - E_{S_1^*} = E_{W_t} + E_{W_t, S_1^*} \quad (7.6.16)$$

$$E_{\hat{S}_1, \hat{S}_3} - E_{S_1^*, S_3^*} = E_{W_t, \hat{W}_t} - d_{t,W_t} = E_{W_t, S_3^*} - E_{W_t} - E_{W_t, S_1^*} - d_{t,\bar{U}} - d_{t,V_1} \quad (7.6.17)$$

$$E_{\hat{S}_1, V_1} - E_{S_1^*, V_1} = E_{W_t, V_1} \quad (7.6.18)$$

$$E_{\hat{S}_2, V_1} - E_{S_2^*, V_1} = d_{t,V_1}. \quad (7.6.19)$$

Because (7.2.2) is uniquely optimized for S_1^* , S_2^* and S_3^* ,

$$\begin{aligned}
 &|\hat{S}_1| - |\hat{S}_2| - \frac{2E_{\hat{S}_1} + E_{\hat{S}_1, \hat{S}_3} + E_{\hat{S}_1, V_1} - E_{\hat{S}_2, V_1}}{\tau-1} \\
 &< |S_1^*| - |S_2^*| - \frac{2E_{S_1^*} + E_{S_1^*, S_3^*} + E_{S_1^*, V_1} - E_{S_2^*, V_1}}{\tau-1}. \quad (7.6.20)
 \end{aligned}$$

Using (7.6.16)-(7.6.19) this reduces to

$$|W_t| - 1 - \frac{2E_{W_t} - E_{W_t, \hat{W}_t} - d_{t, \bar{U}} - d_{t, S_1^*} - d_{t, V_1}}{\tau - 1} < 0, \tag{7.6.21}$$

or

$$-\tau + (\tau - 1) |W_t| + d_{t, \bar{U}} + d_{t, S_1^*} + d_{t, V_1} - 2E_{W_t} - E_{W_t, \hat{W}_t} < -1. \tag{7.6.22}$$

Using that $\zeta_t = d_{t, V_1} + d_{t, S_1^*}$ shows that the inner integral of (7.6.14) is finite, and equal to

$$Kx_{t-1}^{1-2\tau+d_{t-1, \bar{U}}+(\tau-1)|W_t \cup W_{t-1}|-2E_{W_t \cup W_{t-1}} - E_{W_t \cup W_{t-1}, W_t \cup \hat{W}_{t-1}}}, \tag{7.6.23}$$

for some $K > 0$. Then, choosing $\hat{S}_2 = S_2^* \cup \{t, t-1\}$, $\hat{S}_1 = S_1^* \cup W_t \cup W_{t-1}$ and $\hat{S}_3 = \hat{S}_3^* \setminus (W_t \cup W_{t-1} \cup \{t, t-1\})$, we can again prove using (7.6.20) that integrating (7.6.23) over x_{t-1} from x_{t-2} to ∞ as in (7.6.14) results in a finite function of x_{t-2} . We can continue this process until we arrive at the integral over x_1 and show that this final integral is finite, so that (7.6.14) and therefore also (7.6.13) is finite. Since the ordering $x_1 < x_2 < \dots < x_t$ was arbitrary, the integral over any ordering of x_1, \dots, x_t is finite, so that (7.6.4) is finite as well. \square

Lemma 7.9. *Suppose the optimal solution to (7.2.2) is unique, and attained by S_1^* , S_2^* and S_3^* . Say that $S_2^* = \{1, \dots, t_2\}$ and $S_1^* = \{t_2 + 1, \dots, t_2 + t_1\}$. Then,*

$$\int_0^1 \dots \int_0^1 \int_0^\infty \dots \int_0^\infty \prod_{j \in [t_1+t_2]} x_j^{-\tau+\zeta_j} \prod_{(u,v) \in E_{S_1^*, S_2^*}} \min(x_u x_v, 1) dx_{t_1+t_2} \dots dx_1 < \infty. \tag{7.6.24}$$

Proof. This proof has a similar structure as the proof of Lemma 7.8. We first rewrite the integral as

$$\int_0^1 \dots \int_0^1 \prod_{j \in [t_2]} x_j^{-\tau+\zeta_j} \prod_{i=t_2+1}^{t_1+t_2} \tilde{h}(i, \mathbf{x}) dx_{t_2} \dots dx_1, \tag{7.6.25}$$

where $\mathbf{x} = (x_j)_{j \in [t_2]}$ and

$$\tilde{h}(i, \mathbf{x}) = \int_0^\infty x_i^{-\tau+\zeta_i} \prod_{j \in S_2^*: (i,j) \in E_H} \min(x_i x_j, 1) dx_i. \tag{7.6.26}$$

Similarly to (7.6.9) and (7.6.10), we assume that vertex j has vertices $1, 2, \dots, l$ as neighbors in S_2^* , but now we assume that $x_1 > x_2 > \dots > x_l$. Then, using that $\zeta_i + l = d_i \geq 2$ and that $\zeta_i \leq 1$ by Lemma 7.7, we obtain

$$\begin{aligned} \tilde{h}(i, \mathbf{x}) &= \int_0^\infty x_i^{-\tau+\zeta_i} \min(x_i x_1, 1) \min(x_i x_2, 1) \dots \min(x_i x_l, 1) dx_i \\ &= \int_0^{1/x_1} x_i^{-\tau+\zeta_i+l} x_1 \dots x_l dx_i + \int_{1/x_2}^{1/x_3} x_i^{-\tau+\zeta_i+l-1} x_2 \dots x_l dx_i \\ &\quad + \dots + \int_{1/x_l}^\infty x_i^{-\tau+\zeta_i} dx_i \\ &\leq K \int_0^{1/x_1} x_i^{-\tau+d_i} x_1 \dots x_l dx_i \end{aligned} \tag{7.6.27}$$

for some $K > 0$. For $j \in S_1^*$ define

$$w_j = \min\{i \mid i \in S_2^*, (i, j) \in E_H\}. \quad (7.6.28)$$

By Lemma 7.7, w_j is well defined for all $j \in S_1^*$. Then, using (7.6.27) results in

$$\begin{aligned} & \int_0^1 \int_0^{x_1} \cdots \int_0^{x_{t_2-1}} \prod_{j \in [t_2]} x_j^{-\tau + \zeta_j + d_{j, S_1^*}} \prod_{i=t_2+1}^{t_1+t_2} \tilde{h}(i, \mathbf{x}) dx_{t_2} \cdots dx_1 \\ & \leq \tilde{K} \int_0^1 \int_0^{x_1} \cdots \int_0^{x_{t_2-1}} \prod_{j \in [t_2]} x_j^{-\tau + \zeta_j + d_{j, S_1^*}} \prod_{i=t_2+1}^{t_1+t_2} (1/x_{w_j})^{-\tau+1+d_i} dx_{t_2} \cdots dx_1, \end{aligned} \quad (7.6.29)$$

for some $\tilde{K} > 0$. Define $W_i = \{j \in S_1^* \mid w_j = i\}$ for $i \in S_2^*$ and let $\bar{W}_i = V_H \setminus W_i$. Then, (7.6.29) reduces to

$$\tilde{K} \int_0^1 \int_0^{x_1} \cdots \int_0^{x_{t_2-1}} \prod_{i \in [t_2]} x_i^{-\tau + \zeta_i + d_{i, S_1^*} + (\tau-1)|W_i| - 2E_{W_i} - E_{W_i, W_i}} dx_{t_2} \cdots dx_1. \quad (7.6.30)$$

We set $\hat{S}_1 = S_1^* \setminus W_t$, $\hat{S}_2 = S_2^* \setminus \{t\}$ and $\hat{S}_3 = S_3^* \cup W_t \cup \{t\}$. Notice that $E_{S_1^*} - E_{\hat{S}_1} = E_{W_t} + E_{W_t, S_1^* \setminus W_t}$, $E_{S_1^*, S_3^*} - E_{\hat{S}_1, \hat{S}_3} = E_{W_t, S_3^*} - E_{t, S_1^* \setminus W_t} - E_{W_t, S_1^* \setminus W_t}$, $E_{S_1^*, V_1} - E_{\hat{S}_1, V_1} = E_{W_t, V_1}$ and $E_{S_2^*, V_1} - E_{\hat{S}_2, V_1} = E_{t, V_1}$. Because the optimal solution to (7.2.2) is unique, we obtain using (7.6.20) that

$$-\tau + (\tau-1)|W_t| - 2E_{W_t} + E_{W_t, S_1^* \setminus W_t} + E_{W_t, S_3^*} + E_{W_t, V_1} - d_{t, S_1^* \setminus W_t} - d_{t, V_1} > -1. \quad (7.6.31)$$

Using that $E_{W_t, \bar{W}_t} = E_{W_t, S_1^* \setminus W_t} + E_{W_t, S_3^*} + E_{W_t, S_2^*} + E_{W_t, V_1}$ and that $\zeta_t = d_{t, V_1}$ then shows that

$$-\tau + (\tau-1)|W_t| - 2E_{W_t} + E_{W_t, \bar{W}_t} - E_{W_t, S_2^*} - d_{t, S_1^* \setminus W_t} - \zeta_t > -1. \quad (7.6.32)$$

By definition of W_t , $E_{W_t, S_2^*} = d_{t, W_t}$. Then using that $d_{t, S_1^* \setminus W_t} + d_{t, W_t} = d_{t, S_1^*}$ results in

$$-\tau + (\tau-1)|W_t| - 2E_{W_t} + E_{W_t, \bar{W}_t} - d_{t, S_1^*} - \zeta_t > -1, \quad (7.6.33)$$

which shows that the inner integral in (7.6.30) is finite. A similar argument, setting $\hat{S}_1 = S_1^* \setminus (W_t \cup W_{t-1})$ and $\hat{S}_2 = S_2^* \setminus \{t, t-1\}$ shows that the second integral is also finite, and we can proceed to show that the outer integral of (7.6.30) is finite. Because the ordering $x_1 > x_2 > \cdots > x_t$ was arbitrary, the integral is finite over any reordering, so that (7.6.24) is finite. \square

Proof of Theorem 7.1(i). Because $D_{\max} = O_{\mathbb{P}}(n^{1/(\tau-1)})$, the contribution from vertices with $D_i > n^{1/(\tau-1)}/\varepsilon_n$ to the number of motifs converges to zero. Since we are only interested in the order of magnitude of the number of motifs, we assume for ease of notation that $D_{\max} = n^{1/(\tau-1)}$. We define

$$\gamma_i^u(n) = \begin{cases} n^{1/(\tau-1)} & \text{if } i \in S_2^* \\ n^{\alpha_i^*}/\varepsilon_n & \text{else,} \end{cases} \quad (7.6.34)$$

and

$$\gamma_i^l(n) = \begin{cases} 1 & \text{if } i \in V_1 \\ \varepsilon_n n^{\alpha_i^*} & \text{else.} \end{cases} \quad (7.6.35)$$

We then show that the expected number of motifs where the degree of at least one vertex i satisfies $D_i \notin [\gamma_i^l(n), \gamma_i^u(n)]$ is small, similarly to the proof of Theorem 7.2. We first study the expected number of copies of H where vertex 1 has degree in $[1, \gamma_1(n)]$ and all other vertices satisfy $D_i \in [\gamma_i^l(n), \gamma_i^u(n)]$. We bound the expected number of such copies of H by

$$\begin{aligned} & \mathbb{E} \left[N(H) \mathbb{1}_{\{D_1 < \gamma_1^l(n), D_i \in [\gamma_i^l(n), \gamma_i^u(n)] \forall i > 1\}} \right] \\ & \leq c^k \int_1^{\gamma_1^l(n)} \int_{\gamma_2^l(n)}^{\gamma_2^u(n)} \cdots \int_{\gamma_k^l(n)}^{\gamma_k^u(n)} (x_1 \cdots x_k)^{-\tau} \prod_{(i,j) \in E_H} \min\left(\frac{x_i x_j}{\mu n}, 1\right) dx_k \cdots dx_1. \end{aligned} \quad (7.6.36)$$

This integral equals zero when vertex 1 is in V_1 . Suppose vertex 1 is in S_2^* . W.l.o.g. assume that $S_2^* = \{1, \dots, t_2\}$, $S_1^* = \{t_2 + 1, \dots, t_1 + t_2\}$ and $S_3^* = \{t_1 + t_2 + 1, \dots, t_1 + t_2 + t_3\}$. We bound the minimum in (7.6.36) by $x_i x_j / (\mu n)$ for $i, j \in S_1^*$, for i or j in V_1 and for $i \in S_1^*, j \in S_3^*$ or vice versa. We bound the minima by 1 for $i, j \in S_2^*$ and $i \in S_2^*, j \in S_3^*$ or vice versa. Applying the change of variables $u_i = x_i / n^{\alpha_i^*}$ results similarly as in (7.3.22) and (7.3.34) in the bound

$$\begin{aligned} & \mathbb{E} \left[N(H) \mathbb{1}_{\{D_1 < \gamma_1^l(n), D_i \in [\gamma_i^l(n), \gamma_i^u(n)] \forall i > 1\}} \right] = O\left(n^{\frac{3-\tau}{2}(k_{2+} + B^{\text{(sub)}}(H)) + k_1}\right) \\ & \quad \times \int_0^{\varepsilon_n} \int_0^1 \cdots \int_0^1 \int_0^\infty \cdots \int_0^\infty \prod_{i \in V_H \setminus V_1} x_i^{-\tau + \zeta_i} \\ & \quad \times \prod_{(u,v) \in E_{S_3^*} \cup E_{S_1^*, S_2^*}} \min(x_u x_v, 1) dx_{t_1+t_2+t_3} \cdots dx_1 \prod_{j \in V_1} \int_1^\infty x_j^{1-\tau} dx_j, \end{aligned} \quad (7.6.37)$$

where the integrals from 0 to 1 correspond to vertices in S_2^* and the integrals from 0 to ∞ to vertices in S_1^* and S_3^* . The last integrals corresponding to vertices in V_1 are finite, because $\tau \in (2, 3)$. The first integral can be split into

$$\begin{aligned} & \int_0^{\varepsilon_n} \int_0^1 \cdots \int_0^1 \int_0^\infty \cdots \int_0^\infty \prod_{i \in S_1^* \cup S_2^*} x_i^{-\tau + \zeta_i} \prod_{(u,v) \in E_{S_1^*, S_2^*}} \min(x_u x_v, 1) dx_{t_1+t_2} \cdots dx_1 \\ & \quad \times \int_0^\infty \cdots \int_0^\infty \prod_{i \in S_3^*} x_i^{-\tau + \zeta_i} \prod_{(u,v) \in E_{S_3^*}} \min(x_u x_v, 1) dx_{t_1+t_2+t_3} \cdots dx_{t_1+t_2+1} \end{aligned} \quad (7.6.38)$$

By Lemma 7.8 the second integral of (7.6.37) is finite. Using Lemma 7.9 shows that the first integral is $o(1)$. Therefore,

$$\mathbb{E} \left[N(H) \mathbb{1}_{\{D_1 < \gamma_1^l(n), D_i \in [\gamma_i^l(n), \gamma_i^u(n)] \forall i > 1\}} \right] = o\left(n^{\frac{3-\tau}{2}(k_{2+} + B^{\text{(sub)}}(H)) + k_1}\right), \quad (7.6.39)$$

when vertex 1 is in S_2^* . Similarly, we can show that the expected contribution from $D_1 < \gamma_1^l(n)$ satisfies the same bound when vertex 1 is in S_1^* or S_3^* . The expected number of

motifs where $D_1 > \gamma_1^u(n)$ if vertex 1 is in S_1^* , S_3^* or V_1 can be bounded similarly, as well as the expected contribution where multiple vertices have $D_i \notin [\gamma_i^l(n), \gamma_i^u(n)]$. Therefore, by the Markov inequality

$$N\left(H, \bar{M}_n^{(\alpha^{(\text{sub})})}(\varepsilon_n)\right) = o_{\mathbb{P}}\left(n^{\frac{3-\tau}{2}(k_{2+} + B^{(\text{sub})}(H)) + k_1}\right), \quad (7.6.40)$$

where $N(H, \bar{M}_n^{(\alpha^{(\text{sub})})}(\varepsilon_n))$ denotes the number of copies of H on vertices not in $M_n^{(\alpha^{(\text{sub})})}(\varepsilon_n)$. Combining this with the fact that for fixed ε

$$N(H) \geq N(H, M_n^{(\alpha^{(\text{sub})})}(\varepsilon)) = O_{\mathbb{P}}\left(n^{\frac{3-\tau}{2}(k_{2+} + B^{(\text{sub})}(H)) + k_1}\right) \quad (7.6.41)$$

by Theorem 7.1(ii) shows that

$$\frac{N(H, M_n^{(\alpha^{(\text{sub})})}(\varepsilon_n))}{N(H)} \xrightarrow{\mathbb{P}} 1. \quad (7.6.42)$$

□

7.7 Graphlets

We now describe how to adapt the analysis of motif counts to graphlet counts. For graphlets we can define a similar optimization problem as (7.3.24). When $\alpha_i + \alpha_j \leq 1$, (7.3.5) results in

$$\mathbb{P}_n(X_{ij} = 0) = e^{-\Theta(n^{\alpha_i + \alpha_j - 1})}(1 + o(1)) = 1 + o(1), \quad (7.7.1)$$

whereas for $\alpha_i + \alpha_j > 1$, (7.3.15) yields

$$\mathbb{P}_n(X_{ij} = 0) = o(1). \quad (7.7.2)$$

Similar to (7.3.17), we can write the probability that H occurs as an induced subgraph on $\mathbf{v} = (v_1, \dots, v_k)$ as

$$\mathbb{P}_n(G|_{\mathbf{v}} = E_H) = \Theta_{\mathbb{P}}\left(\prod_{(v_i, v_j) \in E_H: \alpha_i + \alpha_j < 1} n^{\alpha_i + \alpha_j - 1} \prod_{(i, j) \notin E_H: \alpha_i + \alpha_j > 1} o(e^{-n^{\alpha_i + \alpha_j - 1}/2})\right). \quad (7.7.3)$$

Thus, the probability that H is an induced subgraph on (v_1, \dots, v_k) is stretched exponentially small in n when two vertices i and j with $\alpha_i + \alpha_j > 1$ are not connected in H . Then the corresponding optimization problem to (7.3.23) for graphlets becomes

$$\begin{aligned} & \max(1 - \tau) \sum_i \alpha_i + \sum_{(i, j) \in E_H: \alpha_i + \alpha_j < 1} \alpha_i + \alpha_j - 1, \\ & \text{s.t. } \alpha_i + \alpha_j \leq 1 \quad \forall (i, j) \notin E_H. \end{aligned} \quad (7.7.4)$$

The following lemma shows that this optimization problem attains its optimum for very specific values of α (similarly to Lemma 7.2 for motifs):

Lemma 7.10 (Maximum contribution to graphlets). *Let H be a connected graph on k vertices. If the solution to (7.7.5) is unique, then the optimal solution satisfies $\alpha_i \in \{0, \frac{\tau-2}{\tau-1}, \frac{1}{2}, \frac{1}{\tau-1}\}$ for all i . If it is not unique, then there exist at least 2 optimal solutions with $\alpha_i \in \{0, \frac{\tau-2}{\tau-1}, \frac{1}{2}, \frac{1}{\tau-1}\}$ for all i . In any optimal solution, $\alpha_i = 0$ if and only if vertex i has degree one in H .*

Proof. This proof is similar to the proof of Lemma 7.2. First, we again define $\beta_i = \alpha_i - \frac{1}{2}$, so that (7.7.4) becomes

$$\begin{aligned} & \max \frac{1-\tau}{2}k + (1-\tau) \sum_i \beta_i + \sum_{(i,j) \in E_H: \beta_i + \beta_j < 0} \beta_i + \beta_j, \\ & \text{s.t. } \beta_i + \beta_j \leq 0 \quad \forall (i,j) \notin E_H. \end{aligned} \quad (7.7.5)$$

The proof of Step 1 from Lemma 7.2 then also holds for graphlets. Now we prove that if the optimal solution to (7.7.5) is unique, it satisfies $\beta_i \in \{-\frac{1}{2}, \frac{\tau-3}{2(\tau-1)}, 0, \frac{3-\tau}{2(\tau-1)}\}$ for all i . We take $\tilde{\beta}$ as in (7.3.28), and assume that $\tilde{\beta} < \frac{3-\tau}{2(\tau-1)}$. The contribution of the vertices with $|\beta_i| = \tilde{\beta}$ is as in (7.3.29). By increasing $\tilde{\beta}$ or by decreasing it to zero, the constraints on $\beta_i + \beta_j$ are still satisfied for all (i,j) . Thus, we can use the same argument as in Lemma 7.2 to conclude that $\beta_i \in \{\frac{\tau-3}{2(\tau-1)}, 0, \frac{3-\tau}{2(\tau-1)}\}$ for all i with $d_i \geq 2$. A similar argument as in Step 3 of Lemma 7.2 shows that if the solution to (7.7.5) is not unique, it can be transformed into two optimal solutions that satisfy $\beta_i \in \{-\frac{1}{2}, \frac{\tau-3}{2(\tau-1)}, 0, \frac{3-\tau}{2(\tau-1)}\}$ for all i with degree at least 2. \square

Following the same lines as the proof of Theorem 7.1(ii) for motifs, Theorem 7.1(ii) for graphlets follows, where we now use Lemma 7.10 instead of 7.2. We now state an equivalent lemma to Lemma 7.3 for graphlets:

Lemma 7.11 (Convergence of major contribution to graphlets). *Let H be a connected graph on $k > 2$ vertices such that (7.2.3) is uniquely optimized at 0. Then,*

(i) *the number of graphlets with vertices in $W_n^k(\varepsilon)$ satisfies*

$$\begin{aligned} & \frac{N^{(\text{ind})}(H, W_n^k(\varepsilon))}{n^{\frac{k}{2}(3-\tau)}} \xrightarrow{\mathbb{P}} c^k \mu^{-\frac{k}{2}(\tau-1)} \int_{\varepsilon}^{1/\varepsilon} \dots \int_{\varepsilon}^{1/\varepsilon} (x_1 \dots x_k)^{-\tau} \\ & \quad \times \prod_{(i,j) \in E_H} (1 - e^{-x_i x_j}) \prod_{(i,j) \notin E_H} e^{-x_i x_j} dx_1 \dots dx_k. \end{aligned} \quad (7.7.6)$$

(ii) *$A^{(\text{ind})}(H)$ defined in (7.2.11) satisfies $A^{(\text{ind})}(H) < \infty$.*

The proof of Theorem 7.3 is similar to the proof of Theorem 7.2, and uses Lemma 7.11 instead of Lemma 7.3 for motifs. The proof of Lemma 7.11(i) in its turn follows from straightforward extensions of Lemmas 7.4 7.6 and 7.5 for graphlets, now also using that the probability that an edge $\{i, j\}$ not present in H is not present in the subgraph can be approximated by $\exp(-D_i D_j / L_n)$. Lemma 7.11(ii) is an application of the following equivalent lemma to Lemma 7.8 for $S_3^* = V_H$:

Lemma 7.12. *Suppose that the maximum in (7.2.3) is uniquely attained for $|S_3^*| = s > 0$, and say $S_3^* = \{1, \dots, s\}$. Then*

$$\int_0^\infty \cdots \int_0^\infty \prod_{i \in [s]} x_i^{-\tau + \zeta_i} \prod_{(u,v) \in E_{S_3^*}} \min(x_u x_v, 1) \prod_{(u,v) \notin E_{S_3^*}} e^{-x_u x_v} dx_s \cdots dx_1 < \infty. \quad (7.7.7)$$

Proof. This integral is finite if

$$\int_0^\infty \cdots \int_0^\infty \prod_{i \in [s]} x_i^{-\tau + \zeta_i} \prod_{(u,v) \in E_{S_3^*}} \min(x_u x_v, 1) \prod_{(u,v) \notin E_{S_3^*}} \mathbb{1}_{\{x_u x_v < 1\}} dx_s \cdots dx_1 < \infty, \quad (7.7.8)$$

since if

$$\int_a^b \int_0^{1/x_1} x_1^{\gamma_1} x_2^{\gamma_2} e^{-x_1 x_2} dx_2 dx_1 < \infty, \quad (7.7.9)$$

then also

$$\int_a^b \int_{1/x_1}^\infty x_1^{\gamma_1} x_2^{\gamma_2} e^{-x_1 x_2} dx_2 dx_1 < \infty. \quad (7.7.10)$$

We can show similarly to (7.6.5) and (7.6.6) that the integral is finite when all integrands are larger than one, or when all are smaller than one. We compute the contribution to (7.7.8) where the integrand runs from 1 to ∞ for vertices in some nonempty set U , and from 0 to 1 for vertices in $\bar{U} = S_3^* \setminus U$. W.l.o.g., assume $U = \{1, \dots, t\}$ for some $t \geq 1$. Define for $i \in \bar{U}$

$$\hat{h}(i, \mathbf{x}) = \int_0^1 x_i^{-\tau + \zeta_i + d_{i, \bar{U}}} \prod_{j \in U: (i,j) \in E_H} \min(x_i x_j, 1) \prod_{v \in U: (i,v) \notin E_H} \mathbb{1}_{\{x_i x_v < 1\}} dx_i. \quad (7.7.11)$$

Then (7.7.7) results in

$$\int_1^\infty \cdots \int_1^\infty \prod_{j \in [t]} x_j^{-\tau + \zeta_j} \prod_{u,v \in U: (u,v) \notin E_H} \mathbb{1}_{\{x_u x_v < 1\}} \prod_{i=t+1}^k \hat{h}(i, \mathbf{x}) dx_t \cdots dx_1. \quad (7.7.12)$$

When U does not induce a complete graph on H , this integral equals zero. Thus, we assume that U induces a complete graph on H so that $\{(u,v) \in U \mid (u,v) \notin E_H\} = \emptyset$. Assume that $1 < x_2 < \dots < x_t$. Then, similarly to (7.6.10),

$$\hat{h}(i, \mathbf{x}) \leq K \int_0^{1/x_t} x_i^{-\tau + \zeta_i + d_{i, S_3^*}} \prod_{j \in [t]: (i,j) \in E_H} x_j dx_i, \quad (7.7.13)$$

for some $K > 0$. By a similar argument as in Lemma 7.7, $\zeta_i + d_{i, S_3^*} \geq 2$ for $i \in S_3^*$ so that this integral is finite. Thus,

$$\begin{aligned} & \int_1^\infty \cdots \int_{x_{t-1}}^\infty \prod_{j \in [t]} x_j^{-\tau + \zeta_j} \prod_{u,v \in U: (u,v) \notin E_H} \mathbb{1}_{\{x_u x_v < 1\}} \prod_{i=t+1}^k \hat{h}(i, \mathbf{x}) dx_t \cdots dx_1 \\ & \leq \tilde{K} \int_1^\infty \int_{x_1}^\infty \cdots \int_{x_{t-1}}^\infty \prod_{j=1}^{t-1} x_j^{-\tau + \zeta_j + d_{j, \bar{U}}} x_t^{-\tau + \zeta_t + d_{t, \bar{U}} + (\tau-1)|\bar{U}| - \sum_{s \in \bar{U}} \zeta_s + d_{s, S_3^*}} dx_t \cdots dx_1 \\ & \leq \tilde{K} \int_1^\infty \int_{x_1}^\infty \cdots \int_{x_{t-1}}^\infty \prod_{j=1}^{t-1} x_j^{-\tau + \zeta_j + d_{j, \bar{U}}} x_t^{-\tau + \zeta_t + d_{t, \bar{U}} + (\tau-1)|\bar{U}| - 2E_{\bar{U}} - E_{\bar{U}, \bar{U}}} dx_t \cdots dx_1, \end{aligned} \quad (7.7.14)$$

for some $\tilde{K} > 0$, where $\hat{U} = V_1 \cup S_1^* \cup S_3^* \setminus \bar{U}$. We can now show that the integral over x_t is finite in a similar manner as in Lemma 7.8. Define $\hat{S}_1 = S_1^* \cup \bar{U}$, $\hat{S}_2 = S_2^* \cup \{t\}$ and $\hat{S}_3 = U \setminus \{t\}$. Because U induces a complete graph in H , and $t \in U$, vertex t is connected to all vertices in \hat{S}_3 , so that these newly defined sets still satisfy the constraint in (7.2.3), and we may proceed as in Lemma 7.8 using (7.6.20) to show that the integral over x_t finite. Iterating this proves Lemma 7.12. □

The following lemma is the counterpart of Lemma 7.9 for graphlets:

Lemma 7.13. *Suppose that the optimal solution to (7.2.3) is unique, and attained by S_1^* , S_2^* and S_3^* . Say that $S_2^* = \{1, \dots, t_2\}$ and $S_1^* = \{t_2 + 1, \dots, t_2 + t_1\}$. Then,*

$$\int_0^1 \cdots \int_0^1 \int_0^\infty \cdots \int_{0, j \in [t_1+t_2]}^\infty \prod x_j^{-\tau+\zeta_j} \prod_{(u,v) \in E_{S_1^*, S_2^*}} \min(x_u x_v, 1) \times \prod_{(u,v) \notin E_{S_1^*, S_2^*}} e^{-x_u x_v} dx_{t_1+t_2} \cdots dx_1 < \infty. \tag{7.7.15}$$

Proof. This lemma can be proven along similar lines as Lemma 7.9, with similar adjustments as the adjustments to prove Lemma 7.12 for graphlets from its counterpart for motifs, Lemma 7.8. □

From these lemmas, the proof of Theorem 7.1(i) for graphlets follows along the same lines as the proof of Theorem 7.1(i) for motifs.

8 Subgraph fluctuations in inhomogeneous random graphs

Based on:

Variational principle for scale-free network motifs
C. Stegehuis, R. van der Hofstad, J.S.H. van Leeuwen

In this chapter, we again focus on subgraph counts. In Chapter 7 we have investigated the number of subgraphs in a typical network observation. We now study the *expected* number of subgraphs instead of the typical number of subgraphs by presenting a variational principle that identifies the dominant structure of any given subgraph as the solution of an optimization problem. The unique optimizer describes the vertex degrees that contribute most to the expected motif count, resulting in explicit asymptotic formulas for the expected motif count and its fluctuations. Interestingly, the structures that optimize the contribution to the expected motif count may be different than the most likely motif structure of Chapter 7. We then classify all motifs into two categories: self-averaging motifs with relatively small fluctuations signaling stable network structures, and motifs that are inherently non-self-averaging. We further divide the non-self-averaging motifs into two classes with substantially different concentration properties. We compare these theoretical findings with empirical network data.

8.1 Introduction

In this chapter, we again study the presence of network motifs. With the goal to explain the occurrence of motifs beyond counting, we develop a variational principle designed to identify, for any given motif, the composition that dominates the expected motif count as the solution of an optimization problem. The unique optimizer describes the degrees of the vertices that form the largest contribution to the expected motif count, as well as predicts the leading asymptotic order for the motif count in the large-network limit. The variational principle extends the analysis of constrained triangle counts of Chapter 6 to general subgraphs.

The variational principle can potentially be applied to various random network models, as we show in Chapter 9 but here we focus on the rank-1 inhomogeneous random graph described in Section 1.1.4. We focus on the Chung-Lu version of the rank-1 inhomogeneous random graph, with scale-free vertex weights. Thus, each vertex has a weight h drawn from the probability density

$$\rho(h) = Ch^{-\tau} \tag{8.1.1}$$

for some constant C and $h \geq h_{\min}$. Each pair of vertices is joined independently with probability

$$p(h, h') = \min(hh' / (\mu n), 1). \quad (8.1.2)$$

with h and h' the weights associated with the two vertices, and μ the mean of the weights. The variational principle now seeks for the dominant structure of any given motif. The probability $P(H)$ of creating motif H on k uniformly chosen vertices can be written as

$$P(H) = \int_{\mathbf{h}} \mathbb{P}(H \text{ on } h_1, \dots, h_k) \mathbb{P}(h_1, \dots, h_k) d\mathbf{h}, \quad (8.1.3)$$

where the integral is over all possible hidden-variable sequences on k vertices, with $\mathbf{h} = (h_1, \dots, h_k)$ and $\mathbb{P}(h_1, \dots, h_k)$ the density that a randomly chosen set of k weights is proportional to h_1, \dots, h_k . Degrees and weights are provably close (see Section 4.C), so (8.1.3) can be interpreted as a sum over all possible degree sequences. Therefore, the variational principle then needs to settle the following trade-off, inherently present in power-law networks, also described in Chapter 7: On the one hand, large-degree vertices like hubs contribute substantially to the number of motifs, because they are highly connected, and therefore participate in many motifs. On the other hand, large-degree vertices are by definition rare. This should be contrasted with lower-degree vertices that occur more frequently, but take part in fewer connections and hence fewer motifs. Therefore, our variational principle is designed to find the selection of vertices with specific degrees that together ‘optimize’ this trade-off and hence maximize the expected number of such motifs.

We leverage the variational principle in three ways. First, we derive sharp asymptotic expressions for the expected motif counts in the large-network limit in terms of the network size and the power-law exponent. Second, we use the variational principle to identify the fluctuations of motif counts. In this way we can determine for any given motif whether it is self-averaging or not. Third, we compare the optimal motifs composition with real-world network data, and find empirical confirmation of typical motif degrees.

We present two versions of the variational principle arising from analyzing (8.1.3) in Sections 8.2 and 8.3, related to different ways of restricting the variation (free and typical). Free variation corresponds to computing the average number of motifs over many samples of the random network model. Typical variation corresponds to the number of motifs in one single instance of the random graph null model, and yields the same optimization problems as in Chapter 7. Remarkably, for $\tau \in (2, 3)$ these can be highly different. We then apply the variational principle to study motif count fluctuations in Section 8.4 and classify all motifs into self-averaging and non-self-averaging motifs. We further show that non-self-averaging motifs can be further divided into two classes, with considerably different concentration properties. Finally, in Section 8.5 we compare the results of the variational principle to real-world data.

8.2 Free variation

We first show that only hidden-variable sequences \mathbf{h} with weights of specific orders give the largest contribution to (8.1.3), using a similar approach as in Chapter 7. Write

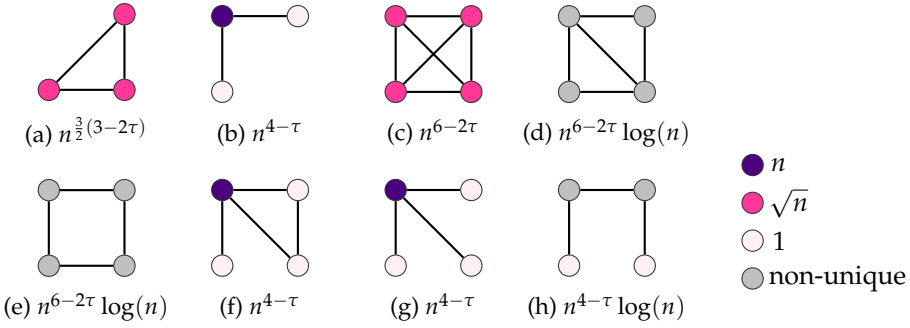


Figure 8.1: Order of magnitude of the expected number of motifs on 3 and 4 vertices, where the vertex color indicates the optimal vertex degree. Vertices where the optimizer is not unique are gray.

the weights as $h_i \propto n^{\alpha_i}$ for some $\alpha_i \geq 0$ for all i . Then, using (8.1.2), the probability that motif H exists on vertices with weights $\mathbf{h} = (n^{\alpha_1}, \dots, n^{\alpha_k})$ satisfies

$$\mathbb{P}(H \text{ on } \mathbf{h}) \propto \prod_{(i,j) \in E_H: \alpha_i + \alpha_j < 1} n^{\alpha_i + \alpha_j - 1}. \tag{8.2.1}$$

The weights are an i.i.d. sample from a power-law distribution, so that the probability that k uniformly chosen weights satisfy $(h_1, \dots, h_k) \propto (n^{\alpha_1}, \dots, n^{\alpha_k})$ is of the order $n^{(1-\tau)\sum_i \alpha_i}$ (see (7.3.21)). Taking the product of this with (8.2.1) shows that the maximum contribution to the summand in (8.1.3) is obtained for those $\alpha_i \geq 0$ that maximize the exponent

$$(1 - \tau) \sum_i \alpha_i + \sum_{(i,j) \in E_H: \alpha_i + \alpha_j < 1} (\alpha_i + \alpha_j - 1), \tag{8.2.2}$$

which is a piecewise-linear function in α . In Appendix 8.B, we show that the maximizer of this optimization problem satisfies $\alpha_i \in \{0, \frac{1}{2}, 1\}$ for all i . Thus, the maximal value of (8.2.2) is attained by partitioning the vertices of H into the sets S_1, S_2, S_3 such that vertices in S_1 have $\alpha_i = 0$, vertices in S_2 have $\alpha_i = 1$ and vertices in S_3 have $\alpha_i = \frac{1}{2}$. Then, the edges with $\alpha_i + \alpha_j < 1$ are edges inside S_1 and edges between S_1 and S_3 . If we denote the number of edges inside S_1 by E_{S_1} and the number of edges between S_1 and S_3 by E_{S_1, S_3} , then maximizing (8.2.2) is equivalent to maximizing

$$B_f(H) = \max_{\mathcal{P}} \left[|S_1| - |S_2| - \frac{2E_{S_1} + E_{S_1, S_3}}{\tau - 1} \right] \tag{8.2.3}$$

over all partitions \mathcal{P} of the vertices of H into S_1, S_2, S_3 . This gives the following theorem (a more elaborate version is proven in Appendix 8.B):

Theorem 8.1 (Expected motif count). *Let H be a motif on k vertices such that the solution to (8.2.3) is unique. As $n \rightarrow \infty$, the expected number of motifs H grows as*

$$\mathbb{E}[N(H)] = n^k P(H) \propto n^{\frac{3-\tau}{2}k + \frac{\tau-1}{2}B_f(H)}, \tag{8.2.4}$$

and is thus fully determined by the partition \mathcal{P}^* that optimizes (8.2.3).

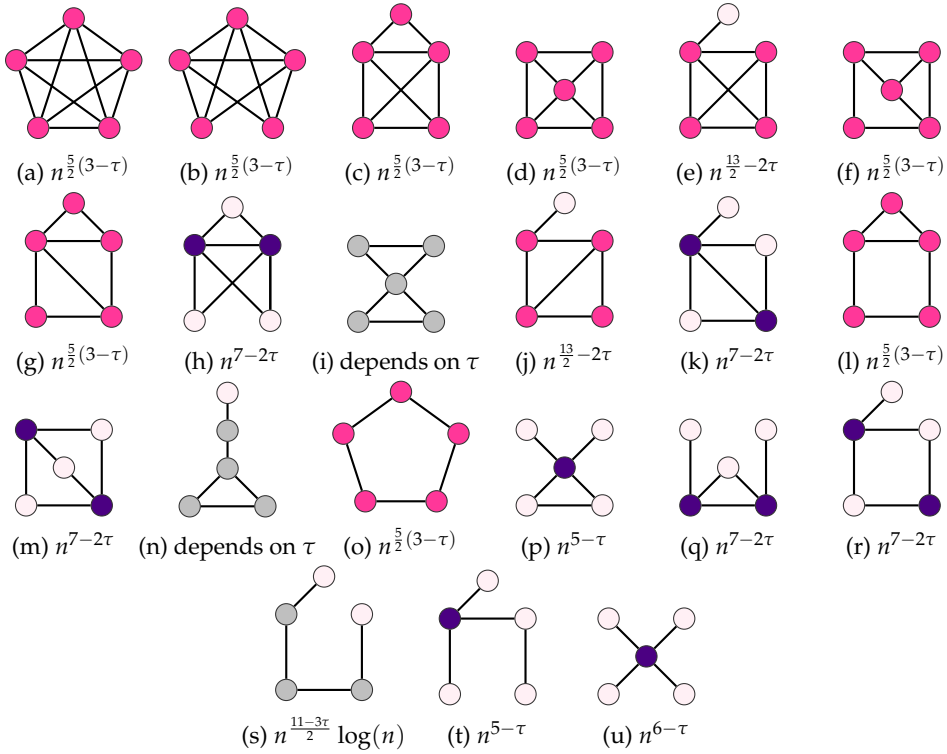


Figure 8.2: Order of magnitude of the expected number of motifs on 5 vertices, where the vertex color indicates the optimal vertex degree, as in Figure 8.1.

Theorem 8.1 implies that the expected number of motifs is dominated by motifs on vertices with weights (and thus degrees) of specific orders of magnitude: constant degrees, degrees proportional to \sqrt{n} or degrees proportional to n . Figure 8.1 and 8.2 show the partitions \mathcal{P}^* that form the optimal motif structures and dominate the expected number of motifs on three, four and five vertices.

Graphlets. It is also possible to only count the number of times H appears as an induced subgraph, also called graphlet counting. This means that an edge that is not present in graphlet H , should also be absent in the network subgraph. In Appendix 8.D we classify the expected number of graphlets with a similar variational principle as for motifs. Figure 8.7 shows the expected number of graphlets on 4 vertices. This figure also shows that graphlet counting is more detailed than motif counting. For example, counting all square motifs is equivalent to counting all graphlets that contain the square as an induced subgraph: the square, the diamond and K_4 . Indeed, the expected number of square motifs scales as $n^{6-2\tau} \log(n)$ by adding the number of square, diamond and K_4 graphlets from Figure 8.7. This shows that the main contribution to the expected number of square motifs is actually from the diamond graphlets of Figure 8.7d. Thus, graphlet counting gives more detailed

information than motif counting.

8.3 Typical variation

The largest degrees (hubs) in typical samples of the rank-1 inhomogeneous random graph scale as $n^{1/(\tau-1)}$ with high probability. The expected number of motifs, however, may be dominated by network samples where the largest degree is proportional to n (see Theorem 8.1). These samples contain many motifs because of the high degrees, and therefore contribute significantly to the expectation. Nevertheless, the probability of observing such a network tends to zero as n grows large. In Chapter 7 we investigated the typical number of motifs in erased configuration models, using a similar variational principle. There we have also assumed the degrees to be proportional to n^{a_i} , but limit to degree sequences where the maximal degree is of order $n^{1/(\tau-1)}$, the natural cutoff in view of the typical hub degrees. Vertices with degrees above the natural cutoff are rare, so do not contribute to the typical motif count, but our variational principle shows that they may still contribute to the average motif count. In the proof of Theorem 7.1, the variational principle for the erased configuration model, the connection probability is estimated by $\min(D_i D_j / (\mu n), 1)$, which is the same as the connection probability in the rank-1 inhomogeneous random graph. Therefore, the optimal typical motif structures in the rank-1 inhomogeneous random graph are the same as in Figure 7.2 and the optimal typical graphlet structures are given by Figures 7.1 and 7.3.

Observe that the optimal motif structures for the expected number of motifs of Figure 8.1 and for the typical number of motifs of Figure 7.2 may differ. For example, the scaling of the expected number of claws (Figure 8.1g) and the typical number of claws (Figure 7.1e) is different. This is caused by the left upper vertex that has degree proportional to n in the free optimal structure, whereas its typical degree is proportional to $n^{1/(\tau-1)}$. Only when the solution to (8.2.3) does not involve hub vertices, the two scalings coincide. When the optimal structure contains hub vertices, this is not the case. While typical hub degrees scale as $n^{1/(\tau-1)}$, expected hub degrees may be much larger, causing the number of such motifs with hubs to scale faster in the free variation setting than in the typical variation setting. This indicates that the average and median motif count can differ dramatically.

8.4 Fluctuations

Self-averaging network properties have relative fluctuations that tend to zero as the network size n tends to infinity. Several physical quantities in for example Ising models, fluid models and properties of the galaxy display non-self-averaging behavior [2, 69, 134, 171, 179, 219, 220]. We consider motif counts $N(H)$ and call $N(H)$ self-averaging when $\text{Var}(N(H)) / \mathbb{E}[N(H)]^2 \rightarrow 0$ as $n \rightarrow \infty$. Essential understanding of $N(H)$ can then be obtained by taking a large network sample, since the sample-to-sample fluctuations vanish in the thermodynamic limit. In contrast, if $\text{Var}(N(H)) / \mathbb{E}[N(H)]^2$ approaches a constant or tends to infinity as $n \rightarrow \infty$, the

motif count is called non-self-averaging, in which case $N(H)$ shows strong sample-to-sample fluctuations that cannot be mitigated by taking more network samples.

Our variational principle facilitates a systematic study of such fluctuations, and leads to a classification into self-averaging and non-self-averaging motifs. It turns out that whether $N(H)$ is self-averaging or not depends on the power-law exponent τ and the optimal structure of H . We also show that non-self-averaging behavior of motif counts may not have the intuitive explanation described above. In some cases, motif counts in two instances are similar with high probability, but rare network samples behave differently, causing the variance of the motif count to be large, leading to non-self-averaging behavior. Thus, the classification of network motifs into self-averaging and non-self-averaging motifs does not give a complete picture of the motif count fluctuations. We therefore further divide the non-self-averaging motifs into two classes based on the type of fluctuations in the motif counts.

8.4.1 Triangle fluctuations

We first illustrate how we can apply the variational principle to obtain the variance of the number of subgraphs by computing the variance of the number of triangles in the rank-1 inhomogeneous random graph. Let Δ denote the number of triangles, and let $\Delta_{i,j,k}$ denote the event that vertices i, j and k form a triangle. Then, we can write the number of triangles as

$$\Delta = \frac{1}{6} \sum'_{i,j,k \in [n]} \mathbb{1}_{\Delta_{i,j,k}}, \quad (8.4.1)$$

where \sum' denotes the sum over distinct indices. Thus, the variance of the number of triangles can be written as

$$\text{Var}(\Delta) = \sum'_{i,j,k \in [n]} \sum'_{s,t,u \in [n]} \mathbb{P}(\Delta_{i,j,k}, \Delta_{s,t,u}) - \mathbb{P}(\Delta_{i,j,k})\mathbb{P}(\Delta_{s,t,u}). \quad (8.4.2)$$

When i, j, k and s, t, u do not overlap, the weights of i, j, k and s, t, u are independent, so that the event that i, j and k form a triangle and the event that s, t and u form a triangle are independent. Thus, when i, j, k, s, t, u are all distinct, $\mathbb{P}(\Delta_{i,j,k}, \Delta_{s,t,u}) = \mathbb{P}(\Delta_{i,j,k})\mathbb{P}(\Delta_{s,t,u})$, so that the contribution from 6 distinct indices to (8.4.2) is zero. On the other hand, when $i = u$ for example, the first term in (8.4.2) denotes the probability that a bow tie (see Figure 8.2(i)) is present with i as middle vertex. Furthermore, since the degrees are i.i.d. and the edge statuses are independent as well, $\mathbb{P}(\Delta_{i,j,k})$ is the same for any $i \neq j \neq k$, so that

$$\mathbb{P}(\Delta_{i,j,k}) = \frac{\mathbb{E}[\Delta]}{6\binom{n}{3}} = \frac{\mathbb{E}[\Delta]}{6n^3}(1 + o(1)). \quad (8.4.3)$$

This results in

$$\begin{aligned} \text{Var}(\Delta) &= 9\mathbb{E}[\# \text{ bow-ties}] - 9n^{-1}\mathbb{E}[\Delta]^2 + 18\mathbb{E}[\# \text{ diamonds}] \\ &\quad - 18n^{-2}\mathbb{E}[\Delta]^2 + 6\mathbb{E}[\Delta] - 6n^{-3}\mathbb{E}[\Delta]^2 \\ &= 9\mathbb{E}[\# \text{ bow-ties}] + 18\mathbb{E}[\# \text{ diamonds}] + 6\mathbb{E}[\Delta] + \mathbb{E}[\Delta]^2 O(n^{-1}), \end{aligned} \quad (8.4.4)$$

where the diamond motif is as in Figure 8.1d. The combinatorial factors 9,18 and 6 arise because there are 9 ways to construct a bow tie (18 for a diamond, and 6 for a triangle) by letting two triangles overlap. The diamond motif does not satisfy the assumption in Theorem 8.1 that the optimal solution to (8.2.3) is unique. However, we can show the following result:

Lemma 8.1. $\mathbb{E}[\text{number of diamonds}] = \Theta(n^{6-2\tau}) \log(n)$.

Proof. The function (8.2.3) is optimized for $\alpha_i = \beta$, $\alpha_j = 1 - \beta$, $\alpha_k = \beta$ and $\alpha_s = 1 - \beta$ for all values of $\beta \in [1/2, 1]$. All these optimizers together give the major contribution to the number of diamonds. Thus, we need to find the number of sets of four vertices, satisfying

$$h_i h_j = \Theta(n), \quad h_i > h_j, \quad h_k = \Theta(h_i), \quad h_s = \Theta(h_j). \quad (8.4.5)$$

Given h_i and h_j , the number of sets of two vertices k, s with $h_k = \Theta(h_i)$ and $h_s = \Theta(h_j)$ is given by $n^2 h_i^{1-\tau} h_j^{1-\tau} = \Theta(n^{3-\tau})$, where we used that $h_i h_j = \Theta(n)$. The number of sets of vertices i, j such that $h_i h_j = \Theta(n)$ can be found using that the product of two independent power-law random variables is again distributed as a power law, with an additional logarithmic term [110]. Thus, the number of sets of vertices with $h_i h_j = \Theta(n)$ scales as $n^2 n^{1-\tau} \log(n)$. Then, the expected number of sets of four vertices satisfying all constraints on the degrees scales as $n^{6-2\tau} \log(n)$. By (8.2.1), the probability that a diamond exists on degrees satisfying (8.4.5) is asymptotically constant, so that the expected number of diamonds also scales as $n^{6-2\tau} \log(n)$. \square

Theorem 8.3 gives for the number of bow ties that

$$\mathbb{E}[\#\text{ bow ties}] = \begin{cases} \Theta(n^{\frac{5}{2}(3-\tau)}) & \tau < \frac{7}{3}, \\ \Theta(n^{4-\tau}) & \tau \geq \frac{7}{3}, \end{cases} \quad (8.4.6)$$

and that $\mathbb{E}[\Delta] = \Theta(n^{3(3-\tau)/2})$. Combining this with (8.4.4) results in

$$\text{Var}(\Delta) = \begin{cases} \Theta(n^{\frac{5}{2}(3-\tau)}) & \tau < \frac{7}{3}, \\ \Theta(n^{4-\tau}) & \tau \geq \frac{7}{3}. \end{cases} \quad (8.4.7)$$

To investigate whether the triangle motif is self-averaging, we need to compare the variance to the second moment of the number of triangles, which results in

$$\frac{\text{Var}(\Delta)}{\mathbb{E}[\Delta]^2} = \begin{cases} \Theta(n^{\frac{1}{2}(\tau-3)}), & \tau < \frac{7}{3}, \\ \Theta(n^{2\tau-5}), & \tau \geq \frac{7}{3}. \end{cases} \quad (8.4.8)$$

Therefore,

$$\lim_{n \rightarrow \infty} \frac{\text{Var}(\Delta)}{\mathbb{E}[\Delta]^2} = \begin{cases} 0 & \tau < \frac{5}{2}, \\ \infty & \tau > \frac{5}{2}, \end{cases} \quad (8.4.9)$$

which proves the following proposition:

Proposition 8.1. *The number of triangles in rank-1 inhomogeneous random graphs is self-averaging as long as $\tau < \frac{5}{2}$. When $\tau \geq \frac{5}{2}$ the number of triangles is not self-averaging.*

8.4.2 General motif fluctuations

We now compute the variance of general motifs, similar to the triangle example (see also [84, 147, 171, 180]).

Lemma 8.2. *Let H_1, \dots, H_l denote all motifs that can be constructed by merging two copies of H at at least one vertex. We can then write the variance of the motif count as*

$$\text{Var}(N(H)) = C_1 \mathbb{E}[N(H_1)] + \dots + C_l \mathbb{E}[N(H_l)] + \mathbb{E}[N(H)]^2 O(n^{-1}). \quad (8.4.10)$$

for constants C_1, \dots, C_l .

Proof. Let $\mathbf{i} = (i_1, \dots, i_k)$ be such that $i_p \neq i_q$ when $p \neq q$. We write the variance as

$$\text{Var}(N(H)) = \sum_{\mathbf{i} \in [n]^k} \sum_{\mathbf{j} \in [n]^k} (\mathbb{P}(H_{\mathbf{i}}, H_{\mathbf{j}} \text{ present}) - \mathbb{P}(H_{\mathbf{i}} \text{ present}) \mathbb{P}(H_{\mathbf{j}} \text{ present})), \quad (8.4.11)$$

where $H_{\mathbf{i}}$ present denotes the event that motif H is present on vertices \mathbf{i} . The sum splits into several cases, depending on the overlap of \mathbf{i} and \mathbf{j} . The term where \mathbf{i} and \mathbf{j} do not overlap equals zero, since edge presences of non-overlapping vertices are independent.

Now suppose \mathbf{i} and \mathbf{j} overlap at i_{t_1}, \dots, i_{t_r} and j_{s_1}, \dots, j_{s_r} for some $r > 0$. Then $\mathbb{P}(H_{\mathbf{i}}, H_{\mathbf{j}} \text{ present})$ equals the probability that motif \tilde{H} is present on vertices $\{i_1, \dots, i_k, j_1, \dots, j_k\} \setminus \{j_{s_1}, \dots, j_{s_r}\}$, where \tilde{H} denotes the motif that is constructed by merging two copies of H at i_{t_1} with j_{s_1} , at i_{t_2} with j_{s_2} and so on. Thus, this term can be written as

$$\sum'_{\mathbf{t} \in [n]^{2k-r}} \mathbb{P}(\tilde{H}_{\mathbf{t}} \text{ present}) = \mathbb{E}[N(\tilde{H})], \quad (8.4.12)$$

where \sum' denotes a sum over distinct indices. Furthermore, since the degrees are i.i.d., $\mathbb{P}(H_{\mathbf{i}} \text{ present}) = \mathbb{E}[N(H)] / \binom{n}{k}$, similarly to (8.4.3). Thus,

$$\sum'_{\mathbf{i}, \mathbf{j} \in [n]^k: |\mathbf{i} \cap \mathbf{j}|=r} \mathbb{P}(H_{\mathbf{i}} \text{ present}) \mathbb{P}(H_{\mathbf{j}} \text{ present}) = n^{-r} \mathbb{E}[N(H)]^2 O(1). \quad (8.4.13)$$

Thus, when H_1, \dots, H_l denote all motifs that can be constructed by merging two copies of H at at least one vertex, we can write the variance of the motif count as

$$\text{Var}(N(H)) = C_1 \mathbb{E}[N(H_1)] + \dots + C_l \mathbb{E}[N(H_l)] + \mathbb{E}[N(H)]^2 O(n^{-1}). \quad (8.4.14)$$

where C_i is a combinatorial constant that denotes the number of distinct ways to merge two copies of H into H_i . These constants satisfy [84]

$$\sum_{i=1}^l C_i = \sum_{s=0}^{k-1} \binom{k}{s}^2 (k-s)!. \quad (8.4.15)$$

□

self-averaging for	subfigures of Figure 8.1	subfigures of Figure 8.2
(2,3)	c	a,b,c,d
(2,5/2)	a	f,g,l,o
(2,7/3)		i
-	b,d,e,f,g,h	e,h,j,k,m,n,p,q,r,s,t,u

Table 8.1: The values of $\tau \in (2,3)$ where the motifs of Figures 8.1 and 8.2 are self-averaging.

Using (8.4.10), we can determine for any motif H whether it is self-averaging or not. First, we find all motifs that are created by merging two copies of H . For the triangle motif for example, these motifs are the bow-tie, where two triangles are merged at one single vertex, the diamond of Figure 8.1d, and the triangle itself. We find the order of magnitude of the expected number of these motifs using Theorem 8.1 to obtain the variance of $N(H)$. We divide by $\mathbb{E}[N(H)]^2$, also obtained by Theorem 8.1, and check whether this fraction is diverging or not. Table 8.1 shows for which values of $\tau \in (2,3)$ the motifs on 3, 4 and 5 vertices are self-averaging. For example, the triangle turns out to be self-averaging only for $\tau \in (2,5/2)$.

The following observation shows the importance of the optimal motif structure:

Theorem 8.2. *All self-averaging motifs for any $\tau \in (2,3)$ have optimal free variation structures with vertices of weights $\Theta(\sqrt{n})$ only.*

We prove this theorem in Appendix 8.E. The condition on the optimal motif structure is a necessary condition for being self-averaging, but it is not a sufficient one, as the triangle example shows. Combining the classification of the motifs into self-averaging and non-self-averaging with the classification based on the value of $B_f(H)$ from (8.2.3) as well as the difference between the expected and typical number of motifs yields a classification into the following three types of motifs:

Type I: Motifs with small variance. $\text{Var}(N(H)) / \mathbb{E}[N(H)]^2 \rightarrow 0$ and $B_f(H) = 0$. These motifs only contain vertices of degrees $\Theta(\sqrt{n})$. The number of such rescaled motifs converges to a constant by Theorem 7.2. Furthermore, the variance of the number of motifs is small compared to the second moment, so that the fluctuations of these types of motifs are quite small and vanish in the large network limit. The triangle for $\tau < 5/2$ is an example of such a motif, illustrated in Figures 8.3b and 8.3e.

Type II: Concentrated, non-self-averaging motifs. $\text{Var}(N(H)) / \mathbb{E}[N(H)]^2 \not\rightarrow 0$ and $B_f(H) = 0$. These motifs also only contain vertices of degrees \sqrt{n} . Again, the rescaled number of such motifs converges to a constant in probability by Theorem 7.2. Thus, most network samples contain a similar amount of motifs as n grows large, even though these motifs are non-self-averaging. Still, in rare network samples the number of motifs significantly deviates from its typical number, causing the variance of the number of motifs to be large. Figures 8.3a and 8.3d illustrate this for triangle counts for $\tau \geq 5/2$. The fluctuations are larger than for the concentrated motifs, but most of the samples have motif counts close to the expected value.

Type III: Non-concentrated motifs. $B_f(H) > 0$. These motifs contain hub vertices. The expected and typical number of such motifs therefore scale differently in n . By

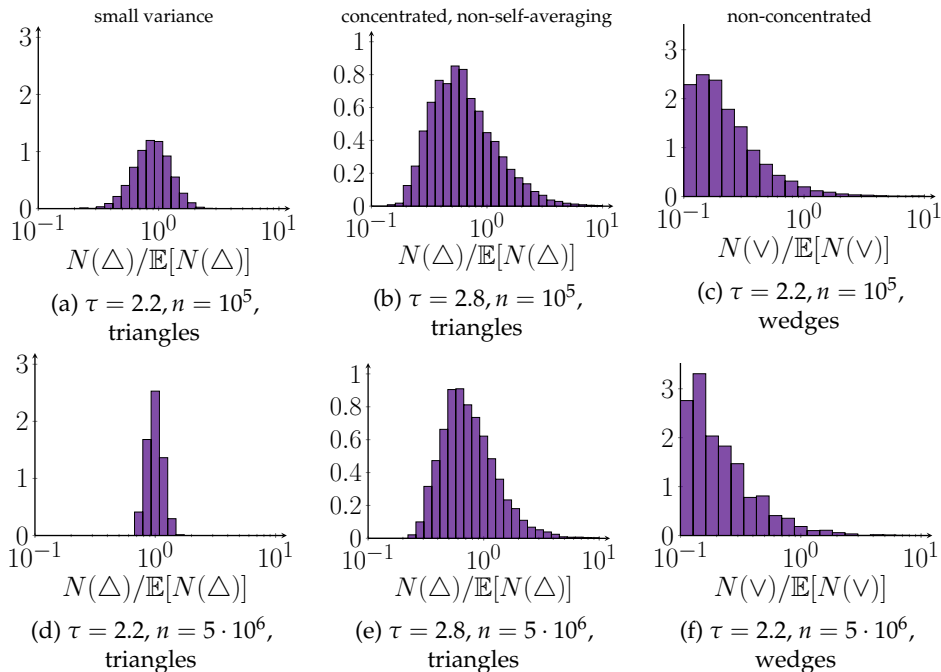


Figure 8.3: Density approximation of the normalized triangle and wedge counts for various values of τ and n , obtained over 10^4 network samples.

	n	m	τ
Gowalla	196591	950327	2.65
Oregon	11174	23409	2.08
Enron	36692	183831	1.97
PGP	10680	24316	2.24
Hep	9877	25998	3.50

Table 8.2: Statistics of the five data sets, where n is the number of vertices, m the number of edges, and τ the power-law exponent fitted by the procedure of [66].

Theorem 8.2, these motifs are non-self-averaging. The rescaled number of such motifs may not converge to a constant, so that two network samples contain significantly different motif counts. Figures 8.3c and 8.3f show that the fluctuations of these motifs are indeed of a different nature, since most network samples have motif counts that are far from the expected value.

8.5 Data

We now compare our results to five real-world networks with heavy-tailed degree distributions, the Gowalla social network [137], the Oregon autonomous systems network [137], the Enron email network [129, 137], the PGP web of trust [40] and the

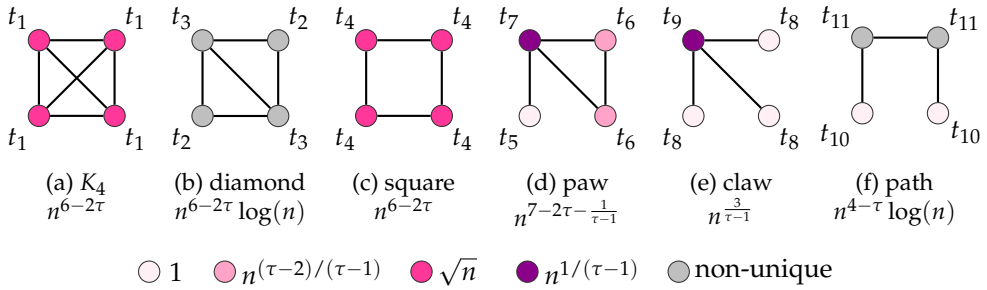


Figure 8.4: Order of magnitude of the typical number of graphlets on 4 vertices. The vertex color indicates the typical vertex degree. The vertex labels indicate the vertex types used in Figure 8.6

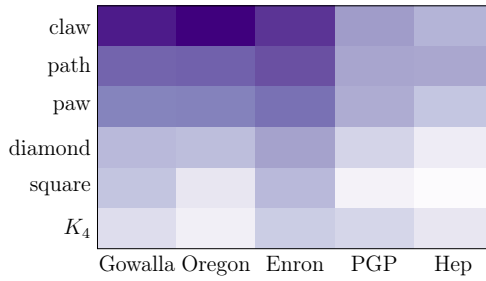


Figure 8.5: Number of graphlets on four vertices in five data sets on logarithmic scale: $\log(N(H)) / \log(n)$. The darker the color, the larger the graphlet count. The ordering of the six different graphlets is from the most occurring in the rank-1 inhomogeneous random graph to the least.

High Energy Physics collaboration network (HEP) [137]. Table 8.2 provides detailed statistics of these data sets. Because the number of motifs can be obtained from the number of graphlets, we focus on graphlet counts. Furthermore, since network data sets contain one single observation of graphlet counts, we compare the data sets to the results on the typical number of graphlets instead of the average number of graphlets, as presented in Figure 8.4. Figure 8.5 shows the graphlet counts in the data sets on a logarithmic scale. The order of the graphlets is from the most occurring graphlet (the claw), to the least occurring graphlet (the complete graph and K_4) in the rank-1 inhomogeneous random graph, see Figure 8.4. The colors indicate that in most data sets the ordering of the motifs follows that of the rank-1 inhomogeneous random graph. It also sheds some light on real-world graphlet formation. Indeed, when deviations arise, these can be often directly linked to the specific nature of the real-world network. In the HEP collaboration network, for example, K_4 occurs more frequently than the square. While this is not predicted by the rank-1 inhomogeneous random graph, it naturally arises due to the frequently occurring collaboration between four authors, which creates K_4 instead of the square.

Figure 8.4 enumerates all possible vertex types in graphlets on 4 vertices. By our theorems, vertex types t_7 and t_9 have typical degrees proportional to $n^{1/(\tau-1)}$

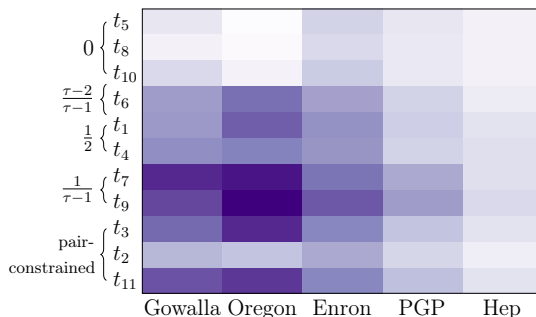


Figure 8.6: Typical degrees of the vertex types displayed in Figure 8.4 in 5 data sets on logarithmic scale. The darker the color, the larger the degree. The curly brackets indicate the typical degree exponent of the vertex type in the rank-1 inhomogeneous random graph.

in the rank-1 inhomogeneous random graph, vertex types t_1 and t_4 typically have degrees proportional to \sqrt{n} , vertex type t_6 typically has degree proportional to $n^{(\tau-2)/(\tau-1)}$ and vertex types t_5, t_8, t_{10}, t_{11} typically have constant degree in the rank-1 inhomogeneous random graph. Vertex types t_2, t_3 and t_{11} do not have a unique optimizer. The degrees of these vertex types are pair-constrained (see the proof of Lemma 8.1). Figure 8.6 shows the typical degree of all 11 vertex types in real-world data. We see that vertices that have typical degree $n^{1/(\tau-1)}$ in the rank-1 inhomogeneous random graph also have the highest degree in the data sets (the darkest colors). Vertices with typical degree 1 in the rank-1 inhomogeneous random graph have the lowest degree in the five data sets. Vertices of typical degrees \sqrt{n} and $n^{(\tau-2)/(\tau-1)}$ in the rank-1 inhomogeneous random graph have moderate degrees in the five data sets. Thus, even though real-world networks are not formed by the rank-1 inhomogeneous random graph, the typical degrees of vertices in a graphlet roughly follow the same orders of magnitude.

Furthermore, the HEP collaboration network does not have a large distinction between the degrees of the different vertex types. This may be related to the fact that this network has less heavy-tailed degrees than the other networks (see Table 8.2). In general, the typical degrees of graphlets in these data sets roughly follow the same orders of magnitude as the typical degrees predicted by the variational principle in the rank-1 inhomogeneous random graph.

8.6 Conclusion

By developing a variational principle for the optimal degree composition of motifs in the rank-1 inhomogeneous random graph, we have identified the asymptotic growth of expected motif counts and their fluctuations for all motifs. This has allowed us to determine for which values of the degree exponent $\tau \in (2, 3)$ the number of motifs is self-averaging. We have further divided the non-self-averaging motifs into two classes with substantially different concentration properties.

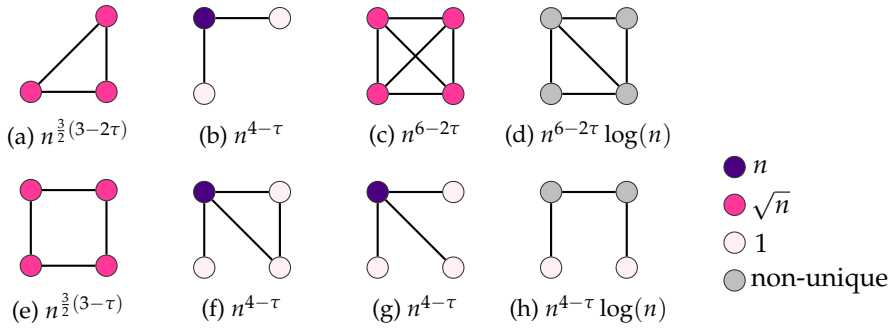


Figure 8.7: Order of magnitude of the expected number of graphlets on 3 and 4 vertices, where the vertex color indicates the optimal vertex degree. Vertices where the optimizer is not unique are gray.

Hub vertices in optimal motif structures cause wild degree fluctuations and non-self-averaging behavior, so that large differences between the average motif count and the motif count in one sample of the random network model arise. Non-self-averaging motifs without a hub vertex show milder fluctuations.

We expect that the variational principle can be extended to different random graph models, such as the hyperbolic random graph and random intersection graphs. For example, for the hyperbolic random graph, the optimal structure of complete graphs is known to be \sqrt{n} degrees [85] like in the rank-1 inhomogeneous random graph, but the optimal structures of other motifs are yet unknown. In Chapter 9 we show how to apply the variational principle to the preferential attachment model.

8.A Optimal graphlet structures

8.B Proof of Theorem 8.1

We now investigate the relation between the expected number of motifs and the optimization problem (8.2.3). Let $N(H, \alpha, \varepsilon)$ denote the number of times motif H occurs on vertices with degrees $[\varepsilon, 1/\varepsilon](n^{\alpha_i})_{i \in [k]}$ and let α^* be defined as

$$\alpha_i^* = \begin{cases} 0 & \text{if } i \in S_1, \\ 1 & \text{if } i \in S_2, \\ \frac{1}{2} & \text{if } i \in S_3. \end{cases} \tag{8.B.1}$$

Then, the following theorem provides a more detailed version of Theorem 8.1:

Theorem 8.3 (General motifs, expectation). *Let H be a motif on k vertices such that the solution to (8.B.13) is unique. Then, for any $\alpha \neq \alpha^*$ and $0 < \varepsilon < 1$,*

$$\frac{\mathbb{E}[N(H, \alpha, \varepsilon)]}{\mathbb{E}[N(H, \alpha^*, \varepsilon)]} \rightarrow 0. \tag{8.B.2}$$

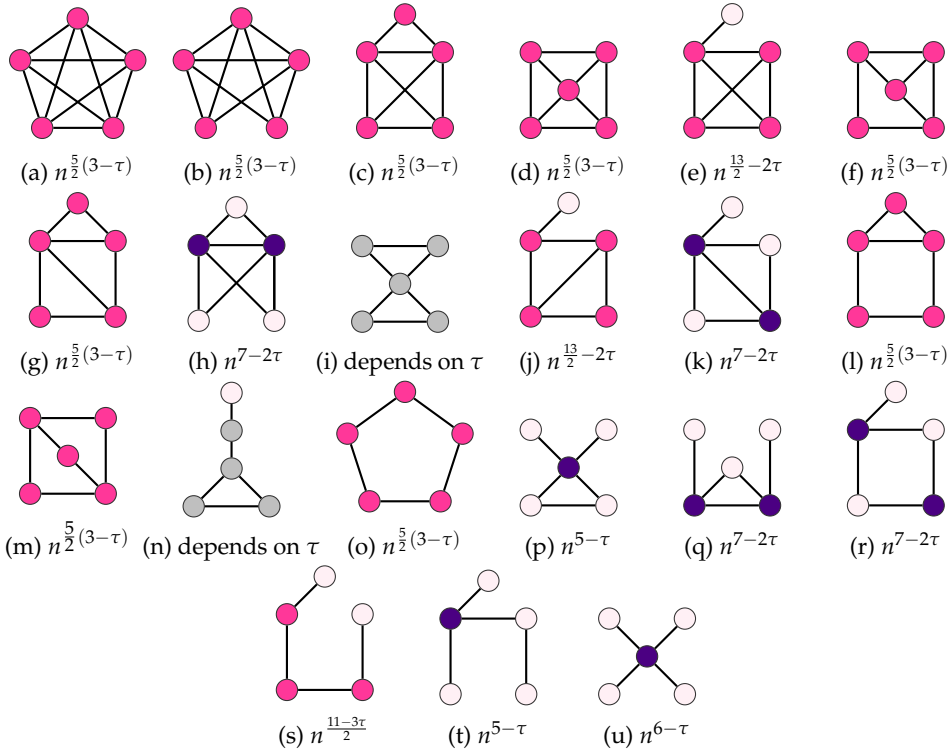


Figure 8.8: Order of magnitude of the expected number of graphlets on 5 vertices, where the vertex color indicates the optimal vertex degree, as in Figure 8.7.

Furthermore, for some function $f(\varepsilon)$ not depending on n ,

$$\frac{\mathbb{E}[N(H, \boldsymbol{\alpha}^*, \varepsilon)]}{n^{\frac{3-\tau}{2}k + \frac{\tau-2}{2}B(H)}} = f(\varepsilon)\Theta(1). \tag{8.B.3}$$

We now prove Theorem 8.3. In the rank-1 inhomogeneous random graph, the connection probability of vertices with weights h_i and h_j equals

$$p(h_i, h_j) = \min(h_i h_j / (\mu n), 1). \tag{8.B.4}$$

Assume that $h_i = \Theta(n^{\alpha_i})$ for some $\alpha_i \geq 0$ for all i . Then, the probability that motif H exists on vertices with weights $\mathbf{h} = (n^{\alpha_i})_{i \in [k]}$ satisfies

$$\mathbb{P}(H \text{ present on weights } \mathbf{h}) = \Theta\left(\prod_{(i,j) \in E_H: \alpha_i + \alpha_j < 1} n^{\alpha_i + \alpha_j - 1}\right). \tag{8.B.5}$$

The weights are an i.i.d. sample from a power-law distribution. Therefore,

$$\mathbb{P}(h_i \in [\varepsilon, 1/\varepsilon](\mu n)^\alpha) = \int_{\varepsilon(\mu n)^\alpha}^{1/\varepsilon(\mu n)^\alpha} cx^{-\tau} dx = K(\varepsilon)(\mu n)^{\alpha(1-\tau)} \tag{8.B.6}$$

for some constant $K(\varepsilon)$ not depending on n . The expected number of vertices with degrees in $[\varepsilon, 1/\varepsilon](\mu n)^\alpha$ scales as $\Theta(n^{(1-\tau)\alpha+1})$. Then, the expected number of sets of vertices with degrees in $[\varepsilon, 1/\varepsilon](n^{\alpha_i})_{i \in [k]}$ scales as

$$\Theta\left(n^{k+(1-\tau)\sum_i \alpha_i}\right). \tag{8.B.7}$$

Combining (8.B.5) and (8.B.7) yields that the contribution from vertices with degrees $n^\alpha = (n^{\alpha_i})_{i \in [k]}$ to the expected number of motifs, $\mathbb{E}[N(H, \alpha), \varepsilon]$ is

$$\mathbb{E}[N(H, \alpha, \varepsilon)] = \Theta\left(n^{k+(1-\tau)\sum_i \alpha_i} \prod_{(i,j) \in E_H: \alpha_i + \alpha_j < 1} n^{\alpha_i + \alpha_j - 1}\right). \tag{8.B.8}$$

The maximum contribution is obtained for α_i that maximize

$$\max(1 - \tau) \sum_i \alpha_i + \sum_{(i,j) \in E_H: \alpha_i + \alpha_j < 1} \alpha_i + \alpha_j - 1 \tag{8.B.9}$$

for $\alpha_i \geq 0$. The following lemma shows that this optimization problem attains its maximum for highly specific values of α :

Lemma 8.3 (Maximum contribution to expected number of motifs). *Let H be a connected graph on k vertices. If the solution to (8.B.9) is unique, then the optimal solution satisfies $\alpha_i \in \{0, \frac{1}{2}, 1\}$ for all i . If it is not unique, then there exist at least 2 optimal solutions with $\alpha_i \in \{0, \frac{1}{2}, 1\}$ for all i .*

Proof. Defining $\beta_i = \alpha_i - \frac{1}{2}$ yields for (8.B.9)

$$\max \frac{1-\tau}{2}k + (1-\tau) \sum_i \beta_i + \sum_{(i,j) \in E_H: \beta_i + \beta_j < 0} \beta_i + \beta_j, \tag{8.B.10}$$

over all possible values of $\beta_i \geq -\frac{1}{2}$. Then, we have to prove that $\beta_i \in \{-\frac{1}{2}, 0, \frac{1}{2}\}$ for all i in the optimal solution. Since (8.B.10) is a piecewise linear function in β , if (8.B.10) has a unique maximum, it must be attained at the boundary for β_i or at a border of one of the linear sections. Thus, any unique optimal value of β_i satisfies $\beta_i = -\frac{1}{2}$ or $\beta_i + \beta_j = 0$ for some j . This implies that $\beta_i \leq \frac{1}{2}$.

From there, we can follow the lines of Step 2 and 3 of the proof of Lemma 7.2, which analyzes the same equation. This shows that any unique solution does not have vertices with $|\beta_i| \in (0, \frac{1}{2})$, and that any non-unique optimal solution can be transformed into two different optimal solutions with $\beta_i \in \{-\frac{1}{2}, 0, \frac{1}{2}\}$. \square

Proof of Theorem 8.3. We first rewrite (8.B.9) using Lemma 8.3. By Lemma 8.3, the maximal value of (8.B.9) is attained by partitioning V_H into the sets S_1, S_2, S_3 such that vertices in S_1 have $\alpha_i = 0$, vertices in S_2 have $\alpha_i = 1$ and vertices in S_3 have $\alpha_i = \frac{1}{2}$. Then, the edges with $\alpha_i + \alpha_j < 1$ are edges inside S_1 and edges between S_1 and S_3 . If we denote the number of edges inside S_1 by E_{S_1} and the number of edges between S_1 and S_3 by E_{S_1, S_3} , then we can rewrite (8.B.9) as

$$\max_p (1 - \tau)(|S_2| + \frac{1}{2}|S_3|) - E_{S_1} - \frac{1}{2}E_{S_1, S_3} \tag{8.B.11}$$

over all partitions \mathcal{P} of the vertices of H into S_1, S_2, S_3 . Using that $|S_3| = k - |S_1| - |S_2|$ yields

$$\max_{\mathcal{P}} \frac{1-\tau}{2}k + \frac{\tau-1}{2} \left(|S_1| - |S_2| - \frac{2E_{S_1} + E_{S_1, S_3}}{\tau-1} \right), \quad (8.B.12)$$

Since k is fixed and $\tau - 1 > 0$, maximizing (8.B.9) is equivalent to maximizing

$$B_f(H) = \max_{\mathcal{P}} \left[|S_1| - |S_2| - \frac{2E_{S_1} + E_{S_1, S_3}}{\tau-1} \right]. \quad (8.B.13)$$

Furthermore, by Lemma 8.3, the optimal value of (8.B.13) is unique if and only if the solution to (8.B.9) is unique.

Let α^* be the unique optimizer of (8.B.9). Then, by (8.B.8), for any $\alpha \neq \alpha^*$

$$\frac{\mathbb{E}[N(H, \alpha, \varepsilon)]}{\mathbb{E}[N(H, \alpha^*, \varepsilon)]} = \Theta(n^{-\eta}) \quad (8.B.14)$$

for some $\eta > 0$. Combining this with (8.B.13) proves the first part of the theorem. By (8.B.8), the contribution of the maximum is then given by

$$\mathbb{E}[N(H, \alpha^*, \varepsilon)] = n^k n^{\frac{1-\tau}{2}(k+B_f(H))} = n^{\frac{3-\tau}{2}k + \frac{\tau-1}{2}B_f(H)}, \quad (8.B.15)$$

which proves the second part of the theorem. \square

8.C Typical motif counts

We now present an equivalent version of Theorem 8.3 for typical motif counts, which is similar to Theorem 7.1. Define

$$B_t(H) = \max_{\mathcal{P}} |S_1| - |S_2| - \frac{2E_{S_1} + E_{S_1, S_3} + E_{S_1, 1} - E_{S_2, 1}}{\tau-1}, \quad (8.C.1)$$

where the maximum is over all partitions \mathcal{P} of $V_H \setminus V_1$ into three sets, where V_1 denotes the set of vertices in H of degree 1. Vertices in S_1 correspond to vertices that have degree proportional to $n^{(\tau-2)/(\tau-1)}$ in the graph, vertices in S_2 correspond to the maximal degree vertices with degrees proportional to $n^{1/(\tau-1)}$, and S_3 corresponds to the vertices of \sqrt{n} degrees. Let S_1^*, S_2^*, S_3^* denote the optimal sets of (8.C.1) and define

$$\alpha_i^* = \begin{cases} (\tau-2)/(\tau-1) & i \in S_1^*, \\ 1/(\tau-1) & i \in S_2^*, \\ \frac{1}{2} & i \in S_3^*, \\ 0 & i \in V_1. \end{cases} \quad (8.C.2)$$

The following theorem gives the scaling of the typical number of motifs in rank-1 inhomogeneous random graphs:

Theorem 8.4 (Typical graphlets). *Let H be a motif on k vertices such that the solution to (8.C.1) is unique. Then,*

(i) For any ε_n such that $\lim_{n \rightarrow \infty} \varepsilon_n = 0$,

$$\frac{N(H, \boldsymbol{\alpha}^*, \varepsilon_n)}{N(H)} \xrightarrow{\mathbb{P}} 1. \quad (8.C.3)$$

(ii) Furthermore, for any fixed $0 < \varepsilon < 1$,

$$\frac{N(H, \boldsymbol{\alpha}^*, \varepsilon)}{n^{\frac{3-\tau}{2}(k_{2+} + B_t(H)) + \frac{1}{2}k_1}} = f(\varepsilon)\Theta_{\mathbb{P}}(1) \quad (8.C.4)$$

for some function $f(\varepsilon)$ not depending on n . Here k_{2+} denotes the number of vertices in H of degree at least 2, and k_1 the number of degree one vertices in H .

Proof. The proof of Theorem 8.4 follows the same lines as the proof of Theorem 7.1 for the erased configuration model instead of the rank-1 inhomogeneous random graph. That proof relies on the fact that the connection probability between vertices with degrees D_i and D_j in the erased configuration model has the same order of magnitude as $\min(D_i D_j / (\mu n), 1)$, which is the same as the connection probability in the rank-1 inhomogeneous random graph. Thus, the proof of 7.1 also holds for the rank-1 inhomogeneous random graph. \square

8.D Graphlets

We now focus on graphlet counting, or induced subgraph counting. Therefore, edges that are not present in H are also required not to be present in the graph. Because the probability that two edges between vertices of high degree are present equals one (see (8.B.4)), this puts a constraint on the number of vertices that typically have high degree. When we again let the degrees in the graphlet scale as n^{α_i} , we see that the probability that an edge (i, j) is present equals one as soon as $\alpha_i + \alpha_j > 1$. Thus, for the expected number of graphlets, the optimization problem corresponding to (8.B.9) becomes

$$\begin{aligned} & \max(1 - \tau) \sum_i \alpha_i + \sum_{(i,j) \in E_H: \alpha_i + \alpha_j < 1} \alpha_i + \alpha_j - 1 \\ & \text{s.t. } \alpha_i + \alpha_j \leq 1 \quad \forall (i, j) \notin E_H, \end{aligned} \quad (8.D.1)$$

where E_H denotes the edge set of H . Again, this optimization problem is maximized for $\alpha_i \in 0, \frac{1}{2}, 1$, so that similarly, the optimization problem corresponding to (8.B.13) including the extra constraint then becomes

$$\begin{aligned} B_{g,f}(H) &= \max_{\mathcal{P}} \left[|S_1| - |S_2| - \frac{2E_{S_1} + E_{S_1, S_3}}{\tau - 1} \right], \\ & \text{s.t. } (u, v) \in E_H \quad \forall u \in S_2, v \in S_2 \cup S_3. \end{aligned} \quad (8.D.2)$$

This optimization problem again finds the most likely degrees of vertices that together form the graphlet H . Vertices in S_1 have degrees proportional to a constant, vertices in S_2 have degrees proportional to n and vertices in S_3 have degrees proportional to \sqrt{n} . Using this optimization problem, we are also able to find the scaling of the number of graphlets, by replacing the optimum B_f and B_t in Theorems 8.3 and 8.4 by their counterparts for graphlets, $B_{g,f}$ and $B_{g,t}$ respectively.

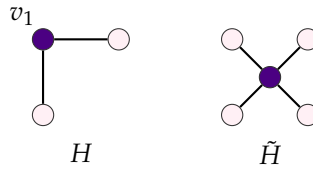


Figure 8.9: Construction of \tilde{H} when H is a path of length 3. \tilde{H} is constructed by merging two copies of H at vertex v_1 with most likely degree n .

Graphlet fluctuations

Fluctuations of the number of graphlets can be studied similarly as motif fluctuations. Again, (8.4.10) holds, but now it only includes graphlets H_1, \dots, H_l that can be constructed by merging two copies of H at one or more vertices, with the additional constraint that after merging the copies of H , both copies still form induced subgraphs of H . As an example, consider 2-path fluctuations. The subgraphs that can be constructed from merging two 2-paths are the subgraphs in Figure 8.8s, 8.8t, 8.8u, 8.7a, 8.7b, 8.7e, 8.7f, Figure 8.7g and 8.7h. However, the merged subgraphs of Figure 8.7a and Figure 8.7f do not contain two induced copies of graphlets, and therefore these subgraphs are excluded from equation (8.4.10). Other than that, the procedure to determine for any graphlet whether it is self-averaging or not is the same as the procedure for motifs, using the variational principle for graphlets (8.D.2) to find the expected order of magnitude of the number of merged graphlets.

8.E Proof of Theorem 8.2

Let H be a motif on k vertices such that the optimal contribution of Theorem 8.1 contains vertices that do not have optimal contribution from \sqrt{n} vertices. A similar analysis as the triangle example shows that the only motif on 2 vertices (the 2-path) is non-self-averaging. Thus, we may assume that $k \geq 3$. Let the expected number of motifs of type H as predicted in Theorem 8.1 be denoted by $n^{f(H)}$, and the optimal contribution from vertices $1, \dots, k$ in H by $(n^{\alpha_i})_{i \in [k]}$. By Theorem 8.1,

$$\mathbb{E}[N(H)] \propto \mathbb{E} \left[\# \text{ vertices of weights } (n^{\alpha_i})_{i \in [k]} \right] \mathbb{P}(H \text{ present on weights } (n^{\alpha_i})_{i \in [k]}) \quad (8.E.1)$$

Suppose vertex $v_1 \in H$ has optimal contribution of weight n vertices, that is $\alpha_1 = 1$. Then, we study the contribution to the variance in (8.4.10) from the motif \tilde{H} which is the motif on $2k - 1$ vertices where two copies of H are merged at their vertex v_1 (see Figure 8.9 for an example). We now investigate the expected number of \tilde{H} motifs. In particular, we study the contribution to the expected number of \tilde{H} motifs from vertices with weights of the order of magnitude $(n^{\beta_i})_{i \in [2k-1]}$ with

$$\beta_i = \alpha_{t(i)}, \quad (8.E.2)$$

where $t(i)$ is the vertex in H corresponding to vertex i in \tilde{H} . Figure 8.9 gives an example of this contribution when H is a path with 3 vertices. Because \tilde{H} is formed

by two copies of H without overlapping edges

$$\mathbb{P}(\tilde{H} \text{ present on weights } (n^{\beta_i})_{i \in [2k-1]}) = \mathbb{P}(H \text{ present on weights } (n^{\alpha_i})_{i \in [k]})^2. \quad (8.E.3)$$

Furthermore, the only difference between the vertices in \tilde{H} and two separate versions of H , is that \tilde{H} contains one less vertex of weight proportional to n . The expected number of vertices with weight proportional to n is given by $n^{2-\tau}$. Therefore,

$$\mathbb{E}[\# \text{ vertices of weights } (n^{\beta_i})_{i \in [2k-1]}] \propto \frac{\mathbb{E}[\# \text{ vertices of weights } (n^{\alpha_i})_{i \in [k]}]^2}{n^{2-\tau}}. \quad (8.E.4)$$

Thus,

$$\begin{aligned} \mathbb{E}[N(\tilde{H})] &\geq \mathbb{P}(H \text{ present on weights } (n^{\alpha_i})_{i \in [k]})^2 \\ &\quad \times \mathbb{E}[\# \text{ vertices of weights } (n^{\alpha_i})_{i \in [k]}]^2 n^{\tau-2} \\ &\propto \mathbb{E}[N(H)]^2 n^{\tau-2}, \end{aligned} \quad (8.E.5)$$

where the last step uses Theorem 8.3. Combining this with (8.4.10) results in

$$\frac{\text{Var}(N(H))}{\mathbb{E}[N(H)]^2} \geq \frac{\mathbb{E}[N(\tilde{H})]}{\mathbb{E}[N(H)]^2} \geq n^{\tau-2}, \quad (8.E.6)$$

which diverges, because $\tau \in (2,3)$. Thus, if the optimal contribution to H satisfies $S_2 \neq \emptyset$, H cannot be self-averaging for $\tau \in (2,3)$.

Now we study the case where H has optimal contribution with $S_1 \neq \emptyset$, but no vertices of weights proportional to n so that $S_2 = \emptyset$. Let $v \in S_1$. The contribution from v to (8.2.3) is

$$1 - \frac{2d_{v,S_1} + d_{v,S_3}}{\tau - 1}, \quad (8.E.7)$$

where d_{v,S_i} denotes the number of edges from v to S_i . Moving v to S_3 would change the contribution to $-d_{v,S_1}/(\tau - 1)$. Because $v \in S_1$ is the optimal contribution,

$$-\frac{d_{v,S_1}}{\tau - 1} < 1 - \frac{2d_{v,S_1} + d_{v,S_3}}{\tau - 1}, \quad (8.E.8)$$

or $d_{v,S_1} + d_{v,S_3} \leq \tau - 1$, so that $d_{v,S_1} + d_{v,S_3} \in \{0, 1\}$. Thus, every vertex in S_1 has at most 1 edge to other vertices in S_1 or vertices in S_3 . Since we have assumed that $S_2 = \emptyset$, and $k > 2$, this means that all vertices in S_1 have degree 1 inside the motif, and are connected to a vertex in S_3 . W.l.o.g. assume that v_1 is a vertex such that $v_1 \in S_3$ and v_1 has at least one connection to a vertex in S_1 . As in the previous proof, we consider \tilde{H} constructed by merging two copies of H at v_1 , as illustrated in Figure 8.10. Define $(\alpha_i)_{i \in [k]}$ as the maximal contribution to H . We define

$$\beta_i = \begin{cases} \alpha_{t(i)} & t(i) \neq v_1, \\ 1 & t(i) = v_1. \end{cases} \quad (8.E.9)$$

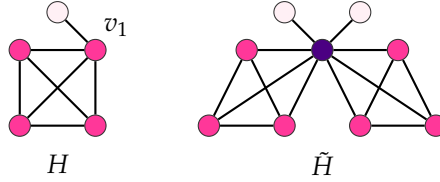


Figure 8.10: Example of the construction of \tilde{H} . \tilde{H} is constructed by merging two copies of H at vertex v_1 : a vertex with most likely degree \sqrt{n} that is connected to a vertex of degree 1.

That is, we study the contribution where all vertices in \tilde{H} except v_1 have the same weight as in their counterpart in H . The weight of v_1 is proportional to n , whereas the counterpart of v_1 in H had weight proportional to \sqrt{n} , illustrated in Figure 8.10. We study the contribution to $\mathbb{E}[N(\tilde{H})]$ from vertices of weights $(n^{\beta_i})_{i \in [2k-1]}$. We now compare the probability that \tilde{H} exists on vertices of weights $(n^{\beta_i})_{i \in [2k-1]}$ to the probability that two copies of H exist on weights $(n^{\alpha_i})_{i \in [k]}$. The difference between these two probabilities is that vertex v_1 in \tilde{H} has weight n instead of \sqrt{n} in H . In H , v_1 is connected to at least one vertex of weight proportional to 1. The probability of this connection to be present is proportional to $n^{-1/2}$. In \tilde{H} , v_1 has weight n , so that the probability that the corresponding connections occur in \tilde{H} is proportional to 1. The connection probabilities of vertices not connected to v_1 do not change, so that

$$\mathbb{P}(\tilde{H} \text{ present on weights } (n^{\beta_i})_{i \in [2k-1]}) \geq \frac{\mathbb{P}(H \text{ present on weights } (n^{\alpha_i})_{i \in [k]})^2}{n^{-1}}. \quad (8.E.10)$$

Since the difference between the vertices of two copies of H and \tilde{H} is that we remove two vertices of weight \sqrt{n} and add one vertex of weight proportional to n , we obtain

$$\mathbb{E}[\#\text{ vertices of weights } (n^{\beta_i})_{i \in [2k-1]}] \propto \frac{\mathbb{E}[\#\text{ vertices of weights } (n^{\alpha_i})_{i \in [k]}]^2 n^{2-\tau}}{n^{3-\tau}}, \quad (8.E.11)$$

using that the number of vertices of weight proportional to \sqrt{n} is $n^{(3-\tau)/2}$ by (8.B.7) and the number of vertices of weight proportional to n scales as $n^{2-\tau}$. This results in

$$\mathbb{E}[N(\tilde{H})] \geq \mathbb{P}(H \text{ present on weights } (n^{\alpha_i})_{i \in [k]})^2 \mathbb{E}[\#\text{ vertices of weights } (n^{\alpha_i})_{i \in [k]}]^2 \propto \mathbb{E}[N(H)]^2 \quad (8.E.12)$$

so that by (8.4.10)

$$\frac{\text{Var}(N(H))}{\mathbb{E}[N(H)]^2} \geq \frac{\mathbb{E}[N(\tilde{H})]}{\mathbb{E}[N(H)]^2} \propto 1, \quad (8.E.13)$$

which does not converge to zero, so that the motif is not self-averaging. \square

9 Subgraphs in preferential attachment models

Based on:
Subgraphs in preferential attachment models
A. Garavaglia and C. Stegehuis, arXiv:1806.10406

In Chapters 7 and 8, we have investigated the typical and the expected number of subgraphs in erased configuration models and rank-1 inhomogeneous random graphs. In this chapter, we consider expected subgraph counts in scale-free preferential attachment models. We find the scaling of the expected number of times that a specific subgraph occurs as a power of the number of vertices. Similarly to Chapters 7 and 8 we define an optimization problem that finds the optimal subgraph structure and enables us to find the expected number of such subgraphs. This optimization problem optimizes the indices of the vertices that together span the subgraph and uses the representation of the preferential attachment model as a Pólya urn model.

9.1 Introduction

In this chapter, we analyze subgraph counts for the preferential attachment model (PAM) with parameters m and δ , described in Section 1.1.5. We focus on the case where $m \geq 2$ is fixed, and $\delta > -m$. Taking $\delta \in (-m, 0)$ results in a power-law degree distribution with exponent $\tau \in (2, 3)$, as observed in many real-world networks. An important difference between the preferential attachment model and most other random graph models is that edges can be interpreted as directed. Thus, analyzing subgraph counts in PAMs allows us to study *directed* subgraphs. This is a major advantage of the PAM over other random graph models, since most real-world network subgraphs in for example biological networks are directed as well [153, 200].

Most existing work on counting subgraphs in PAMs focuses on counting triangles. Bollobás and Riordan [44] prove that for any integer-valued function $T(t)$ there exists a PAM with $T(t)$ triangles, where t denotes the number of vertices in PAM. They further show that the clustering coefficient in the Albert-Barabási model (where $\delta = 0$) is of order $(\log t)^2/t$, while the expected number of triangles is of order $(\log t)^3$ and more generally, the expected number of cycles of length l scales as $(\log t)^l$.

Eggmann and Noble [77] consider $\delta > 0$, so that $\tau > 3$ and investigate the number of subgraphs for $m = 1$ (so subtrees). For $m \geq 2$ they study the number of triangles and the clustering coefficient, proving that the expected number of triangles is of order $\log t$ while the clustering coefficient is of order $\log t/t$, which is different than

the results in [44]. Our result on general subgraphs for any value of δ in Theorem 9.1 explains this difference (in particular, we refer to (9.2.1)).

In a series of papers [183–185], Prokhorenkova et al. proved results on the clustering coefficient and the number of triangles for a broad class of PAMs, assuming general properties on the attachment probabilities. These attachment probabilities are in a form that increases the probability of creating a triangle. They prove that in this setting the number of triangles is of order t , while the clustering coefficient behaves differently depending on the exact attachment probabilities.

9.1.1 Our contribution

For every directed subgraph, we obtain the scaling of the expected number of such subgraphs in the PAM, generalizing the above results on triangles, cycles and subtrees. Furthermore, we identify the most likely degrees of vertices participating in a specific subgraph, which shows that subgraphs in the PAM are typically formed between vertices with degrees of a specific order of magnitude. The order of magnitude of these degrees can be found using an optimization problem. For general subgraphs, our results provide the scaling of the expected number of subgraphs in the network size t . For the triangle subgraph, we obtain precise asymptotic results on the subgraph count, which allows us to study clustering in the PAM.

We use the interpretation of the PAM as a Pólya urn graph as in [20]. This interpretation allows to view the edges as being present independently, so that we are able to obtain the probability that a subgraph H is present on a specific set of vertices.

9.1.2 Organization of the chapter

We first describe the specific PAM we study in Section 9.1.3. After that, we present our result on the scaling on the number of subgraphs in the PAM and the exact asymptotics of the number of triangles in Section 9.2. Section 9.3 provides an important ingredient for the proof of the scaling of the expected number of subgraphs: a lemma that describes the probability that a specific subgraph is present on a subset of vertices. After that, we prove our main results in Sections 9.4–9.6. Finally, Section 9.7 gives the conclusions and the discussion of our results.

9.1.3 Model

As mentioned in Section 9.1, different versions of PAMs exist. In this chapter, we again consider a modification of [20, Model 3]:

Definition 9.1 (Sequential PAM). *Fix $m \geq 1$, $\delta > -m$. Then $(\text{PA}_t(m, \delta))_{t \in \mathbb{N}}$ is a sequence of random graphs defined as follows:*

- for $t = 1$, $\text{PA}_1(m, \delta)$ consists of a single vertex with no edges;
- for $t = 2$, $\text{PA}_2(m, \delta)$ consists of two vertices with m edges between them;
- for $t \geq 3$, $\text{PA}_t(m, \delta)$ is constructed recursively as follows: conditioning on the graph at time $t - 1$, we add a vertex t to the graph, with m new edges. Edges start from vertex t

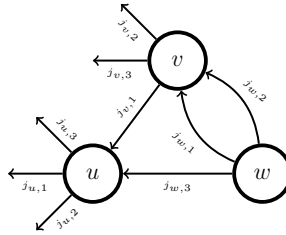


Figure 9.1: Two labeled triangles.

and, for $j = 1, \dots, m$, they are attached sequentially to vertices $E_{t,1}, \dots, E_{t,m}$ chosen with the following probability:

$$\mathbb{P}(E_{t,j} = i \mid \text{PA}_{t-1,j-1}(m, \delta)) = \begin{cases} \frac{D_i(t-1) + \delta}{2m(t-2) + (t-1)\delta} & j = 1, \\ \frac{D_i(t-1, j-1) + \delta}{2m(t-2) + (j-1) + (t-1)\delta} & j = 2, \dots, m. \end{cases} \quad (9.1.1)$$

In (9.1.1), $D_i(t-1)$ denotes the degree of i in $\text{PA}_{t-1}(m, \delta)$, while $D_i(t-1, j-1)$ denotes the degree of vertex i after the first $j-1$ edges of vertex t have been attached. Here we assume that $\text{PA}_{t-1,0} = \text{PA}_{t-1}$.

To keep notation light, we write PA_t instead of $\text{PA}_t(m, \delta)$ throughout the rest of the chapter. The first term in the denominator of (9.1.1) describes the total degree of the first $t-1$ vertices in $\text{PA}_{t-1,j-1}$ when $t-1$ vertices are present and $j-1$ edges have been attached. The term $(t-1)\delta$ in the denominator comes from the fact that there are $t-1$ vertices to which an edge can attach. We do not allow for self-loops, but we do allow for multiple edges.

The PAM of Definition 9.1 generates a random graph where the asymptotic degree sequence is close to a power law [107, Lemma 4.7], where the degree exponent τ satisfies

$$\tau = 3 + \delta/m. \quad (9.1.2)$$

Labeled subgraphs. As mentioned before, the PAM in Definition 9.1 is a multigraph, i.e., any pair of vertices may be connected by m different edges. One could *erase* multiple edges in order to obtain a simple graph, similarly to [50] for the configuration model. In the PAM in Definition 9.1 there are at most m edges between any pair of vertices, so that the effect of erasing multiple edges is small, unlike the in configuration model. We do not erase edges, so that we may count a subgraph on the same set of vertices multiple times. Not erasing edges has the advantage that we do not modify the law of the graph, therefore we can directly use known results on PAM.

More precisely, to count the number of subgraphs, we analyze *labeled* subgraphs, i.e., subgraphs where the edges are specified. In Figure 9.1 we give the example of two labeled triangles on three vertices u, v, w , one consisting of edges $\{j_{v,1}, j_{w,1}, j_{w,3}\}$ and the other one of edges $\{j_{v,1}, j_{w,2}, j_{w,3}\}$. As it turns out, the probability of two labeled subgraphs on the same vertices and different edges being present is *independent of the choice of the edges*. For a more precise explanation, we refer to Section 9.3.1.

9.2 Main results

In this section, we present our results on the number of directed subgraphs in the preferential attachment model. We first define subgraphs in more detail. Let $H = (V_H, E_H)$ be a connected, directed graph. Let $\pi : V_H \mapsto 1, \dots, |V_H|$ be a one-to-one mapping of the vertices of H to $1, \dots, |V_H|$. In the PAM, vertices arrive one by one. We let π correspond to the order in which the vertices in H have appeared in the PAM, that is $\pi(i) < \pi(j)$ if vertex i was created before vertex j . Thus, the pair (H, π) is a directed graph, together with a prescription of the order in which the vertices of H have arrived. We call the pair (H, π) an ordered subgraph.

In the PAM, it is only possible for an older vertex to connect to a newer vertex but not the other way around. This puts constraints on the types of subgraphs that can be formed. We call the ordered subgraphs that can be formed in the PAM attainable. The following definition describes all attainable subgraphs:

Definition 9.2 (Attainable subgraphs). *Let (H, π) be an ordered subgraph with adjacency matrix $A_\pi(H)$, where the rows and columns of the adjacency matrix are permuted by π . We say that (H, π) is attainable if $A_\pi(H)$ defines a directed acyclic graph, where all out-degrees are less or equal than m .*

We now investigate how many of these subgraphs are typically present in the PAM. We introduce the optimization problem

$$\begin{aligned} B(H, \pi) &= \max_{s=0,1,\dots,k} -s + \sum_{i=s+1}^k \left[\frac{\tau-2}{\tau-1} (d_H^{(\text{in})}(\pi^{-1}(i)) - d_H^{(\text{out})}(\pi^{-1}(i))) - d_H^{(\text{in})}(\pi^{-1}(i)) \right] \\ &:= \max_{s=0,1,\dots,k} -s + \sum_{i=s+1}^k \beta(\pi^{-1}(i)), \end{aligned} \quad (9.2.1)$$

where $d_H^{(\text{out})}$ and $d_H^{(\text{in})}$ denote respectively the in- and the out-degree in the subgraph H . Let $N_t(H, \pi)$ denote the number of times the connected graph H with ordering π occurs as a subgraph of a PAM of size t . The following theorem studies the scaling of the expected number of directed subgraphs in the PAM, and relates it to the optimization problem (9.2.1):

Theorem 9.1. *Let H be a directed subgraph on k vertices with ordering π such that (H, π) is attainable and there are r different optimizers to (9.2.1). Then, there exist $0 < C_1 \leq C_2 < \infty$ such that*

$$C_1 \leq \lim_{t \rightarrow \infty} \frac{\mathbb{E}[N_t(H, \pi)]}{t^{k+B(H, \pi)} \log^{r-1}(t)} \leq C_2. \quad (9.2.2)$$

Theorem 9.1 gives the asymptotic scaling of the number of subgraphs where the order in which the vertices appeared in the PAM is known. The total number of copies of H for any ordering, $N_t(H)$, can then easily be obtained from Theorem 9.1:

Corollary 9.1. *Let H be a directed subgraph on k vertices with $\Pi \neq \emptyset$ the set of orderings π such that (H, π) is attainable. Let*

$$B(H) = \max_{\pi \in \Pi} B(H, \pi), \quad (9.2.3)$$

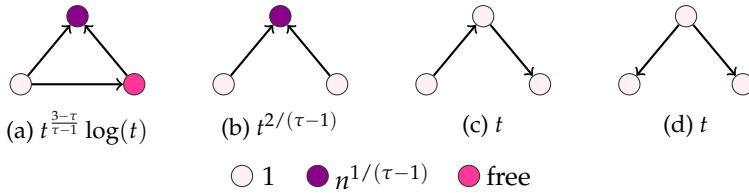


Figure 9.2: Order of magnitude of $N_t(H)$ for all attainable connected directed graphs on 3 vertices and for $2 < \tau < 3$. The vertex color indicates the optimal vertex degree.

and let r^* be the largest number of different optimizers to (9.2.1) among all $\pi \in \Pi$ that maximize (9.2.3). Then, there exist $0 < C_1 \leq C_2 < \infty$ such that

$$C_1 \leq \lim_{t \rightarrow \infty} \frac{\mathbb{E} [N_t(H)]}{t^{k+B(H)} \log^{r^*-1}(t)} \leq C_2. \tag{9.2.4}$$

From Corollary 9.1 it is also possible to obtain the undirected number of subgraphs in a PAM, by summing the number of all possible directed subgraphs that create some undirected subgraph when the directions of the edges are removed.

Interpretation of the optimization problem. The optimization problem (9.2.1) has an intuitive explanation. Assume that π is the identity mapping, so that vertex 1 is the oldest vertex of H , vertex 2 the second oldest and so on. We show in Section 9.3.2 that the probability that an attainable subgraph is present on vertices with indices $u_1 < u_2 < \dots < u_k$ scales as

$$\prod_{i \in [k]} u_i^{\beta(i)}, \tag{9.2.5}$$

with $\beta(i)$ as in (9.2.1). Thus, if for all i , $u_i \propto t^{\alpha_i}$ for some α_i , then the probability that the subgraph is present scales as $t^{\sum_{i \in [k]} \alpha_i \beta(i)}$. The number of vertices with index proportional to t^{α_i} scales as t^{α_i} . Therefore, heuristically, the number of times subgraph H occurs on vertices with indices proportional to $(t^{\alpha_i})_{i \in [k]}$ such that $\alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_k$ scales as

$$t^{\sum_{i \in [k]} (\beta(i)+1)\alpha_i}. \tag{9.2.6}$$

Because the exponent is linear in α_i , the exponent is maximized for $\alpha_i \in \{0, 1\}$ for all i . Because of the extra constraint $\alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_k$ which arises from the ordering of the vertices in the PAM, the maximal value of the exponent is $k + B(H)$. This suggests that the number of subgraphs scales as $t^{k+B(H)}$.

Thus, the optimization problem $B(H)$ finds the most likely configuration of a subgraph in terms of the indices of the vertices involved. If the optimum is unique, the number of subgraphs is maximized by subgraphs occurring on one set of very specific vertex indices. For example, when the maximum contribution is $\alpha_i = 0$, this means that vertices with constant index, i.e., the oldest vertices of the PAM, are most likely to be a member of subgraph H at position i . When $\alpha_i = 1$ is the optimal contribution, vertices with index proportional to t , the newest vertices, are most likely to be a member of subgraph H at position i . When the optimum is not unique, several

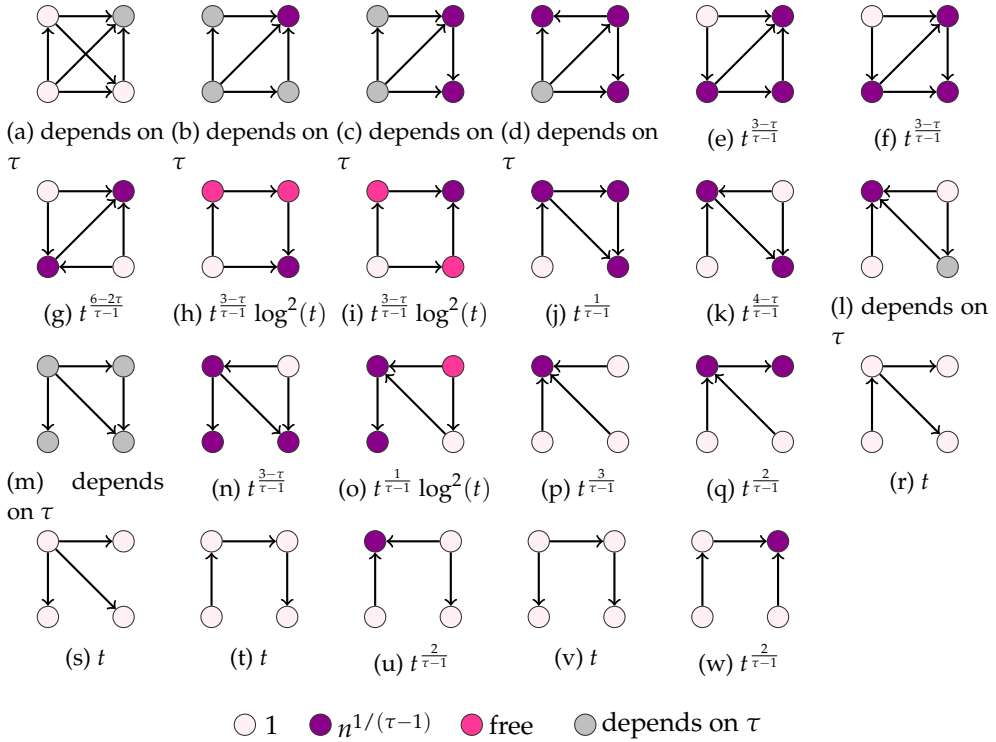


Figure 9.3: Order of magnitude of $N_t(H)$ for all attainable connected directed graphs on 4 vertices and for $2 < \tau < 3$. The vertex color indicates the optimal vertex degree.

maximizers contribute equally to the number of subgraphs, which introduces the extra logarithmic factors in (9.2.2).

Most likely degrees. As mentioned above, the optimization problem (9.2.1) finds the most likely orders of magnitude of the indices of the vertices. When the optimum is unique, the optimum is attained by some vertices of constant index, and some vertices with index proportional to t . The vertices of constant index have degrees proportional to $t^{1/(\tau-1)}$ with high probability [106], whereas the vertices with index proportional to t have degrees proportional to a constant. When the optimum is not unique, the indices of the vertices may have any range, so that the degrees of these vertices in the optimal subgraph structures have degrees ranging between 1 and $t^{1/(\tau-1)}$. Thus, the optimization problem (9.2.1) also finds the optimal subgraph structure in terms of its degrees. The most likely degrees of all directed connected subgraphs on 3 and 4 vertices resulting from Corollary 9.1 and the asymptotic number of such subgraphs for $2 < \tau < 3$ are visualized in Figures 9.2 and 9.3. For some subgraphs, the optimum of (9.2.1) is attained by the same s and therefore the same most likely degrees for all $2 < \tau < 3$, while for other subgraphs the optimum may change with τ .

One such example is the complete graph of size 4. For the directed complete graph, there is only one attainable ordering satisfying Definition 9.2 (as long as $m \geq 3$), so we take the vertices of H to be labeled with this ordering. For $\tau < 5/2$, the optimizer of (9.2.1) is given by $s = 3$ with optimal value $-3 - 3\frac{\tau-2}{\tau-1}$, whereas for $\tau > 5/2$ it is given by $s = 4$ and optimal value -4 . Thus, for $\tau < 5/2$ a complete graph of size four typically contains three hub vertices of degree proportional to $t^{1/(\tau-1)}$ and one vertex of constant degree, and the number of such subgraphs scales as $t^{1-(\tau-2)/(\tau-1)}$ whereas for $\tau > 5/2$ the optimal structure contains four hub vertices instead and the number of such subgraphs scales as a constant.

Fluctuations of the number of subgraphs. In Theorem 9.1 we investigate the expected number of subgraphs, which explains the average number of subgraphs over many PAM realizations. Another interesting question is what the distribution of the number of subgraphs in a PAM realization behaves like. In this chapter, we mainly focus on the expected value of the number of subgraphs, but here we argue that the limiting distribution of the rescaled number of subgraphs may be quite different for different subgraphs.

In Section 9.3.2 we show that by viewing the PAM as a Pólya urn graph, we can associate a sequence of random independent random variables $(\psi_v)_{v \in [t]}$ to the vertices of the PAM, where ψ_v has a Beta distribution with parameters depending on m, δ and v . Once we condition on ψ_1, \dots, ψ_t , the edge statuses of the graph are independent of each other. Furthermore, the degree of a vertex v depends on the index v and ψ_v . The higher ψ_v is, the higher $D_v(t)$ is. Thus, we can interpret ψ_v as a *hidden weight* associated to the vertex v .

Using this representation of the PAM we can view the PAM as a random graph model with two sources of randomness: the randomness of the ψ -variables, and then the randomness of the independent edge statuses determined by the ψ -variables. Therefore, we can define two levels of concentration for the number of ordered subgraphs $N_t(H, \pi)$. Denote by $\mathbb{E}_{\psi_t}[N_t(H, \pi)] := \mathbb{E}[N_t(H, \pi) \mid \psi_1, \dots, \psi_t]$. Furthermore, let $N_{t,\psi}(H, \pi)$ denote the number of ordered subgraphs conditionally on ψ . Then, the ordered subgraph (H, π) can be in the following three classes of subgraphs:

- ▷ *Concentrated:* $N_{t,\psi}(H, \pi)$ is concentrated around its conditional expectation $\mathbb{E}_{\psi_t}[N_t(H, \pi)]$, i.e., as $t \rightarrow \infty$,

$$\frac{N_{t,\psi}(H, \pi)}{\mathbb{E}_{\psi_t}[N_t(H, \pi)]} \xrightarrow{\mathbb{P}} 1, \tag{9.2.7}$$

and as $t \rightarrow \infty$,

$$\frac{N_t(H, \pi)}{\mathbb{E}[N_t(H, \pi)]} \xrightarrow{\mathbb{P}} 1. \tag{9.2.8}$$

- ▷ *Only conditionally concentrated:* condition (9.2.7) holds, and as $t \rightarrow \infty$

$$\frac{N_t(H, \pi)}{\mathbb{E}[N_t(H, \pi)]} \xrightarrow{d} X, \tag{9.2.9}$$

for some random variable X .

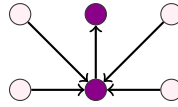


Figure 9.4: The order of magnitude of this subgraph containing two merged copies of the subgraph of Figure 9.3q is $t^{\frac{4}{r-1}}$, so that the condition in Proposition 9.1 is not satisfied for the subgraph in Figure 9.3q.

▷ *Non-concentrated*: condition (9.2.7) does not hold.

For example, it is easy to see that the number of subgraphs as shown in Figure 9.2d satisfies $N(H)/t \xrightarrow{\mathbb{P}} m(m-1)/2$, so that it is a subgraph that belongs to the class of concentrated subgraphs. Below we argue that the triangle belongs to the class of only conditionally concentrated subgraphs. We now give a criterion for the conditional convergence of (9.2.7) in the following proposition:

Proposition 9.1 (Criterion for conditional convergence). *Consider an attainable subgraph (H, π) such that $\mathbb{E}[N_t(H, \pi)] \rightarrow \infty$ as $t \rightarrow \infty$. Denote by $\hat{\mathcal{H}}$ the set of all possible subgraphs composed by two distinct copies of (H, π) with at least one edge in common. Then, as $t \rightarrow \infty$,*

$$\sum_{\hat{H} \in \hat{\mathcal{H}}} \mathbb{E}[N_t(\hat{H})] = o\left(\mathbb{E}[N_t(H, \pi)]^2\right) \implies \frac{N_{t,\psi}(H, \pi)}{\mathbb{E}_{\psi_i}[N_t(H, \pi)]} \xrightarrow{\mathbb{P}} 1. \quad (9.2.10)$$

Proposition 9.1 gives a simple criterion for conditional convergence for a subgraph (H, π) , and it is proved in Section 9.6. The condition in (9.2.10) is simple to evaluate in practice. We denote the subgraphs consisting of two overlapping copies of (H, π) sharing at least one edge by $\hat{H}_1, \dots, \hat{H}_r$. To identify the order of magnitude of $\mathbb{E}[\hat{H}_i]$, we apply Corollary 9.1 to \hat{H}_i or, in other words, we apply Theorem 9.1 to all possible orderings $\hat{\pi}$ of \hat{H}_i . Once we have all orders of magnitude of $(\hat{H}_i, \hat{\pi})$ for all orderings $\hat{\pi}$, and for all \hat{H}_i , it is immediate to see if hypothesis of Proposition 9.1 is satisfied.

There are subgraphs where the condition in Proposition 9.1 does not hold. For example, merging two copies of the subgraph of Figure 9.3q as in Figure 9.4 violates the condition in Proposition 9.1. We show in Section 9.6 that this subgraph is in the class of non-concentrated subgraphs with probability close to one.

9.2.1 Exact constants: triangles

Theorem 9.1 allows to identify the order of magnitude of the expected number of subgraphs in PAM. In particular, for a subgraph H with ordering π , it assures the existence of two constants $0 < C_1 \leq C_2 < \infty$ as in (9.2.2). A more detailed analysis is necessary to prove a stronger result than Theorem 9.1 of the type

$$\lim_{t \rightarrow \infty} \frac{\mathbb{E}[N_t(H, \pi)]}{t^{k+B(H,\pi)} \log^{r-1}(t)} = C,$$

for some constant $0 < C < \infty$. In other words, given an ordered subgraph (H, π) , we want to identify the constant $C > 0$ such that

$$\mathbb{E}[N_t(H, \pi)] = Ct^{k+B(H,\pi)} \log^{r-1}(t)(1 + o(1)). \quad (9.2.11)$$

We prove (9.2.11) for triangles to show the difficulties in the evaluation of the precise constant C for general subgraphs. The following theorem provides the detailed scaling of the expected number of triangles:

Theorem 9.2 (Phase transition for the number of triangles). *Let $m \geq 2$ and $\delta > -m$ be parameters for $(PA_t)_{t \geq 1}$. Denote the number of labeled triangles in PA_t by Δ_t . Then, as $t \rightarrow \infty$,*

(i) *if $\tau > 3$, then*

$$\mathbb{E}[\Delta_t] = \frac{m^2(m-1)(m+\delta)(m+\delta+1)}{\delta^2(2m+\delta)} \log(t)(1+o(1));$$

(ii) *if $\tau = 3$, then*

$$\mathbb{E}[\Delta_t] = \frac{m(m-1)(m+1)}{48} \log^3(t)(1+o(1));$$

(iii) *if $\tau \in (2,3)$, then*

$$\mathbb{E}[\Delta_t] = \frac{m^2(m-1)(m+\delta)(m+\delta+1)}{\delta^2(2m+\delta)} t^{(3-\tau)/(\tau-1)} \log(t)(1+o(1)).$$

Theorem 9.2 in the case $\delta = 0$ coincides with [44, Theorem 14]. For $\delta > 0$ we retrieve the result in [77, Proposition 4.3], noticing that the additive constant β in the attachment probabilities in the Móri model considered in [77] coincides with (9.1) for $\beta = \delta/m$.

The proof of Theorem 9.2 in Section 9.5 shows that to identify the constant in (9.2.11) we need to evaluate the precise expectations involving the attachment probabilities of edges. The equivalent formulation of PAM given in Section 9.3.1 simplifies the calculations, but it is still necessary to evaluate rather complicated expectations involving products of several terms as in (9.3.10). For a more detailed discussion, we refer to Remark 9.1.

The distribution of the number of triangles. Theorem 9.2 shows the behavior of the expected number of triangles. The distribution of the number of triangles across various PAM realizations is another object of interest. We prove the following result for the number of triangles Δ_t :

Corollary 9.2 (Conditional concentration of triangles). *For $\tau \in (2,3)$, the number of triangles Δ_t is conditionally concentrated in the sense of (9.2.7).*

Corollary 9.2 is a direct consequence of Proposition 9.1, and the atlas of the order of magnitudes of all possible realizations of the subgraphs consisting of two triangles sharing one or two edges, presented in Figure 9.5. Figure 9.6 shows a density approximation of the number of triangles obtained by simulations. These figures suggest that the rescaled number of triangles converges to a random limit, since the width of the density plots does not decrease in t . Thus, while the number of triangles

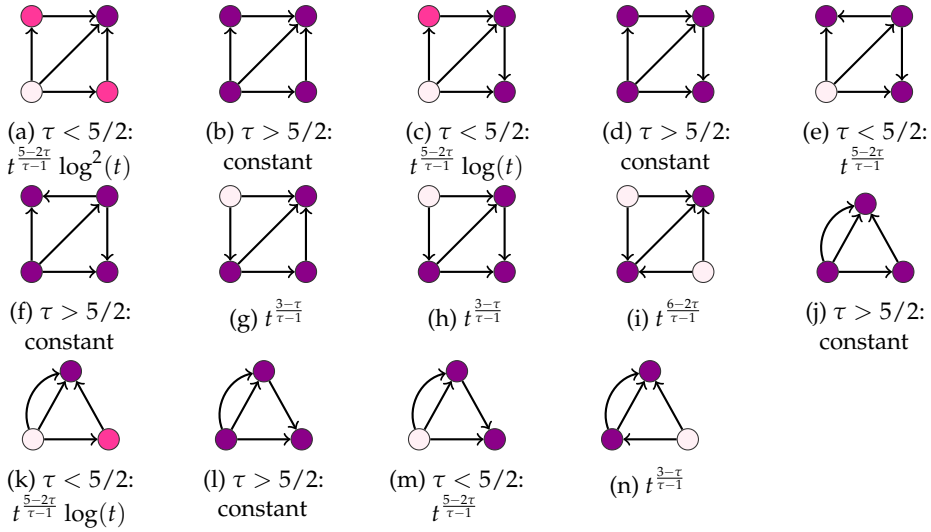


Figure 9.5: Order of magnitude of $N_t(H)$ for all merged triangles on 4 vertices and for $2 < \tau < 3$. The vertex color indicates the optimal vertex degree as in Figure 9.3.

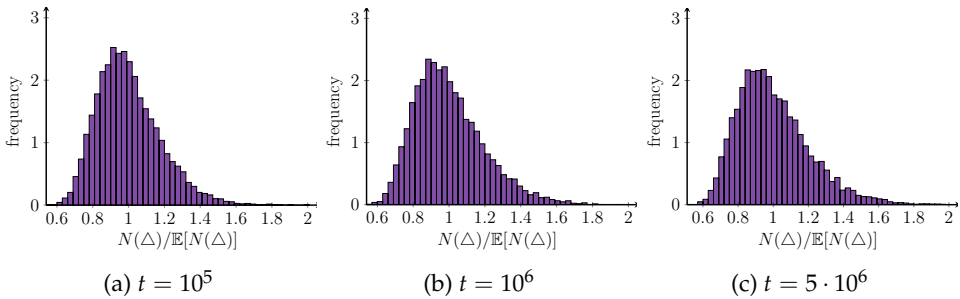


Figure 9.6: Density approximation of the number of triangles in 10^4 realizations of the preferential attachment model with $\tau = 2.5$ and various values of t .

concentrates conditionally, it does not seem to converge to a constant when taking the random ψ -variables into account. This would put the triangle subgraph in the class of only conditionally concentrated subgraphs. Proving this and identifying the limiting random variable of the number of triangles is an interesting open question.

9.3 The probability of a subgraph being present

In this section, we prove the main ingredient for the proof of Theorem 9.1, the probability of a subgraph being present on a given set of vertices. The most difficult part of evaluating the probability of a subgraph H being present in PA_t is that the PAM is constructed recursively. We consider triangles as an example. We write the event of a labeled triangle being present by $\{u \xrightarrow{j^1} v, u \xrightarrow{j^2} w, v \xrightarrow{j^3} w\}$, where $\{u \xrightarrow{j} v\}$

denotes the event that the j -th edge of vertex v is attached to vertex u . Notice that in this way we express precisely which edges we consider in the triangle construction. Then,

$$\begin{aligned} & \mathbb{P}(u \overset{j_1}{\leftarrow} v, u \overset{j_2}{\leftarrow} w, v \overset{j_3}{\leftarrow} w) \\ &= \mathbb{E} \left[\mathbb{P}(u \overset{j_1}{\leftarrow} v, u \overset{j_2}{\leftarrow} w, v \overset{j_3}{\leftarrow} w \mid \text{PA}_{t-1, j_3-1}) \right] \\ &= \mathbb{E} \left[\mathbb{1}\{u \overset{j_1}{\leftarrow} v, u \overset{j_2}{\leftarrow} w\} \frac{D_v(w-1, j_3-1) + \delta}{2m(w-2) + (j_3-1) + (w-1)\delta} \right]. \end{aligned} \tag{9.3.1}$$

In (9.3.1), the indicator function $\mathbb{1}\{u \overset{j_1}{\leftarrow} v, u \overset{j_2}{\leftarrow} w\}$ and $D_v(w-1, j_3-1)$ are not independent, therefore evaluating the expectation on the right-hand side of (9.3.1) is not easy. A possible solution for the evaluation of the expectation in (9.3.1) is to rescale $D_v(w-1, j_3-1)$ with an appropriate constant to obtain a martingale, and then recursively use the conditional expectation. For a detailed explanation of this, we refer to [45, 205] and [106, Section 8.3]. This method is hardly tractable due to the complexity of the constants appearing (see Remark 9.1 for a more detailed explanation).

We use a different approach to evaluate of the expectation in (9.3.1) using the interpretation of the PAM as a Pólya urn graph, focusing mainly on the *the age (the indices) of the vertices*, and not on precise constants. We give a lower and upper bound of the probability of having a finite number of edges present in the graph, as formulated in the following lemma:

Lemma 9.1 (Probability of finite set of labeled edges). *Fix $\ell \in \mathbb{N}$. For vertices $\mathbf{u}_\ell = (u_1, \dots, u_\ell) \in [t]^\ell$ and $\mathbf{v}_\ell = (v_1, \dots, v_\ell) \in [t]^\ell$ and edge labels $\mathbf{j}_\ell = (j_1, \dots, j_\ell) \in [m]^\ell$, consider the corresponding finite set of ℓ distinct labeled edges $M_\ell(\mathbf{u}_\ell, \mathbf{v}_\ell, \mathbf{j}_\ell)$. Assume that the subgraph defined by set $M_\ell(\mathbf{u}_\ell, \mathbf{v}_\ell, \mathbf{j}_\ell)$ is attainable in the sense of Definition 9.2. Define $\chi = (m + \delta) / (2m + \delta)$. Then:*

- (i) *There exist two constants $c_1 = c_1(m, \delta, \ell), c_2 = c_2(m, \delta, \ell) > 0$ such that*

$$c_1 \prod_{l=1}^{\ell} u_l^{\chi-1} v_l^{-\chi} \leq \mathbb{P}(M_\ell(\mathbf{u}_\ell, \mathbf{v}_\ell, \mathbf{j}_\ell) \subseteq E(\text{PA}_t)) \leq c_2 \prod_{l=1}^{\ell} u_l^{\chi-1} v_l^{-\chi}. \tag{9.3.2}$$

- (ii) *Define the set*

$$J(\mathbf{u}_\ell, \mathbf{v}_\ell) = \left\{ \mathbf{j}_\ell \in [m]^\ell : M_\ell(\mathbf{u}_\ell, \mathbf{v}_\ell, \mathbf{j}_\ell) \subseteq E(\text{PA}_t) \right\}. \tag{9.3.3}$$

Then, there exist two constants $\hat{c}_1(m, \delta, \ell), \hat{c}_2(m, \delta, \ell) > 0$ such that

$$\hat{c}_1(m, \delta, \ell) \prod_{l=1}^{\ell} u_l^{\chi-1} v_l^{-\chi} \leq \mathbb{E}[|J(\mathbf{u}_\ell, \mathbf{v}_\ell)|] \leq \hat{c}_2(m, \delta, \ell) \prod_{l=1}^{\ell} u_l^{\chi-1} v_l^{-\chi}. \tag{9.3.4}$$

Formula (9.3.2) in the above lemma bounds the probability that a subgraph is present on vertices \mathbf{u}_ℓ and \mathbf{v}_ℓ such that the j_i -th edge from u_i connects to v_i . Notice

that (9.3.2) is independent of the precise edge labels (j_1, \dots, j_ℓ) . To be able to count all subgraphs, and not only subgraphs where the edge labels have been specified, (9.3.4) bounds the expected number of times a specific subgraph is present on vertices u_ℓ and v_ℓ . This number is given exactly by the elements in set $J(u_\ell, v_\ell)$ as in (9.3.3). The expectation in (9.3.4) may be larger than one, due to the fact that the PAM is a multigraph.

Lemma 9.1 gives a bound on the probability of presence of $\ell \in \mathbb{N}$ distinct edges in the graph as function of the indices $(u_1, v_1), \dots, (u_\ell, v_\ell)$ of the endpoints of the ℓ edges. Due to the properties of PAM, the index of a vertex is an indicator of its degree, due to the old-get-richer effect. Lemma 9.1 is a stronger result than [73, Corollary 2.3], which gives an upper bound of the form in (9.3.2) only for self-avoiding paths.

The proof of Lemma 9.1 is based on the interpretation of the PAM in Definition 9.1 as a urn experiment as proposed in [20]. We now introduce urn schemes and state the preliminary results we need for the proof of Lemma 9.1, which is given in Section 9.3.2.

9.3.1 Pólya urn graph

An urn scheme consists of an urn, with blue balls and red balls. At every time step, we draw a ball from the urn and we replace it by two balls of the same color. We start with $B_0 = b_0$ blue balls and $R_0 = r_0$ red balls. We consider two weight functions

$$W_b(k) = a_b + k, \quad \text{and} \quad W_r(k) = a_r + k. \quad (9.3.5)$$

Conditionally on the number of blue balls B_n and red balls R_n , at time $n + 1$ the probability of drawing a blue ball is equal to

$$\frac{W_b(B_n)}{W_b(B_n) + W_r(R_n)}.$$

The evolution of the number of balls $((B_n, R_n))_{n \in \mathbb{N}}$ obeys [107, Theorem 4.2]

$$\mathbb{P}(B_n = B_0 + k) = \mathbb{E}[\mathbb{P}(\text{Bin}(n, \psi) = k | \psi)], \quad (9.3.6)$$

where ψ has a Beta distribution with parameters $B_0 + a_b$ and $R_0 + a_r$. In other words, the number of blue balls (equivalently, of red balls) is given by a Binomial distribution with a random probability of success ψ (equivalently, $1 - \psi$). Sometimes we call the random variable ψ the *intensity* or *strength* of the blue balls in the urn. We can also see the urn process as two different urns, one containing only blue balls and the other only red balls, and we choose a urn proportionally to the number of balls in the urns. In this case, the result is the same, but we can say that ψ is the strength of the blue balls urn and $1 - \psi$ is the strength of the red balls urn.

The sequential model PA_t can be interpreted as experiment with t urns, where the number of balls in each urn represent the degree of a vertex in the graph. First, we introduce a random graph model:

Definition 9.3 (Pólya urn graph). Fix $m \geq 1$ and $\delta > -m$. Let $t \in \mathbb{N}$ be the size of the graph. Let $\psi_1 = 1$, and consider ψ_2, \dots, ψ_t independent random variables, where

$$\psi_k \stackrel{d}{=} \text{Beta}(m + \delta, m(2k - 3) + (k - 1)\delta). \quad (9.3.7)$$

Define

$$\varphi_j = \psi_j \prod_{i=j+1}^t (1 - \psi_i), \quad S_k = \sum_{j=1}^k \varphi_j, \quad I_k = [S_{k-1}, S_k). \quad (9.3.8)$$

Conditioning on ψ_1, \dots, ψ_t , let $\{U_{k,j}\}_{k=2, \dots, t}^{j=1, \dots, m}$ be independent random variables, with $U_{k,j}$ uniformly distributed on $[0, S_{k-1}]$. Then, the corresponding Pólya urn graph PU_t is the graph of size t where, for $u < v$, the number of edges between u and v is equal to the number of variables $U_{v,j}$ in I_u , for $j = 1, \dots, m$ (multiple edges are allowed).

The two sequences of graphs $(\text{PA}_t)_{t \in \mathbb{N}}$ and $(\text{PU}_t)_{t \in \mathbb{N}}$ have the same distribution [20, Theorem 2.1], [107, Chapter 4]. The Beta distributions in Definition 9.3 come from the Pólya urn interpretation of the sequential model, using urns with affine weight functions.

The formulation in Definition 9.3 in terms of urn experiments allows us to investigate the presence of subgraphs in an easier way than with the formulation given in Definition 9.1 since the dependent random variables in (9.3.1), are replaced by the product of independent random variables. We now state two lemmas that are the main ingredients for proving Lemma 9.1:

Lemma 9.2 (Attachment probabilities). *Consider PU_t as in Definition 9.3. Then,*

(i) for $k \in [t]$,

$$S_k = \prod_{h=k+1}^t (1 - \psi_h); \quad (9.3.9)$$

(ii) conditioning on ψ_1, \dots, ψ_t , the probability that the j -th edge of k is attached to v is equal to

$$\mathbb{P}\left(U_{k,j} \in I_v \mid \psi_1, \dots, \psi_t\right) = \psi_v \frac{S_v}{S_{h-1}} = \psi_v \prod_{h=v+1}^{k-1} (1 - \psi_h). \quad (9.3.10)$$

The proof of Lemma 9.2 follows from Definition 9.3, and the fact that $(S_k)_{k \in [t]}$ as in (9.3.8) can be written as in (9.3.9) (see the proof of [20, Theorem 2.1]).

Before proving Lemma 9.1, we state a second result on the concentration of the positions $\{S_k\}_{k \in [t]}$ in the urn graph $(\text{PU}_t)_{t \in \mathbb{N}}$. In particular, it shows that these positions concentrate around deterministic values:

Lemma 9.3 (Position concentration in PU_t). *Consider a Pólya urn graph as in Definition 9.3. Let $\chi = (m + \delta) / (2m + \delta)$. Then, for every $\omega, \varepsilon > 0$ there exists $N_0 = N_0(\omega, \varepsilon) \in \mathbb{N}$ such that, for every $t \geq N_0$,*

$$\mathbb{P}\left(\bigcap_{i=N_0}^t \left\{ \left| S_i - \left(\frac{i}{t}\right)^\chi \right| \leq \omega \left(\frac{i}{t}\right)^\chi \right\}\right) \geq 1 - \varepsilon \quad (9.3.11)$$

and, for t large enough,

$$\mathbb{P}\left(\max_{i \in [t]} \left| S_i - \left(\frac{i}{t}\right)^\chi \right| \geq \omega\right) \leq \varepsilon \quad (9.3.12)$$

As a consequence, as $t \rightarrow \infty$,

$$\max_{i \in [t]} \left| S_i - \left(\frac{i}{t} \right)^\chi \right| \xrightarrow{\mathbb{P}} 0. \quad (9.3.13)$$

The proof of Lemma 9.3 is given in [20, Lemma 3.1].

9.3.2 Proof of Lemma 9.1

We now prove Lemma 9.1, starting with the proof of (9.3.2). Fix u_ℓ, v_ℓ, j_ℓ . In the proof, we denote $M_\ell(u_\ell, v_\ell, j_\ell)$ simply by M_ℓ to keep notation light. We use the fact that the Pólya urn graph PU_t and PA_t have the same distribution and evaluate $\mathbb{P}(M_\ell \subseteq E(\text{PU}_t))$. We consider ℓ distinct labeled edges, so we can use (9.3.10) to write

$$\mathbb{P}(M_\ell \subseteq E(\text{PU}_t) \mid \psi_1, \dots, \psi_t) = \prod_{l=1}^{\ell} \psi_{u_l} \frac{S_{u_l}}{S_{v_l-1}}. \quad (9.3.14)$$

Now fix $\varepsilon > 0$. Define $\mathcal{E}_\varepsilon := \{\max_{i \in [t]} |S_i - \left(\frac{i}{t}\right)^\chi| \leq \varepsilon\}$. By (9.3.13), and the fact that the product of the random variables in (9.3.14) is bounded by 1,

$$\mathbb{E} \left[\prod_{l=1}^{\ell} \psi_{u_l} \frac{S_{u_l}}{S_{v_l-1}} \right] = \mathbb{E} \left[\mathbb{1}_{\mathcal{E}_\varepsilon} \prod_{l=1}^{\ell} \psi_{u_l} \frac{S_{u_l}}{S_{v_l-1}} \right] + o(1). \quad (9.3.15)$$

On the event \mathcal{E}_ε , we have, for every $l \in [\ell]$,

$$(1 - \varepsilon) \left(\frac{u_l}{v_l} \right)^\chi \leq \frac{S_{u_l}}{S_{v_l-1}} \leq (1 + \varepsilon) \left(\frac{u_l}{v_l} \right)^\chi, \quad (9.3.16)$$

where in (9.3.16) we have replaced $v_l - 1$ with v_l with a negligible error. Notice that since v_l is always the source of the edge, this implies $v_l \geq 2$, therefore this is allowed. Using (9.3.16) in (9.3.15) we obtain

$$\begin{aligned} (1 - \varepsilon)^\ell \prod_{l=1}^{\ell} \left(\frac{u_l}{v_l} \right)^\chi \mathbb{E} \left[\mathbb{1}_{\mathcal{E}_\varepsilon} \prod_{l=1}^{\ell} \psi_{u_l} \right] &\leq \mathbb{P}(M_\ell \subseteq E(\text{PU}_t)) \\ &\leq (1 + \varepsilon)^\ell \prod_{l=1}^{\ell} \left(\frac{u_l}{v_l} \right)^\chi \mathbb{E} \left[\mathbb{1}_{\mathcal{E}_\varepsilon} \prod_{l=1}^{\ell} \psi_{u_l} \right]. \end{aligned} \quad (9.3.17)$$

Even though ψ_1, \dots, ψ_t depend on \mathcal{E}_ε , it is easy to show that we can ignore $\mathbb{1}_{\mathcal{E}_\varepsilon}$ in (9.3.17) and obtain a similar bound. Therefore

$$(1 - \varepsilon)^\ell \prod_{l=1}^{\ell} \left(\frac{u_l}{v_l} \right)^\chi \mathbb{E} \left[\prod_{l=1}^{\ell} \psi_{u_l} \right] \leq \mathbb{P}(M_\ell \subseteq E(\text{PU}_t)) \leq (1 + \varepsilon)^\ell \prod_{l=1}^{\ell} \left(\frac{u_l}{v_l} \right)^\chi \mathbb{E} \left[\prod_{l=1}^{\ell} \psi_{u_l} \right] \quad (9.3.18)$$

What remains is to evaluate the expectation in (9.3.18). We assumed to have ℓ distinct edges, that does not imply that the vertices $u_1, v_1, \dots, u_\ell, v_\ell$ are distinct. The expectation in (9.3.18) depends only on the receiving vertices of the ℓ edges, namely u_1, \dots, u_ℓ .

Let $\bar{u}_1, \dots, \bar{u}_k$ denote the $k \leq \ell$ distinct elements that appear among u_1, \dots, u_ℓ . For $h \in [k]$, the vertex \bar{u}_h appears in the product inside the expectation in (9.3.18) with multiplicity $d_h^{(\text{in})}$, which is the degree of vertex \bar{u}_k in the subgraph defined by M_ℓ . As a consequence, we can write

$$\mathbb{E} \left[\prod_{l=1}^{\ell} \psi_{u_l} \right] = \mathbb{E} \left[\prod_{h=1}^k \psi_{\bar{u}_h}^{d_h^{(\text{in})}} \right] = \prod_{h=1}^k \mathbb{E} \left[\psi_{\bar{u}_h}^{d_h^{(\text{in})}} \right], \quad (9.3.19)$$

where in (9.3.19) we have used the fact that ψ_1, \dots, ψ_t are all independent. Notice that $\mathbb{E}[\psi_1^d] = 1$ for all $d \geq 0$, since $\psi_1 \equiv 1$. Therefore, if $\bar{u}_h = 1$ for some $h \in [k]$, $\mathbb{E}[\psi_{\bar{u}_h}^d] = 1$ and the terms depending on the first vertex contribute to the expectation in (9.3.19) by a constant.

For the terms where $\bar{u}_h \geq 2$, recall that, if $X(\alpha, \beta)$ is a Beta random variable, then, for any integer $d \in \mathbb{N}$,

$$\mathbb{E}[X(\alpha, \beta)^d] = \frac{\alpha(\alpha + 1) \cdots (\alpha + d - 1)}{(\alpha + \beta)(\alpha + \beta + 1) \cdots (\alpha + \beta + d - 1)}.$$

Since $\psi_{\bar{u}_h}$ is Beta distributed with parameters $m + \delta$ and $2(\bar{u}_h - 3) + (\bar{u}_h - 1)\delta$,

$$\begin{aligned} \mathbb{E} \left[\psi_{\bar{u}_h}^{d_h^{(\text{in})}} \right] &= \frac{(m + \delta) \cdots (m + \delta + d_h^{(\text{in})} - 1)}{[m(2\bar{u}_h - 2) + \bar{u}_h \delta] \cdots [m(2\bar{u}_h - 2) + \bar{u}_h \delta + d_h^{(\text{in})} - 1]} \\ &= \bar{u}_h^{-d_h^{(\text{in})}} \frac{(m + \delta) \cdots (m + \delta + d_h^{(\text{in})} - 1)}{[2m + \delta - (2m)/\bar{u}_h] \cdots [2m + \delta + (d_h^{(\text{in})} - 1 - 2m)/\bar{u}_h]}. \end{aligned} \quad (9.3.20)$$

Notice that if $\bar{u}_h \geq 2$, uniformly in t and the precise choice of the ℓ edges,

$$(m + \delta)^{-\ell} \leq \left([2m + \delta - (2m)/\bar{u}] \cdots [2m + \delta + (d_h^{(\text{in})} - 1 - 2m)/\bar{u}] \right)^{-1} \leq (2m + \delta + \ell)^{-\ell}.$$

As a consequence, we can find two constants $c_1(m, \delta, \ell), c_2(m, \delta, \ell)$ such that

$$c_1(m, \delta, \ell) \prod_{h=1}^k \bar{u}_h^{-d_h^{(\text{in})}} \leq \prod_{h=1}^k \mathbb{E} \left[\psi_{\bar{u}_h}^{d_h^{(\text{in})}} \right] \leq c_2(m, \delta, \ell) \prod_{h=1}^k \bar{u}_h^{-d_h^{(\text{in})}}. \quad (9.3.21)$$

We now use (9.3.21) in (9.3.18) to obtain

$$\begin{aligned} c_1(m, \delta, \ell)(1 - \varepsilon)^\ell \prod_{l=1}^{\ell} \left(\frac{u_l}{v_l} \right)^\chi \prod_{h=1}^k \bar{u}_h^{-d_h^{(\text{in})}} &\leq \mathbb{P}(M_\ell \subseteq E(\text{PU}_t)) \\ &\leq c_2(m, \delta, \ell)(1 + \varepsilon)^\ell \prod_{l=1}^{\ell} \left(\frac{u_l}{v_l} \right)^\chi \prod_{h=1}^k \bar{u}_h^{-d_h^{(\text{in})}}. \end{aligned} \quad (9.3.22)$$

In (9.3.22) we can just rename the constants $c_1(m, \delta, \ell) = c_1(m, \delta, \ell)(1 - \varepsilon)^\ell$ and $c_2(m, \delta, \ell) = c_2(m, \delta, \ell)(1 + \varepsilon)^\ell$. Since $d_h^{(\text{in})}$ is the multiplicity of vertex \bar{u}_h as receiving vertex, we can write

$$\prod_{h=1}^k \bar{u}_h^{-d_h^{(\text{in})}} = \prod_{l=1}^{\ell} u_l^{-1}.$$

Combining this with (9.3.22) completes the proof of (9.3.2).

The proof of (9.3.4) follows immediately from (9.3.2) and the definition of the set $J(\mathbf{u}_\ell, \mathbf{v}_\ell)$ in (9.3.3). In fact, we can write

$$\mathbb{E}[|J(\mathbf{u}_\ell, \mathbf{v}_\ell)|] = \sum_{\mathbf{j}_\ell \in [m]^\ell} \mathbb{P}(M_\ell(\mathbf{u}_\ell, \mathbf{v}_\ell, \mathbf{j}_\ell) \subseteq E(\text{PA}_t)).$$

Recall that $\mathbb{P}(M_\ell(\mathbf{u}_\ell, \mathbf{v}_\ell, \mathbf{j}_\ell) \subseteq E(\text{PA}_t))$ is independent of the labels \mathbf{j}_ℓ . For a fixed set of source and target vertices \mathbf{u}_ℓ and \mathbf{v}_ℓ , there is only a finite combination of labels \mathbf{j}_ℓ such that the subgraph defined by $M_\ell(\mathbf{u}_\ell, \mathbf{v}_\ell, \mathbf{j}_\ell)$ is attainable in the sense of Definition 9.2. In fact, the number of such labels \mathbf{j}_ℓ is larger than one (since the corresponding subgraph is attainable), and less than m^ℓ (the total number of elements of $[m]^\ell$). As a consequence, taking $\hat{c}_1 = c_1$ and $\hat{c}_2 = c_2 m^\ell$ proves (9.3.4). \square

9.4 The expected number of subgraphs

To prove Theorem 9.1, we write the expected number of subgraphs as multiple integrals. W.l.o.g. we assume throughout this section that π is the identity permutation, so that the vertices of H are labeled as $1, \dots, k$, and therefore drop the dependence of the quantities on π . We first prove a lemma that states that two integrals that will be important in proving Theorem 9.1 are finite:

Lemma 9.4. *Let H be a subgraph such that the optimum of (9.2.1) is attained by s_1, \dots, s_r . Then,*

$$A_1(H) := \int_1^\infty u_1^{\beta(1)} \int_{u_1}^\infty u_2^{\beta(2)} \dots \int_{u_{s_1-1}}^\infty u_{s_1}^{\beta(s_1)} du_{s_1} \dots du_1 < \infty, \quad (9.4.1)$$

$$A_2(H) := \int_0^1 u_k^{\beta(k)} \int_0^{u_k} u_{k-1}^{\beta(k-1)} \dots \int_0^{u_{s_r+1}} u_{s_r+1}^{\beta(s_r+1)} du_{s_r+1} \dots du_k < \infty. \quad (9.4.2)$$

Proof. The first integral is finite as long as

$$z + \sum_{i=s_1-z}^{s_1} \beta(i) < 0 \quad (9.4.3)$$

for all $z \in [s_1]$. Suppose that (9.4.3) does not hold for some $z^* \in [s_1]$. Then, the difference between the contribution to (9.2.1) for $\tilde{s} = s_1 - z^*$ and s_1 is

$$-(s_1 - z^*) + \sum_{i=s_1-z^*}^k \beta(i) + s_1 - \sum_{i=s_1}^k \beta(i) = z^* + \sum_{i=s_1-z^*}^{s_1} \beta(i) \geq 0, \quad (9.4.4)$$

which would imply that $s_1 - z^*$ is also an optimizer of (9.2.1), which is in contradiction with s_1 being the smallest optimum. Thus, (9.4.3) holds for all $r \in [s]$ and $A_1(H) < \infty$.

The second integral is finite as long as

$$z - s_r + \sum_{i=s_r+1}^z \beta(i) > 0 \quad (9.4.5)$$

for all $z \in \{s_r + 1, \dots, k\}$. Suppose that this does not hold for some $z^* \in \{s_r + 1, \dots, k\}$. Set $\tilde{s} = z^* > s_r$. Then, the difference between the contribution to (9.2.1) for $\tilde{s} = z^*$ and s_r is

$$-z^* + s_r - \sum_{i=s_r+1}^{z^*} \beta(i) \geq 0, \quad (9.4.6)$$

which is a contradiction with s_r being the largest optimizer. Therefore, $A_2(H) < \infty$. \square

We now use this lemma to prove Theorem 9.1:

Proof of Theorem 9.1. Again, we assume that π is the identity mapping, so that we may drop all dependencies on π . Suppose the optimal solution to (9.2.1) is attained by s_1, s_2, \dots, s_r for some $r \geq 1$. Let the ℓ edges of H be denoted by (u_l, v_l) for $l \in [\ell]$. Let $N_t(H, i_1, \dots, i_k)$ denote the number of times subgraph H is present on vertices i_1, \dots, i_k . We then use Lemma 9.1, which proves that, for some $0 < C < \infty$,

$$\begin{aligned} \mathbb{E}[N_t(H)] &= \sum_{i_1 < \dots < i_k \in [t]} \mathbb{E}[N_t(H, i_1, \dots, i_k)] \\ &\leq C \sum_{i_1 < \dots < i_k \in [t]} \prod_{l=1}^{\ell} i_{u_l}^{\chi-1} i_{v_l}^{-\chi} = C \sum_{i_1 < \dots < i_k \in [t]} \prod_{q=1}^k i_q^{\beta(q)}. \end{aligned} \quad (9.4.7)$$

We then bound the sums by integrals as

$$\begin{aligned} \mathbb{E}[N_t(H)] &\leq \tilde{C} \int_1^t u_1^{\beta(1)} \dots \int_{u_{k-1}}^t u_k^{\beta(k)} du_k \dots du_1 \\ &\leq \tilde{C} \int_1^\infty u_1^{\beta(1)} \dots \int_{u_{s-1}}^\infty u_{s_1}^{\beta(s_1)} du_{s_1} \dots du_1 \\ &\quad \times \int_1^t u_{s_1+1}^{\beta(s_1+1)} \int_{u_{s_1+1}}^\infty u_{s_1+2}^{\beta(s_1+2)} \dots \int_{u_{s_2-1}}^t u_{s_2}^{\beta(s_2)} du_{s_2} \dots du_{s_1+1} \\ &\quad \times \int_1^t u_{s_2+1}^{\beta(s_2+1)} \int_{u_{s_2+1}}^\infty u_{s_2+2}^{\beta(s_2+2)} \dots \int_{u_{s_3-1}}^t u_{s_3}^{\beta(s_3)} du_{s_3} \dots du_{s_2+1} \times \dots \\ &\quad \times \int_1^t u_{s_{r-1}+1}^{\beta(s_{r-1}+1)} \int_{u_{s_{r-1}+1}}^\infty u_{s_{r-1}+2}^{\beta(s_{r-1}+2)} \dots \int_{u_{s_r-1}}^t u_{s_r}^{\beta(s_r)} du_{s_r} \dots du_{s_{r-1}+1} \\ &\quad \times \int_0^t u_{s_r+1}^{\beta(s_r+1)} \int_{s_r+1}^t u_{s_r+2}^{\beta(s_r+2)} \dots \int_{u_{k-1}}^t u_k^{\beta(k)} du_k \dots du_{s_r+1}, \end{aligned} \quad (9.4.8)$$

for some $0 < \tilde{C} < \infty$. The first set of integrals is finite by Lemma 9.4 and independent of t . For the last set of integrals, we obtain

$$\begin{aligned} &\int_0^t u_{s_r+1}^{\beta(s_r+1)} \int_{u_{s_r+1}}^t u_{s_r+2}^{\beta(s_r+2)} \dots \int_{u_{k-1}}^t u_k^{\beta(k)} du_k \dots du_{s_r+1} \\ &= t^{k-s_r+\sum_{i=s_r+1}^k \beta(i)} \int_0^1 w_{s_r+1}^{\beta(s_r+1)} \int_{w_{s_r+1}}^1 w_{s_r+2}^{\beta(s_r+2)} \dots \int_{w_{k-1}}^1 w_k^{\beta(k)} dw_k \dots dw_{s_r+1} \\ &= K t^{k+B(H)}, \end{aligned} \quad (9.4.9)$$

for some $0 < K < \infty$, where we have used the change of variables $w = u/t$ and Lemma 9.4. For $r = 1$, this finishes the proof, because then the middle integrals in (9.4.8) are empty. We now investigate the behavior of the middle sets of integrals for $r > 1$. Because the optimum to (9.2.1) is attained for s_1 as well as s_2 ,

$$-s_1 + \sum_{i=s_1+1}^k \beta(i) + s_2 - \sum_{i=s_2+1}^k \beta(i) = s_2 - s_1 + \sum_{i=s_1+1}^{s_2} \beta(i) = 0. \quad (9.4.10)$$

Therefore, when $s_2 = s_1 + 1$, the second set of integrals in (9.4.8) equals

$$\int_1^t u_{s_1}^{-1} du_{s_1} = \log(t). \quad (9.4.11)$$

Now suppose that $s_1 < s_2 + 1$. Then, any $\tilde{s} \in [s_1 + 1, s_2 - 1]$ is a non-optimal solution to (9.2.1), and therefore

$$-s_2 + \sum_{i=s_2+1}^k \beta(i) + \tilde{s} - \sum_{i=\tilde{s}+1}^k \beta(i) = \tilde{s} - s_2 - \sum_{i=\tilde{s}+1}^{s_2} \beta(i) > 0, \quad (9.4.12)$$

or

$$\sum_{i=\tilde{s}+1}^{s_2} \beta(i) < s_2 - \tilde{s}. \quad (9.4.13)$$

This implies that

$$\begin{aligned} & \int_1^t u_{s_1+1}^{\beta(s_1+1)} \int_{u_{s_1+1}}^\infty u_{s_1+2}^{\beta(s_1+2)} \cdots \int_{u_{s_2-1}}^t u_{s_2}^{\beta(s_2)} du_{s_2} \cdots du_{s_1+1} \\ &= \tilde{K} \int_1^t u_{s_1+1}^{\sum_{i=s_1+1}^{s_2} \beta(i) + s_2 - s_1 - 1} du_{s_1+1} \\ &= \tilde{K} \int_1^t u_{s_1+1}^{-1} du_{s_1+1} = \tilde{K} \log(t), \end{aligned} \quad (9.4.14)$$

for some $0 < C < \infty$. A similar reasoning holds for the other integrals, so that combining (9.4.8), (9.4.9) and (9.4.14) yields

$$\lim_{t \rightarrow \infty} \frac{\mathbb{E}[N_t(H)]}{t^{k+B(H)} \log^{r-1}(t)} \leq C_2, \quad (9.4.15)$$

for some $0 < C_2 < \infty$.

We now proceed to prove a lower bound on the expected number of subgraphs. Again, by Lemma 9.1 and lower bounding the sums by integrals as in (9.4.7), we obtain that, for some $0 < C < \infty$

$$\mathbb{E}[N_t(H)] \geq C \int_1^t u_1^{\beta(1)} \cdots \int_{u_{k-1}}^t u_k^{\beta(k)} du_k \cdots du_1. \quad (9.4.16)$$

Fix $\varepsilon > 0$. We investigate the contribution where vertices $1, \dots, s_1$ have index in $[1, 1/\varepsilon]$, vertices $s_1 + 1, \dots, s_2$ have index in $[1/\varepsilon, \varepsilon t^{1/r}]$, vertices $s_2 + 1, \dots, s_3$ have

index in $[t^{1/r}, \varepsilon t^{2/r}]$ and so on, and vertices $s_r + 1, \dots, s_k$ have index in $[\varepsilon t, t]$. Thus, we bound

$$\begin{aligned}
 \mathbb{E} [N_t(H)] &\geq C \int_1^{1/\varepsilon} u_1^{\beta(1)} \int_{u_1}^{1/\varepsilon} u_2^{\beta(2)} \dots \int_{u_{s_1-1}}^{1/\varepsilon} u_{s_1}^{\beta(s)} du_{s_1} \dots du_1 \\
 &\quad \times \int_{1/\varepsilon}^{\varepsilon t^{1/r}} u_{s_1+1}^{\beta(s_1+1)} \int_{u_{s_1+1}}^{u_{s_1+1}/\varepsilon} u_{s_1+2}^{\beta(s_1+2)} \dots \int_{u_{s_2-1}}^{u_{s_2-1}/\varepsilon} u_{s_2}^{\beta(s_2)} du_{s_2} \dots du_{s_1+1} \\
 &\quad \times \int_{t^{1/r}}^{\varepsilon t^{2/r}} u_{s_2+1}^{\beta(s_2+1)} \int_{u_{s_2+1}}^{u_{s_2+1}/\varepsilon} u_{s_2+2}^{\beta(s_2+2)} \dots \int_{u_{s_3-1}}^{u_{s_3-1}/\varepsilon} u_{s_3}^{\beta(s_3)} du_{s_3} \dots du_{s_2+1} \times \dots \\
 &\quad \times \int_{t^{(r-2)/r}}^{\varepsilon t^{(r-1)/r}} u_{s_{r-1}+1}^{\beta(s_{r-1}+1)} \int_{u_{s_{r-1}+1}}^{u_{s_{r-1}+1}/\varepsilon} u_{s_{r-1}+2}^{\beta(s_{r-1}+2)} \dots \int_{u_{s_r-1}}^{u_{s_r-1}/\varepsilon} u_{s_r}^{\beta(s_r)} du_{s_r} \dots du_{s_{r-1}+1} \\
 &\quad \times \int_{\varepsilon t}^t u_{s_r+1}^{\beta(s_r+1)} \int_{u_{s_r+1}}^t u_{s_r+2}^{\beta(s_r+2)} \dots \int_{u_{k-1}}^t u_k^{\beta(k)} du_k \dots du_{s_r+1} \tag{9.4.17}
 \end{aligned}$$

The first set of integrals equals $A_1(H)$ plus terms that vanishes as ε becomes small by Lemma 9.4. For the last set of integrals, we use the change of variables $w = u/t$ to obtain

$$\begin{aligned}
 &\int_{\varepsilon t}^t u_{s_r+1}^{\beta(s_r+1)} \int_{u_{s_r+1}}^t u_{s_r+2}^{\beta(s_r+2)} \dots \int_{u_{k-1}}^t u_k^{\beta(k)} du_k \dots du_{s_r+1} \\
 &= t^{k-s_r+\sum_{i=s_r+1}^k \beta(i)} \int_{\varepsilon}^1 w_{s_r+1}^{\beta(s_r+1)} \int_{w_{s_r+1}}^1 w_{s_r+2}^{\beta(s_r+2)} \dots \int_{w_{k-1}}^1 w_k^{\beta(k)} dw_k \dots dw_{s_r+1} \\
 &= t^{k+B(H)} (A_2(H) - h_1(\varepsilon)), \tag{9.4.18}
 \end{aligned}$$

for some function $h_1(\varepsilon)$. By Lemma 9.4 $h_1(\varepsilon)$ satisfies $\lim_{\varepsilon \rightarrow 0} h_1(\varepsilon) = 0$. Again, if $r = 1$, the middle sets of integrals in (9.4.17) are empty, so we are done.

We now investigate the second set of integrals in (9.4.17) for $r > 1$. Using the substitution $w_{s_1+1} = u_{s_1+1}$ and $w_i = u_i/u_{i-1}$ for $i > s_1 + 1$, we obtain

$$\begin{aligned}
 &\int_{1/\varepsilon}^{\varepsilon t^{1/r}} u_{s_1+1}^{\beta(s_1+1)} \int_{u_{s_1+1}}^{u_{s_1+1}/\varepsilon} u_{s_1+2}^{\beta(s_1+2)} \dots \int_{u_{s_2-1}}^{u_{s_2-1}/\varepsilon} u_{s_2}^{\beta(s_2)} du_{s_2} \dots du_{s_1+1} \\
 &= \int_{1/\varepsilon}^{\varepsilon t^{1/r}} w_{s_1+1}^{s_2-s_1-1+\sum_{i=s_1+1}^{s_2} \beta(i)} dw_{s_1+1} \int_1^{1/\varepsilon} w_{s_1+2}^{s_2-s_1-2+\sum_{i=s_1+2}^{s_2} \beta(i)} dw_{s_2+1} \dots \int_1^{1/\varepsilon} w_{s_2}^{\beta(s_2)} dw_{s_2}. \tag{9.4.19}
 \end{aligned}$$

The first integral equals by (9.4.10)

$$\int_{1/\varepsilon}^{\varepsilon t^{1/r}} w_{s_1+1}^{-1} dw_{s_1+1} = \frac{1}{r} \log(t) + \log(\varepsilon^2). \tag{9.4.20}$$

The integrand in all other integrals in (9.4.19) equals $w_i^{\gamma_i}$ for some $\gamma_i < -1$ by (9.4.13). Therefore, these integrals equal a constant plus a function of ε that vanishes as ε

becomes small so that

$$\begin{aligned} & \int_{1/\varepsilon}^{\varepsilon t^{1/r}} u_{s_1+1}^{\beta(s_1+1)} \int_{u_{s_1+1}}^{u_{s_1+1}/\varepsilon} u_{s_1+2}^{\beta(s_1+2)} \cdots \int_{u_{s_2-1}}^{u_{s_2-1}/\varepsilon} u_{s_2}^{\beta(s_2)} du_{s_2} \cdots du_{s_1+1} \\ &= \left(\frac{1}{r} \log(t) + \log(\varepsilon^2) \right) (K + h_2(\varepsilon)), \end{aligned} \quad (9.4.21)$$

for some $0 < K < \infty$ and some $h_2(\varepsilon)$ such that $\lim_{\varepsilon \rightarrow 0} h_2(\varepsilon) = 0$. The other integrals in (9.4.17) can be estimated similarly.

Combining (9.4.17), (9.4.18) and (9.4.21) we obtain

$$\lim_{t \rightarrow \infty} \frac{\mathbb{E}[N_t(H)]}{t^{k+B(H)} \log^{r-1}(t)} \geq C_1 + h(\varepsilon), \quad (9.4.22)$$

for some constant $0 < C_1 < \infty$ and some function $h(\varepsilon)$ such that $\lim_{\varepsilon \rightarrow 0} h(\varepsilon) = 0$. Taking the limit for $\varepsilon \rightarrow 0$ then proves the theorem. \square

9.5 The expected number of triangles

Fix $m \geq 2$ and $\delta > -m$. The first step of the proof consists of showing that

$$\begin{aligned} \mathbb{E}[\Delta_t] &= \frac{\tau - 2}{\tau - 1} \frac{m^2(m-1)(m+\delta)(m+\delta+1)}{(2m+\delta)^2} \\ &\quad \times \sum_{u=1}^{t-2} [(u - (2m)/(2m+\delta))(u - (2m-1)/(2m+\delta))]^{-1} \\ &\quad \times \frac{\Gamma(u+2 - (2m)/(2m+\delta))}{\Gamma(u+2 - (3m+\delta)/(2m+\delta))} \frac{\Gamma(u+2 - (2m-1)/(2m+\delta))}{\Gamma(u+2 - (3m+\delta-1)/(2m+\delta))} \\ &\quad \times \sum_{v=u+1}^{t-1} (v - (3m+\delta-1)/(2m+\delta))^{-1} \\ &\quad \times \sum_{w=v+1}^t \frac{\Gamma(w - (3m+\delta)/(2m+\delta))}{\Gamma(w - (2m)/(2m+\delta))} \frac{\Gamma(w - (3m+\delta-1)/(2m+\delta))}{\Gamma(w - (2m-1)/(2m+\delta))}. \end{aligned} \quad (9.5.1)$$

We can write

$$\Delta_t := \sum_{u=1}^{t-2} \sum_{v=u+1}^{t-1} \sum_{w=v+1}^t \sum_{j_1 \in [m]} \sum_{j_2, j_3 \in [m]} \mathbb{1}\{u \stackrel{j_1}{\leftarrow} v, u \stackrel{j_2}{\leftarrow} w, v \stackrel{j_3}{\leftarrow} w\}. \quad (9.5.2)$$

Since there are $m^2(m-1)$ possible choices for the edges j_1, j_2, j_3 ,

$$\mathbb{E}[\Delta_t] = m^2(m-1) \sum_{u=1}^{t-2} \sum_{v=u+1}^{t-1} \sum_{w=v+1}^t \mathbb{E} \left[\psi_u \frac{S_u}{S_{v-1}} \psi_u \frac{S_u}{S_{w-1}} \psi_v \frac{S_v}{S_{w-1}} \right]. \quad (9.5.3)$$

Recalling (9.3.10), we can write every term in the sum in (9.5.3) as

$$\mathbb{E} \left[\left(\psi_u \prod_{h=u+1}^{v-1} (1 - \psi_h) \right) \left(\psi_u \prod_{k=u+1}^{w-1} (1 - \psi_k) \right) \left(\psi_v \prod_{l=v+1}^{w-1} (1 - \psi_l) \right) \right]. \quad (9.5.4)$$

Since the random variables ψ_1, \dots, ψ_t are independent, we can factorize the expectation to obtain

$$\mathbb{E}[\psi_u^2] \mathbb{E}[\psi_v(1 - \psi_v)] \prod_{k=u+1, k \neq v}^{w-1} \mathbb{E}[(1 - \psi_k)^2] = \mathbb{E}[\psi_u^2] \frac{\mathbb{E}[\psi_v(1 - \psi_v)]}{\mathbb{E}[(1 - \psi_v)^2]} \prod_{k=u+1}^{w-1} \mathbb{E}[(1 - \psi_k)^2]. \tag{9.5.5}$$

Recall that, for a Beta random variable $X(\alpha, \beta)$, we have

$$\begin{aligned} \mathbb{E}[X] &= \frac{\alpha}{\alpha + \beta}, & \mathbb{E}[X(1 - X)] &= \frac{\alpha\beta}{(\alpha + \beta)(\alpha + \beta + 1)}, \\ \mathbb{E}[X^2] &= \frac{\alpha(\alpha + 1)}{(\alpha + \beta)(\alpha + \beta + 1)}, \end{aligned} \tag{9.5.6}$$

and $1 - X(\alpha, \beta)$ is distributed as $X(\beta, \alpha)$. Using (9.5.6), we can rewrite (9.5.5) in terms of the parameters of ψ_1, \dots, ψ_t . Since ψ_k has parameters $\alpha = m + \delta$ and $\beta = \beta_k = m(2k - 3) + (k - 1)\delta$, the first term in (9.5.5) can be written as

$$\begin{aligned} \mathbb{E}[\psi_u^2] &\frac{(m + \delta)(m + \delta + 1)}{(m(2u - 2) + u\delta)(m(2u - 2) + u\delta + 1)} \\ &= \frac{(m + \delta)(m + \delta + 1)}{(2m + \delta)^2} \left[(u - 2m/2m + \delta)(u - (2m - 1)/(2m + \delta)) \right]^{-1}. \end{aligned} \tag{9.5.7}$$

The second term can be written as

$$\frac{\mathbb{E}[\psi_v(1 - \psi_v)]}{\mathbb{E}[(1 - \psi_v)^2]} = \frac{m + \delta}{m(2v - 3) + (v - 1)\delta} = \frac{\tau - 2}{\tau - 1} (v - (3m + \delta - 1)/(2m + \delta))^{-1}. \tag{9.5.8}$$

The last product in (9.5.5), for $k = u + 1, \dots, w - 1$ results in

$$\begin{aligned} \mathbb{E}[(1 - \psi_k)^2] &= \frac{(m(2k - 3) + (k - 1)\delta)(m(2k - 3) + (k - 1)\delta + 1)}{(m(2k - 2) + k\delta)(m(2k - 2) + k\delta + 1)} \\ &= \frac{k - (3m + \delta)/(2m + \delta)}{k - 2m/(2m + \delta)} \frac{k - (3m + \delta - 1)/(2m + \delta)}{k - (2m - 1)/(2m + \delta)}. \end{aligned} \tag{9.5.9}$$

Using the recursive property $\Gamma(a + 1) = a\Gamma(a)$ of the Gamma function,

$$\begin{aligned} \prod_{k=u+1}^{w-1} \mathbb{E}[(1 - \psi_k)^2] &= \frac{\Gamma(u + 2 - (2m)/(2m + \delta))}{\Gamma(u + 2 - (3m + \delta)/(2m + \delta))} \frac{\Gamma(u + 2 - (2m - 1)/(2m + \delta))}{\Gamma(u + 2 - (3m + \delta - 1)/(2m + \delta))} \\ &\times \frac{\Gamma(w - (3m + \delta)/(2m + \delta))}{\Gamma(w - (2m)/(2m + \delta))} \frac{\Gamma(w - (3m + \delta - 1)/(2m + \delta))}{\Gamma(w - (2m - 1)/(2m + \delta))}. \end{aligned} \tag{9.5.10}$$

Equation (9.5.3) follows by combining (9.5.5), (9.5.7), (9.5.8), (9.5.9) and (9.5.10).

The last step of the proof is to evaluate the sum in (9.5.3), and combining the result with the multiplicative constant in front in (9.5.3). By Stirling's formula

$$\frac{\Gamma(x + a)}{\Gamma(x + b)} = x^{a-b}(1 + O(1/x)).$$

As a consequence, recalling that $\chi = (m + \delta)/(2m + \delta)$, the sum in (9.5.3) can be written as

$$\sum_{u=1}^{t-2} u^{2\chi-2}(1 + O(1/u)) \sum_{v=u+1}^{t-1} v^{-1}(1 + O(1/v)) \sum_{w=v+1}^t w^{-2\chi}(1 + O(1/w)). \tag{9.5.11}$$

We can approximate the sum in (9.5.11) with the corresponding integral using Euler-Maclaurin formula, thus obtaining

$$\int_1^t u^{2\chi-2} du \int_u^t v^{-1} dv \int_v^t w^{-2\chi} dw. \tag{9.5.12}$$

As $t \rightarrow \infty$, the order of magnitude of the integral in (9.5.12) is predicted by Theorem 9.1. If we evaluate the integral, then we obtain that the coefficient of the dominant term in (9.5.12) is $(2m + \delta)^2/\delta^2$ for $\tau > 2, \tau \neq 3$, and $1/6$ for $\tau = 3$.

Putting together these coefficients with the constant in front of the sum in (9.5.1) completes the proof of Theorem 9.2. \square

Remark 9.1 (Constant for general subgraphs). In the proof of Theorem 9.2, the hardest step is to prove (9.5.2), i.e., to find the expectation of the indicator functions in (9.5.1). This is the reason why for a general ordered subgraph (H, π) on k vertices it is hard to find the explicit constant as in (9.2.11). In fact, as we have done to move from (9.5.3) to (9.5.4), it is necessary to identify precisely, for every $v \in [t]$, how many times the terms ψ_v and $(1 - \psi_v)$ appear in the product inside the expectations in (9.5.3). This makes the evaluation of such terms complicated.

Typically, as it shown in (9.5.5), (9.5.7), (9.5.8), (9.5.9) and (9.5.10), the product of the constants obtained by evaluating the probability of an ordered subgraph (H, π) being present can be written as ratios of Gamma functions. The same constants can be found using the martingale approach as in [45, 205] and [106, Section 8.3], even though in this case constants are obtained through a recursive use of conditional expectation.

We remark that our method and the martingale method are equivalent. We focused on the Pólya urn interpretation of the graph since it highlights the dependence of the presence of edges on the *age* of vertices, that is directly related to the order of magnitude of degrees.

9.6 Conditional concentration: proof of Proposition 9.1

In the previous sections, we have considered the order of magnitude of the expectation of the number of occurrences of ordered subgraphs in PAM. In other words, for an ordered subgraph (H, ψ) we are able to identify the order of magnitude $f(t)$ of the expected number of occurrences $N_t(H, \pi)$, so that $\mathbb{E}[N_t(H, \pi)] = O(f(t))$. We now show how these orders of magnitude of the expected number of subgraphs determines the conditional convergence given in (9.2.7).

9.6.1 Bound with overlapping subgraphs

The Pólya urn graph in Definition 9.3 consists of a function of uniform random variables $(U_{v,j})_{v \in [t]}^{j \in [m]}$ and an independent sequence of Beta random variables $(\psi_v)_{v \in [t]}$.

We can interpret the sequence $(\psi_v)_{v \in [t]}$ as a sequence of *intensities* associated to the vertices, where a higher intensity corresponds to a higher probability of receiving a connection. The sequence $(U_{v,j})_{v \in [t]}^{j \in [m]}$ determines the attachment of edges. In particular, conditionally on the sequence $(\psi_v)_{v \in [t]}$, every edge is present *independently* (but with different probabilities).

For $t \in \mathbb{N}$, denote $\mathbb{P}_{\psi_t}(\cdot) = \mathbb{P}(\cdot | \psi_1, \dots, \psi_t)$, and similarly $\mathbb{E}_{\psi_t}[\cdot] = \mathbb{E}[\cdot | \psi_1, \dots, \psi_t]$. Furthermore, let $N_{t,\psi}(H, \pi)$ denote the number of times subgraph (H, π) appears conditionally on the ψ -variables. We now apply a conditional second moment method to $N_{t,\psi}(H, \pi)$. We use the notation introduced in Section 9.3, so that every possible realization of H in PAM corresponds to a finite set of edges $M_\ell(\mathbf{u}_\ell, \mathbf{v}_\ell, \mathbf{j}_\ell)$, where ℓ is the number of edges in H such that $v_h \xrightarrow{j_h} u_h$, i.e., u_h is the receiving vertex, and j_h is the label of the edge. For simplicity, we denote the set $M_\ell(\mathbf{u}_\ell, \mathbf{v}_\ell, \mathbf{j}_\ell)$ by M . For ease of notation, we assume that π is the identity map and drop the dependence on π . We prove the following results:

Lemma 9.5 (Bound on conditional variance). *Consider subgraph H . Then, \mathbb{P} -a.s.,*

$$\text{Var}_{\psi_t}(N_t(H)) \leq \mathbb{E}_{\psi_t}[N_t(H)] + \sum_{\hat{H} \in \hat{\mathcal{H}}} \mathbb{E}_{\psi_t}[N_t(\hat{H})],$$

where $\hat{\mathcal{H}}$ denotes the set of all possible attainable subgraphs \hat{H} that are obtained by merging two copies of H such that they share at least one edge.

Lemma 9.5 gives a bound on the conditional variance in terms of the conditional probabilities of observing two overlapping of the subgraph H at the same time. Notice that we require these copies to overlap at at least one edge, which is different than requiring that they are disjoint (they can share one or more vertices but no edges).

Proof of Lemma 9.5. We prove the bound in Lemma 9.5 by evaluating the conditional second moment of $N_t(H)$ as

$$\begin{aligned} \mathbb{E}_{\psi_t}[N_t(H)^2] &= \mathbb{E}_{\psi_t} \left[\sum_{M, M'} \mathbb{1}_{\{M \subseteq E(\text{PA}_t)\}} \mathbb{1}_{\{M' \subseteq E(\text{PA}_t)\}} \right] \\ &= \sum_{M, M'} \mathbb{P}_{\psi_t} \left(M \subseteq E(\text{PA}_t), M' \subseteq E(\text{PA}_t) \right), \end{aligned}$$

where M and M' are two sets of edges corresponding to two possible realizations of the subgraph H . Notice that M and M' are not necessarily distinct. We then have to evaluate the conditional probability of having both the sets M and M' simultaneously present in the graph. As a consequence, we conditional variance in Lemma 9.5 can be written as

$$\sum_{M \neq M'} \mathbb{P}_{\psi_t} \left(M \subseteq E(\text{PA}_t), M' \subseteq E(\text{PA}_t) \right) - \mathbb{P}_{\psi_t}(M \subseteq E(\text{PA}_t)) \mathbb{P}_{\psi_t}(M' \subseteq E(\text{PA}_t)). \tag{9.6.1}$$

We define

$$\mathcal{M} := \left\{ (M, M') : \exists (u, v, j) : (u, v, j) \in M, (u, v, j) \in M' \right\},$$

$$M \neq M', (M \cup M') \text{ defines an attainable subgraph} \}. \quad (9.6.2)$$

We then consider two different cases, i.e., whether (M, M') is in \mathcal{M} or not. If $(M, M') \notin \mathcal{M}$, then one of the three following situations occurs:

- ▷ $M \cup M'$ defines a subgraph that is not attainable (for instance, M and M' require that the same edge is attached to different vertices);
- ▷ $M \cup M'$ defines a subgraph that is attainable, M and M' are disjoint sets of labeled edges (they are allowed to share vertices);
- ▷ M and M' define the same attainable subgraph (so $M = M'$, thus labels of edges coincide).

When $M = M'$ we have that

$$\mathbb{P}_{\psi_t}(M \subseteq E(\text{PA}_t), M' \subseteq E(\text{PA}_t)) = \mathbb{P}_{\psi_t}(M \subseteq E(\text{PA}_t)),$$

so that the corresponding contribution in the sum in (9.6.1) is

$$\mathbb{P}_{\psi_t}(M \subseteq E(\text{PA}_t)) - \mathbb{P}_{\psi_t}(M \subseteq E(\text{PA}_t))^2 \leq \mathbb{P}_{\psi_t}(M \subseteq E(\text{PA}_t)),$$

and the sum over M gives the term $\mathbb{E}_{\psi_t}[N_t(H)]$ in the statement of Lemma 9.5. When $M \neq M'$ and $M \cup M'$ is attainable and their sets of edges are disjoint it follows directly from the independence of $(U_{v,j})_{v \in [t]}^{j \in [m]}$ and $(\psi_v)_{v \in [t]}$ that

$$\mathbb{P}_{\psi_t}(M \subseteq E(\text{PA}_t), M' \subseteq E(\text{PA}_t)) = \mathbb{P}_{\psi_t}(M \subseteq E(\text{PA}_t))\mathbb{P}_{\psi_t}(M' \subseteq E(\text{PA}_t)).$$

Thus, in this situation the corresponding contribution is zero. When (M, M') is not attainable the corresponding contribution is negative. When $(M, M') \in \mathcal{M}$ we bound the corresponding terms in (9.6.1) by $\mathbb{P}_{\psi_t}(M \subseteq E(\text{PA}_t), M' \subseteq E(\text{PA}_t))$, thus obtaining

$$\text{Var}_{\psi_t}(N_t(H)) \leq \mathbb{E}_{\psi_t}[N_t(H)] + \sum_{(M, M') \in \mathcal{M}} \mathbb{P}_{\psi_t}(M \cup M' \subseteq E(\text{PA}_t)), \quad (9.6.3)$$

We then rewrite this as

$$\text{Var}_{\psi_t}(N_t(H, \pi)) \leq \mathbb{E}_{\psi_t}[N_t(H)] + \sum_{\hat{H} \in \hat{\mathcal{H}}} \mathbb{E}_{\psi_t}[N_t(\hat{H})], \quad (9.6.4)$$

which proves the lemma. □

9.6.2 Criterion for conditional convergence

We now prove Proposition 9.1 using Lemma 9.5 and Lemma 9.7:

Proof of Proposition 9.1. It sufficient to show that for every fixed $\varepsilon > 0$,

$$\mathbb{P} \left(|N_{t,\psi}(H, \pi) - \mathbb{E}_t[N_t(H, \pi)]| > \varepsilon \mathbb{E}[N_t(H, \pi)] \right) = o(1).$$

We now apply Lemma 9.5, which yields

$$\begin{aligned} \mathbb{P} \left(|N_{t,\psi}(H, \pi) - \mathbb{E}_{\psi_t}[N_t(H, \pi)]| > \varepsilon \mathbb{E}[N_t(H, \pi)] \right) & \\ & \leq \frac{1}{\varepsilon^2 \mathbb{E}[N_t(H, \pi)]^2} \mathbb{E}[\text{Var}_{\psi_t}(N_t(H, \pi))] \\ & \leq \frac{\mathbb{E} \left[\mathbb{E}_{\psi_t}[N_t(H, \pi)] + \sum_{\hat{H} \in \hat{\mathcal{H}}} \mathbb{E}_{\psi_t}[N_t(\hat{H})] \right]}{\varepsilon^2 \mathbb{E}[N_t(H, \pi)]^2} \\ & = \frac{\mathbb{E}[N_t(H, \pi)] + \mathbb{E}[N_t(\hat{H})]}{\varepsilon^2 \mathbb{E}[N_t(H, \pi)]^2} = o(1). \end{aligned}$$

□

As an example, we consider triangles. Theorem 9.2 identifies the expected number of triangles, and by Theorem 9.1 we can show that $\mathbb{E}[\Delta_t^2] = \Theta(\mathbb{E}[\Delta_t]^2)$, so we are not able to apply the second moment method to Δ_t . Figure 9.6 suggests that $\Delta_t/\mathbb{E}[\Delta_t]$ converges to a limit that is not deterministic, i.e., in (9.2.9) the limiting X is a random variable.

However, we can prove that Δ_t is conditionally concentrated, as stated in Corollary 9.2. The proof of Corollary 9.2 follows directly from Proposition 9.1, the fact that $\mathbb{E}[\Delta_t] = \Theta(t^{(3-\tau)/(\tau-1)} \log(t))$ as given by Theorem 9.2, and Figure 9.5, that contains the information on the subgraphs consisting of two triangles sharing one or two edges.

9.6.3 Non-concentrated subgraphs

We now show that for most ψ -sequences, the other direction in Proposition 9.1 also holds. That is, if there exists a subgraph composed of two merged copies of H such that the condition in Proposition 9.1 does not hold, then for most ψ -sequences, H is not conditionally concentrated.

Proposition 9.2. *Consider a subgraph (H, π) such that $\mathbb{E}[N_t(H, \pi)] \rightarrow \infty$ as $t \rightarrow \infty$. Suppose that there exists a subgraph \hat{H} , composed of two distinct copies of (H, π) with at least one edge in common such that $\mathbb{E}[N_t(\hat{H})]/\mathbb{E}[N_t(H, \pi)] \rightarrow 0$ as $t \rightarrow \infty$. Then, for any $\varepsilon > 0$, there exists $\eta > 0$ such that*

$$\mathbb{P} \left(\frac{\text{Var}_{\psi_t}(N_t(H, \pi))}{\mathbb{E}[N_t(H, \pi)]^2} > \eta \right) \geq 1 - \varepsilon. \tag{9.6.5}$$

To prove Proposition 9.2 we need a preliminary result on the maximum intensity of the Pólya urn graph:

Lemma 9.6. *For every $\varepsilon > 0$ there exists $K = K(\varepsilon) \in \mathbb{N}$ such that*

$$\mathbb{P}\left(\bigcap_{k \geq K} \left\{ \psi_k \leq \frac{(\log k)^2}{(2m + \delta)k} \right\}\right) \geq 1 - \varepsilon.$$

Lemma 9.6 is a part of a more general coupling result between $(\psi_k)_{k \in \mathbb{N}}$ and a sequence of i.i.d. Gamma random variables. We refer to [20, Lemma 3.2] and [107, Lemma 4.10] for more detail. We now state the lemma we need to prove Proposition 9.2:

Lemma 9.7 (Maximum intensity). *For every $\varepsilon > 0$ there exists $\omega = \omega(\varepsilon) \in (0, 1)$ such that, for every $t \in \mathbb{N}$,*

$$\mathbb{P}\left(\max_{i \in 2, \dots, t} \psi_i < \omega\right) \geq 1 - \varepsilon.$$

Proof. Fix $\varepsilon > 0$, and consider $K(\varepsilon/2)$ as given by Lemma 9.6. For every $\omega \in (0, 1)$ we can write

$$\mathbb{P}\left(\max_{i \in 2, \dots, t} \psi_i < \omega\right) = \mathbb{P}\left(\max_{i \in 2, \dots, K} \psi_i < \omega\right) \mathbb{P}\left(\max_{i \in [t] \setminus [K]} \psi_i < \omega\right), \quad (9.6.6)$$

where we used the independence of ψ_2, \dots, ψ_t . If $t > K$ the second term in the right-hand side of (9.6.6) is well defined, otherwise we only have the first term. Define,

$$\omega_1 = \begin{cases} \frac{(\log K)^2}{(2m + \delta)K} & \text{if } t > K, \\ 0 & \text{if } t \leq K. \end{cases}$$

Notice that, since the function $k \mapsto \frac{(\log k)^2}{(2m + \delta)k}$ is decreasing, it follows that

$$\mathbb{P}\left(\max_{i \in [t] \setminus [K]} \psi_i < \omega_1\right) \geq 1 - \varepsilon/2. \quad (9.6.7)$$

Define the random variable $X_K = \max_{i \in 2, \dots, K} \psi_i$, denote its distribution function by F_K and the inverse of its distribution function by F_K^{-1} . Consider $\omega_2 = F_K^{-1}(1 - \varepsilon/2)$, that implies

$$\mathbb{P}\left(\max_{i \in [K]} \psi_i < \omega_2\right) = 1 - \varepsilon/2. \quad (9.6.8)$$

Consider then $\omega = \max\{\omega_1, \omega_2\}$. Using (9.6.7) and (9.6.8) with ω in (9.6.6), it follows that

$$\mathbb{P}\left(\max_{i \in 2, \dots, K} \psi_i < \omega\right) \mathbb{P}\left(\max_{i \in [t] \setminus [K]} \psi_i < \omega\right) \geq (1 - \varepsilon/2)^2 \geq 1 - \varepsilon,$$

which completes the proof. \square

Proof of Proposition 9.2. We use the expression of the conditional variance of (9.6.1). We first study the term in the conditional variance corresponding to \hat{H} . Let $\tilde{\mathcal{M}}$ denote the set of labeled edges M, M' that together form subgraph \hat{H} . Let the edges that M and M' share be denoted by M_s . Furthermore, let $\tilde{\mathcal{M}}_1$ denote the set of labeled edges

M, M' that together form subgraph \hat{H} that do not use vertex 1. We can then write this term as

$$\begin{aligned} & \sum_{M, M' \in \tilde{\mathcal{M}}} \mathbb{P}_{\psi_t} (M \cup M' \subseteq E(PA_t)) (1 - \mathbb{P}_{\psi_t} (M_s \subseteq E(PA_t))) \\ & \geq \sum_{M, M' \in \tilde{\mathcal{M}}_1} \mathbb{P}_{\psi_t} (M \cup M' \subseteq E(PA_t)) (1 - \psi_{\max}) \\ & = (1 - \psi_{\max}) \mathbb{E}_{\psi_t} [N_t(\hat{H})] \end{aligned} \tag{9.6.9}$$

where the inequality uses (9.3.10), and $\psi_{\max} = \max_{i \in 2, \dots, t} \psi_i$. Here we excluded vertex 1 from the number of subgraphs with negligible error. By Lemma 9.7 there exists ω such that with probability at least $1 - \varepsilon$, $\psi_{\max} < \omega < 1$.

By the assumption on \hat{H} , $\mathbb{E} [N_t(\hat{H})] \geq \tilde{C} \mathbb{E} [N_t(H, \pi)]^2$ for some $\tilde{C} > 0$. We use that $\mathbb{E}_{\psi_t} [N_t(\hat{H})] = O_{\mathbb{P}}(\mathbb{E} [N_t(\hat{H})])$. Thus, for t sufficiently large, we can bound the contribution from subgraph \hat{H} to the conditional variance from below with probability at least $1 - \varepsilon$ by

$$\sum_{M, M' \in \tilde{\mathcal{M}}} \mathbb{P}_{\psi_t} (M \cup M' \subseteq E(PA_t)) (1 - \mathbb{P}_{\psi_t} (M_s \subseteq E(PA_t))) \geq C \mathbb{E}_{\psi_t} [N_t(H, \pi)]^2, \tag{9.6.10}$$

for some $C > 0$.

The only terms that have a negative contribution to (9.6.1) are the terms where $M \cup M'$ is a non-attainable subgraph. In that situation, $\mathbb{P}_{\psi_t} (M \subseteq E(PA_t), M' \subseteq E(PA_t)) = 0$. Furthermore, the sum over $\mathbb{P}_{\psi_t} (M \subseteq E(PA_t)) \mathbb{P}_{\psi_t} (M' \subseteq E(PA_t)) \leq \mathbb{E}_{\psi_t} [N_t(H, \pi)]^2 / n^2$, since the two subgraphs share at least two vertices. Therefore, the negative terms in the conditional variance scale as most as $\mathbb{E}_{\psi_t} [N_t(H, \pi)] / n^2$. We therefore obtain that with probability at least $1 - \varepsilon$,

$$\text{Var}_{\psi_t} (N_t(H, \pi)) \geq \eta \mathbb{E}_{\psi_t} [N_t(H, \pi)]^2, \tag{9.6.11}$$

for some $\eta > 0$, which proves the proposition. \square

9.7 Discussion

In this chapter, we have investigated the expected number of times a graph H appears as a subgraph of a PAM for any degree exponent τ . We find the scaling of the expected number of such subgraphs in terms of the graph size t and the degree exponent τ by defining an optimization problem that finds the optimal structure of the subgraph in terms of the ages of the vertices that form subgraph H and by using the interpretation of the PAM as a Pólya urn graph.

We derive the asymptotic scaling of the number of subgraphs. For the triangle subgraph, we obtain more precise asymptotics. It would be interesting to obtain precise asymptotics of the expected number of other types of subgraphs as well. In particular, this is necessary to compute the variance of the number of subgraphs, which may allow us to derive laws of large numbers for the number of subgraphs.

We show that different subgraphs may have significantly different concentration properties. Therefore, identifying the distribution of the number of rescaled subgraphs for any type of subgraph remains a challenging open problem.

Another interesting extension would be to investigate other types of PAMs, for example models that allow for self-loops, or models that include extra triangles.

We further prove results for the number of subgraphs of fixed size k , while the graph size tends to infinity. It would also be interesting to let the subgraph size grow with the graph size, for example by counting the number of cycles of a certain length that grows in the graph size.

Finally, we investigate the number of times H appears as a subgraph of a PAM. It is also possible to count the number of times H appears as an *induced* subgraph instead, forbidding edges that are not present in H to be present in the larger graph. It would be interesting to see whether the optimal subgraph structure is different from the optimal induced subgraph structure.

10 Finding induced subgraphs in inhomogeneous random graphs

Based on:
E. Cardinaels, J.S.H. van Leeuwen, C. Stegehuis
Finding induced subgraphs in scale-free inhomogeneous random graphs
Algorithms and Models for the Web Graph p. 1-15 (2018)

In Chapters 7 and 8 we have shown that induced subgraphs in rank-1 inhomogeneous random graphs typically occur on vertices with specific degree ranges. A wide class of induced subgraphs appears most frequently on vertices with degrees scaling as \sqrt{n} . In this chapter, we show how we can use this knowledge to find a set of vertices that induce some connected graph H in a rank-1 inhomogeneous random graphs with infinite-variance power-law degrees. We provide a fast algorithm that determines, for any connected graph H on k vertices, whether it exists as induced subgraph in a random graph with n vertices and returns an instance of H if it does. By exploiting the scale-free graph structure, the algorithm runs in nk time with high probability for small values of k . We test our algorithm on several real-world data sets.

10.1 Introduction

The induced subgraph isomorphism problem asks whether a large graph G contains a connected graph H as an induced subgraph. When k is allowed to grow with the graph size n , this problem is NP-hard in general. For example, k -clique and k -induced cycle, which are special cases of H , are known to be NP-hard [90, 125]. For fixed k , they can be solved in polynomial time $O(n^k)$ by searching for H on all possible combinations of k vertices. Several randomized and non-randomized algorithms exist to improve upon this trivial way of finding H [98, 167, 195, 218].

On real-world networks, many algorithms were observed to run much faster than predicted by the worst-case running time of algorithms. This may be ascribed to some of the properties that many real-world networks share [48], such as the power-law degree distribution found in many networks [6, 79, 122, 212]. One way of exploiting these power-law degree distributions is to design algorithms that work well on random graphs with power-law degree distributions. For example, finding the largest clique in a network is NP-complete for general networks [125]. However, in random graph models such as the Erdős-Rényi random graph and the inhomogeneous random graph, their specific structures can be exploited to design

fixed parameter tractable (FPT) algorithms that efficiently find a clique of size k [83, 86] or the largest independent set [105].

In this chapter, we study algorithms that are designed to perform well for the rank-1 inhomogeneous random graph. The inhomogeneous random graph has a densely connected core containing many cliques, consisting of vertices with degrees $\sqrt{n \log(n)}$ and larger. In this densely connected core, the probability of an edge being present is close to one, so that it contains many complete graphs [119]. This observation was exploited in [85] to efficiently determine whether a clique of size k occurs as a subgraph in an inhomogeneous random graph. When searching for *induced* subgraphs however, some edges are required not to be present. Therefore, searching for induced subgraphs in the entire core is not efficient. We show that a connected subgraph H can be found as an induced subgraph by scanning only vertices that are on the boundary of the core: vertices with degrees proportional to \sqrt{n} .

We present an algorithm that first selects the set of vertices with degrees proportional to \sqrt{n} , and then randomly searches for H as an induced subgraph on a subset of k of those vertices. The first algorithm we present does not depend on the specific structure of H . For general sparse graphs, the best known algorithms to solve subgraph isomorphism on 3 or 4 vertices run in $O(n^{1.41})$ or $O(n^{1.51})$ time with high probability [218]. For small values of k , our algorithm solves subgraph isomorphism on k nodes in linear time with high probability on rank-1 inhomogeneous random graphs. However, the graph size needs to be very large for our algorithm to perform well. We therefore present a second algorithm that again selects the vertices with degrees proportional to \sqrt{n} , and then searches for induced subgraph H in a more efficient way. This algorithm has the same performance guarantee as our first algorithm, but performs much better in simulations.

We test our algorithm on large inhomogeneous random graphs, where it indeed efficiently finds induced subgraphs. We also test our algorithm on real-world network data with power-law degrees. There our algorithm does not perform well, probably due to the fact that the densely connected core of some real-world networks may not be the vertices of degrees at least proportional to \sqrt{n} . We then show that a slight modification of our algorithm that looks for induced subgraphs on vertices of degrees proportional to n^γ for some other value of γ performs better on real-world networks, where the value of γ depends on the specific network.

10.1.1 Model

As a random graph null model, we use the rank-1 inhomogeneous random graph, described in Section 1.1.4. We assume that the weights are an i.i.d. sample from the power-law distribution

$$\mathbb{P}(w_i > k) = Ck^{1-\tau} \quad (10.1.1)$$

for some constant C and for $\tau \in (2, 3)$. In this chapter, we focus on the Chung-Lu setting of the rank-1 inhomogeneous random graph, so that two vertices with weights w and w' are connected with probability

$$p(w, w') = \min\left(\frac{ww'}{\mu n}, 1\right), \quad (10.1.2)$$

Algorithm 1: Finding induced subgraph H (random search)

Input : $H, G = (V, E), \mu, f_1 = f_1(n), f_2 = f_2(n)$.
Output: Location of H in G or fail.

- 1 Define $n = |V|, I_n = [\sqrt{f_1\mu n}, \sqrt{f_2\mu n}]$ and set $V' = \emptyset$.
- 2 **for** $i \in V$ **do**
- 3 | **if** $D_i \in I_n$ **then** $V' = V' \cup i$;
- 4 **end**
- 5 Divide the vertices in V' randomly into $\lfloor |V'|/k \rfloor$ sets $S_1, \dots, S_{\lfloor |V'|/k \rfloor}$.
- 6 **for** $j = 1, \dots, \lfloor |V'|/k \rfloor$ **do**
- 7 | **if** H is an induced subgraph on S_j **then return** location of H ;
- 8 **end**

where μ denotes the mean value of the power-law distribution (10.1.1). Choosing the connection probability in this way ensures that the expected degree of a vertex with weight w is w .

10.1.2 Algorithms

We now describe two randomized algorithms that determine whether a connected graph H is an induced subgraph in an inhomogeneous random graph and finds the location of such a subgraph if it exists. Algorithm 1 first selects the vertices that are on the boundary of the core of the graph: vertices with degrees scaling as \sqrt{n} . Then, the algorithm randomly divides these vertices into sets of k vertices. If one of these sets contains H as an induced subgraph, the algorithm terminates and returns the location of H . If this is not the case, then the algorithm fails. In the next section, we show that for k small enough, the probability that the algorithm fails is asymptotically small. This means that H is present as an induced subgraph on vertices that are on the boundary of the core with high probability.

Algorithm 1 is similar to the algorithm in [86] designed to find cliques in random graphs. The major difference is that the algorithm to find cliques selects all vertices with degrees larger than $\sqrt{f_1\mu n}$ for some function f_1 . This algorithm is not efficient for detecting other induced subgraphs than cliques, since vertices with high degrees will be connected with probability close to one. For this reason, the algorithm also removes vertices with degrees larger than $\sqrt{f_2\mu n}$.

The following theorem gives a bound for the performance of Algorithm 1 for small values of k and a wide range of functions f_1 and f_2 :

Theorem 10.1. *Let $k = o(\sqrt{\log(n)/\log(\log(n))})$, and choose $f_1 = f_1(n) \geq 1/\log(n)$ and $f_1 < f_2 < 1$. Then, with high probability, Algorithm 1 detects induced subgraph H on k vertices in an inhomogeneous random graph with n vertices and weights distributed as in (10.1.1) in time nk .*

Thus, for small values of k , Algorithm 1 finds an instance of H in linear time.

A problem with parameter k is called fixed parameter tractable (FPT) if it can be solved in $f(k)n^{O(1)}$ time for some function $f(k)$, and it is called typical FPT (typFPT) if

it can be solved in $f(k)n^{g(n)}$ for some function $g(n) = O(1)$ with high probability [83]. As a corollary of Theorem 10.1 we obtain that the k -induced subgraph problem on the inhomogeneous random graph is in *typFPT* for any subgraph H , similarly to the k -clique problem [86].

Corollary 10.1. *The induced subgraph problem on the rank-1 inhomogeneous random graph is in *typFPT*.*

In theory Algorithm 1 detects any motif on k vertices in linear time for small k . However, this only holds for large values of n , which can be understood as follows. In Lemma 10.2, we show that $|V'| = \Theta(n^{(3-\tau)/2})$, thus tending to infinity as n grows large. However, when $n = 10^7$ and $\tau = 2.5$, this means that the size of the set V' is only proportional to $10^{1.75} = 56$ vertices. Therefore, the number of sets S_j constructed in Algorithm 1 is also small. Even though the probability of finding motif H in any such set is proportional to a constant, this constant may be small, so that for finite n the algorithm almost always fails. Thus, for Algorithm 1 to work, n needs to be large enough so that $n^{(3-\tau)/2}$ is large as well.

The algorithm can be significantly improved by changing the search for H on vertices in the set V' . In Algorithm 2 we propose a search for motif H similar to the Kashtan motif sampling algorithm [127]. Rather than sampling k vertices randomly, it samples one vertex randomly, and then randomly increases the set S by adding vertices in its neighborhood. This already guarantees the vertices in list S_j to be connected, making it more likely for them to form a specific connected motif together. In particular, we expand the list S_j in such a way that the vertices in S_j are guaranteed to form a spanning tree of H as a subgraph. This is ensured by choosing the list T^H that specifies at which vertex in S_j we expand S_j by adding a new vertex. For example, if $k = 4$ and we set $T^H = [1, 2, 3]$ we first add a random neighbor of the first vertex, then we look for a random neighbor of the previously added vertex, and then we add a random neighbor of the third added vertex. Thus, setting $T^H = [1, 2, 3]$ ensures that the set S_j contains a path of length three, whereas setting $T^H = [1, 1, 1]$ ensures that the set S_j contains a star-shaped subgraph. Depending on which subgraph H we are looking for, we can define T^H so that it ensures that the set S_j at least contains a spanning tree of motif H in Step 6 of the algorithm.

The selection on the degrees ensures that the degrees are sufficiently high so that the probability of finding such a connected set on k vertices is high, as well as that the degrees are sufficiently low to ensure that we do not only find complete graphs because of the densely connected core of the inhomogeneous random graph. The probability that Algorithm 2 indeed finds the desired motif H in any check is of constant order of magnitude, similar to Algorithm 1. Therefore, the performance guarantee of both algorithms is similar. However, Section 10.3 shows that in practice Algorithm 2 performs much better, since for finite n , k connected vertices are more likely to form a motif than k randomly chosen vertices.

The following theorem shows that indeed Algorithm 2 has similar performance guarantees as Algorithm 1:

Theorem 10.2. *Choose $f_1 = f_1(n) \geq 1/\log(n)$ and $f_1 < f_2 < 1$. Choose $s = \Omega(n^\alpha)$ for some $0 < \alpha < 1$, such that $s \leq n/k$. Then, Algorithm 2 detects induced subgraph H on*

Algorithm 2: Finding induced subgraph H (neighborhood search)

Input : $H, G = (V, E), \mu, f_1 = f_1(n), f_2 = f_2(n), s$.
Output: Location of H in G or fail.

- 1 Define $n = |V|, I_n = [\sqrt{f_1 \mu n}, \sqrt{f_2 \mu n}]$ and set $V' = \emptyset$.
- 2 **for** $i \in V$ **do**
- 3 | **if** $D_i \in I_n$ **then** $V' = V' \cup i$;
- 4 **end**
- 5 Let G' be the induced subgraph of G on vertices V' .
- 6 Set T^H consistently with motif H .
- 7 **for** $j=1, \dots, s$ **do**
- 8 | Pick a random vertex $v \in V'$ and set $S_j = v$.
- 9 | **while** $|S_j| \neq k$ **do**
- 10 | | Pick a random $v' \in N_{G'}(S_j[T^H[j]]) : v' \notin S_j$
- 11 | | Add v' to S_j .
- 12 | **end**
- 13 | **if** H is an induced subgraph on S_j **then return** location of H ;
- 14 **end**

$k = o(\sqrt{\log(n)/\log(\log(n))})$ vertices on an inhomogeneous random graph with n vertices and weights distributed as in (10.1.1) in time nk with high probability.

The proofs of Theorem 10.1 and 10.2 rely on the fact that for small k , any subgraph on k vertices is present in G' with high probability. This means that after the degree selection step of Algorithms 1 and 2, for small k , any motif finding algorithm can be used to find motif H on the remaining graph G' , such as the Grochow-Kellis algorithm [98], the MAvisto algorithm [195] or the MODA algorithm [167]. In the proofs of Theorem 10.1 and 10.2, we show that G' has $\Theta(n^{(3-\tau)/2})$ vertices with high probability. Thus, the degree selection step reduces the problem of finding a motif H on n vertices to finding a motif on a graph with $\Theta(n^{(3-\tau)/2})$ vertices, significantly reducing the running time of these algorithms.

10.2 Analysis of Algorithms 1 and 2

We prove Theorem 10.1 using two lemmas. The first lemma relates the degrees of the vertices to their weights. The connection probabilities in the inhomogeneous random graph depend on the weights of the vertices. In Algorithm 1, we select vertices based on their degrees instead of their unknown weights. The following lemma shows that the weights of the vertices in V' are close to their degrees. In particular, we show that when we select vertices with degrees in $[\sqrt{f_1 \mu n}, \sqrt{f_2 \mu n}]$, their weights lie in the interval $[(1 - \varepsilon)\sqrt{f_1 \mu n}, (1 + \varepsilon)\sqrt{f_2 \mu n}]$ with high probability:

Lemma 10.1 (Degrees and weights.). *Fix $\varepsilon > 0$, and define $J_n = [(1 - \varepsilon)\sqrt{f_1 \mu n}, (1 +$*

$\varepsilon)\sqrt{f_2\mu n}]$. Then, for some $K > 0$,

$$\mathbb{P}(\exists i \in V' : w_i \notin I_n) \leq Kn \exp\left(-\varepsilon^2\sqrt{\mu n} \min\left(\frac{\sqrt{f_1}}{1-\varepsilon}, \frac{\sqrt{f_2}}{1+\varepsilon}\right)/2\right). \quad (10.2.1)$$

Proof. Fix $i \in V$. Then,

$$\begin{aligned} \mathbb{P}\left(w_i < (1-\varepsilon)\sqrt{f_1\mu n}, D_i \in I_n\right) &= \frac{\mathbb{P}\left(D_i \in I_n \mid w_i < (1-\varepsilon)\sqrt{f_1\mu n}\right)}{\mathbb{P}\left(w_i < (1-\varepsilon)\sqrt{f_1\mu n}\right)} \\ &\leq \frac{\mathbb{P}\left(D_i > \sqrt{f_1\mu n} \mid w_i = (1-\varepsilon)\sqrt{f_1\mu n}\right)}{1 - C((1-\varepsilon)\sqrt{f_1\mu n})^{1-\tau}} \\ &\leq K_1\mathbb{P}\left(D_i > \sqrt{f_1\mu n} \mid w_i = (1-\varepsilon)\sqrt{f_1\mu n}\right), \end{aligned} \quad (10.2.2)$$

for some $K_1 > 0$. Here the first inequality follows because the probability that a vertex with weight w_1 has degree at least $\sqrt{f_1\mu n}$ is larger than the probability that a vertex of weight w_2 has degree at least $\sqrt{f_1\mu n}$ when $w_1 > w_2$. Conditionally on the weights, D_i is the sum of $n-1$ independent indicators indicating the presence of an edge between vertex i and the other vertices and that $\mathbb{E}[D_i] = w_i$. Therefore, by the Chernoff bound

$$\mathbb{P}(D_i > w_i(1+\delta)) \leq \exp(-\delta^2 w_i/2). \quad (10.2.3)$$

Therefore, choosing $\delta = \varepsilon/(1-\varepsilon)$ yields

$$\mathbb{P}\left(D_i > \sqrt{f_1\mu n} \mid w_i = (1-\varepsilon)\sqrt{f_1\mu n}\right) \leq \exp\left(-\frac{\varepsilon^2\sqrt{f_1\mu n}}{2(1-\varepsilon)}\right)(1+o(1)). \quad (10.2.4)$$

Combining this with (10.2.2) and taking the union bound over all vertices then results in

$$\mathbb{P}\left(\exists i : D_i \in I_n, w_i < (1-\varepsilon)\sqrt{f_1\mu n}\right) \leq K_2 n \exp\left(-\frac{\varepsilon^2}{2(1-\varepsilon)}\sqrt{f_1\mu n}\right), \quad (10.2.5)$$

for some $K_2 > 0$. Similarly,

$$\mathbb{P}\left(\exists i : D_i \in I_n, w_i > (1+\varepsilon)\sqrt{f_2\mu n}\right) \leq K_3 n \exp\left(-\frac{\varepsilon^2}{2(1+\varepsilon)}\sqrt{f_2\mu n}\right), \quad (10.2.6)$$

for some $K_3 > 0$, which proves the lemma. \square

The second lemma shows that after deleting all vertices with degrees outside of I_n defined in Step 1 of Algorithm 1, still polynomially many vertices remain with high probability.

Lemma 10.2 (Polynomially many nodes remain.). *There exists $\gamma > 0$ such that*

$$\mathbb{P}(|V'| < \gamma n^{(3-\tau)/2}) \leq 2 \exp(-\Theta(n^{(3-\tau)/2})). \quad (10.2.7)$$

Proof. Let \mathcal{E} denote the event that all vertices $i \in V'$ satisfy $w_i \in J_n$ for some $\varepsilon > 0$, with J_n as in Lemma 10.1. Let W' be the set of vertices with weights in J_n . Under the event \mathcal{E} , $|V'| \leq |W'|$. Then, by Lemma 10.1,

$$\begin{aligned} \mathbb{P}\left(|V'| < \gamma n^{(3-\tau)/2}\right) &\leq \mathbb{P}\left(|V'| < \gamma n^{(3-\tau)/2} \mid \mathcal{E}\right) + \mathbb{P}(\mathcal{E}^c) \\ &\leq \mathbb{P}\left(|W'| < \gamma n^{(3-\tau)/2}\right) + Kn \exp\left(-\frac{\varepsilon^2(1-\varepsilon)}{2(1+\varepsilon)}\sqrt{f_1\mu n}\right). \end{aligned} \tag{10.2.8}$$

Furthermore,

$$\mathbb{P}(w_i \in J_n) = C((1-\varepsilon)\sqrt{f_1\mu n})^{1-\tau} - C((1+\varepsilon)\sqrt{f_2\mu n})^{1-\tau} \geq \alpha_1(\sqrt{\mu n})^{1-\tau} \tag{10.2.9}$$

for some constant $\alpha_1 > 0$ because $f_1 < f_2$. Thus, each of the n vertices is in the set W' independently with probability at least $\alpha_1(\sqrt{\mu n})^{1-\tau}$. Choose $0 < \gamma < \alpha_1$. Applying the Chernoff bound then shows that

$$\mathbb{P}\left(|W'| < \gamma n^{(3-\tau)/2}\right) \leq \exp\left(-\frac{(\alpha_1 - \gamma)^2}{2\alpha_1}n^{(3-\tau)/2}\right), \tag{10.2.10}$$

which together with (10.2.8) and the fact that $\sqrt{f_1\mu n} = \Omega(n^{(3-\tau)/2})$ for $\tau \in (2, 3)$ proves the lemma. \square

We now use these lemmas to prove Theorem 10.1:

Proof of Theorem 10.1. We condition on the event that V' is of polynomial size (Lemma 10.2) and that the weights are within the constructed lower and upper bounds (Lemma 10.1), since both events occur with high probability. This bounds the edge probability between any pair of nodes i and j in V' as

$$p_{ij} < \min\left(\frac{(1+\varepsilon)\sqrt{f_2\mu n}(1+\varepsilon)\sqrt{f_2\mu n}}{\mu n}, 1\right) = f_2(1+\varepsilon)^2, \tag{10.2.11}$$

so that $p_{ij} \leq p_+ = f_2(1+\varepsilon)^2 < 1$ if we choose ε sufficiently small. Similarly,

$$p_{ij} > \min\left(\frac{(1-\varepsilon)^2\sqrt{f_1\mu n^2}}{\mu n}\right) \geq \frac{c_2}{\log(n)}, \tag{10.2.12}$$

for some $c_2 > 0$ by our choice of f_1 . Let $E := |E_H|$ be the number of edges in H . We upper bound the probability of not finding H in one of the partitions of size k of V' as $1 - p_-^E(1 - p_+)^{\binom{k}{2} - E}$. Since all partitions are disjoint we can upper bound the probability of not finding H in any of the partitions as

$$\mathbb{P}(H \text{ not in the partitions}) \leq \left(1 - p_-^E(1 - p_+)^{\binom{k}{2} - E}\right)^{\lceil \frac{|V'|}{k} \rceil}. \tag{10.2.13}$$

Using that $E \leq k^2$, $\binom{k}{2} - E \leq k^2$ and that $1 - x \leq e^{-x}$ results in

$$\mathbb{P}(H \text{ not in the partitions}) \leq \exp\left(-p_-^{k^2}(1 - p_+)^{k^2} \left\lceil \frac{|V'|}{k} \right\rceil\right). \tag{10.2.14}$$

Since $|V'| = \Theta\left(n^{\frac{3-\tau}{2}}\right)$, $\lceil |V'|/k \rceil \geq dn^{\frac{3-\tau}{2}}/k$ for some constant $d > 0$. We fill in the expressions for p_- and p_+ , with $c_3 > 0$ a constant

$$\mathbb{P}(H \text{ not in the partitions}) \leq \exp\left(-\frac{dn^{\frac{3-\tau}{2}}}{k} \left(\frac{c_3}{\log n}\right)^{k^2}\right). \quad (10.2.15)$$

Applying that $k = o(\sqrt{\log(n)/\log(\log(n))})$ yields

$$\begin{aligned} \mathbb{P}(H \text{ not in the partitions}) &\leq \exp\left(-\frac{dn^{\frac{3-\tau}{2}}}{\sqrt{\log(n)/\log(\log(n))}} \left(\frac{c_3}{\log n}\right)^{o\left(\frac{\log(n)}{\log(\log(n))}\right)}\right) \\ &\leq \exp\left(-dn^{\frac{3-\tau}{2}-o(1)}\right). \end{aligned} \quad (10.2.16)$$

Hence, the inner expression grows polynomially so that the probability of not finding H in one of the partitions is negligibly small. The running time of the partial search is given by

$$\frac{|V'|}{k} \binom{k}{2} \leq \frac{n}{k} \binom{k}{2} \leq nk \quad (10.2.17)$$

which concludes the proof. \square

Proof of Corollary 10.1. If $k > \log^{\frac{1}{3}}(n)$, we can determine whether H is an induced subgraph by an exhaustive search in time

$$\binom{n}{k} \binom{k}{2} \leq \frac{n^k k(k-1)}{k^2} \leq kn^k \leq ke^{k^4} \leq ne^{k^4}, \quad (10.2.18)$$

since for all sets of k vertices the presence or absence of $\binom{k}{2}$ edges needs to be checked. For $k \leq \log^{\frac{1}{3}}(n)$, Theorem 10.1 shows that the induced subgraph isomorphism problem can be solved in time $nk \leq ne^{k^4}$. Thus, with high probability, the induced subgraph isomorphism problem can be solved in ne^{k^4} time, which proves that it is in typFPT . \square

Proof of Theorem 10.2. The proof of Theorem 10.2 is similar to the proof of Theorem 10.1. The only way that Algorithm 2 differs from Algorithm 1 is in the selection of the sets S_j . As in the previous theorem, we condition on the event that $|V'| = \Theta(n^{(3-\tau)/2})$ (Lemma 10.2) and that the weights of the vertices in G' are bounded as in Lemma 10.1.

The graph G' constructed in Step 5 of Algorithm 2 then consists of $\Theta(n^{(3-\tau)/2})$ vertices. Furthermore, by the bound (10.2.12) on the connection probabilities of all vertices in G' , the expected degree of a vertex i in G' satisfies $\mathbb{E}[D_{i,G'}] \geq c_1 n^{(3-\tau)/2} / \log(n)$ for some $c_1 > 0$. We can use similar arguments as in Lemma 10.1 to show that $D_{i,G'} \geq c_2 n^{(3-\tau)/2} / \log(n)$ for some $c_2 > 0$ with high probability for all vertices in G' . Therefore, from now on we assume that

$$D_{i,G'} \geq c_2 n^{(3-\tau)/2} / \log(n) \quad \forall i \in V'. \quad (10.2.19)$$

Since G' consists of $\Theta(n^{(3-\tau)/2})$ vertices, $D_{i,G'} = O(n^{(3-\tau)/2})$ as well. This means that for $k = o(\sqrt{\log(n)/\log(\log(n))})$, Steps 8-11 are able to find a connected subgraph on k vertices with high probability.

Algorithm 2 then constructs s sets $(S_j)_{j \in [s]}$ on which it checks whether H is an induced subgraph. However, these sets may overlap, creating dependencies between the edge presences of the overlapping sets. We now show that the probability that these sets overlap is small. We compute the probability that S_j is disjoint with the $j - 1$ previously constructed sets. The probability that the first vertex does not overlap with the previous sets is bounded by $1 - jk/|V'|$, since that vertex is chosen uniformly at random. The second vertex is chosen in a size-biased manner, since it is chosen by following a random edge. Conditionally on the degrees, the probability that vertex i is added can therefore be written as

$$\mathbb{P}(\text{vertex } i \text{ is added}) = \frac{D_{i,G'}}{\sum_{s \in [|V'|]} D_{s,G'}}. \tag{10.2.20}$$

Using (10.2.19) and the fact that $D_{i,G'} = O(n^{(3-\tau)/2})$ for all $i \in V'$ then results in

$$\mathbb{P}(\text{vertex } i \text{ is added}) \leq M \log(n)/|V'| \tag{10.2.21}$$

for some constant $M > 0$. Therefore, the probability that S_j does not overlap with $(S_l)_{l \in [j-1]}$ can be bounded from below by

$$\mathbb{P}(S_j \text{ does not overlap with } (S_l)_{l \in [j-1]}) \geq \left(1 - \frac{kj}{|V'|}\right) \left(1 - \frac{Mkj \log(n)}{|V'|}\right)^{k-1}, \tag{10.2.22}$$

where we used that $(S_l)_{l \in [j-1]}$ contains at most $k(j - 1) \leq kj$ distinct vertices. Thus, the probability that all j sets do not overlap can be bounded as

$$\mathbb{P}(S_j \cap S_{j-1} \cap \dots \cap S_1 = \emptyset) \geq \left(1 - \frac{Mkj \log(n)}{|V'|}\right)^{jk}, \tag{10.2.23}$$

which tends to one when $jk = o(n^{(3-\tau)/4})$.

Let s_{dis} denote the number of disjoint sets out of the s sets constructed in Algorithm 2. Then, by the previous argument, when $s = \Omega(n^\alpha)$ for some $\alpha > 0$, $s_{\text{dis}} > n^\beta$ for some $\beta > 0$ with high probability, because $k = o(\sqrt{\log(n)/\log(\log(n))})$.

The probability that H is present as an induced subgraph is bounded similarly as in Theorem 10.1. We already know that $k - 1$ edges are present. For all other $E - (k - 1)$ edges of H , and all $\binom{k}{2} - E$ edges that are not present in H , we can again use (10.2.11) and (10.2.12) to bound the probability of edges being present or not being present between vertices in V' . Therefore, we can bound the probability that H is not found similarly to (10.2.14) as

$$\begin{aligned} \mathbb{P}(H \text{ not in the partitions}) &\leq \mathbb{P}(H \text{ not in the disjoint partitions}) \\ &\leq \exp\left(-p_-^{k^2}(1 - p_+)^{k^2} s_{\text{dis}}\right). \end{aligned}$$

Because $s_{\text{dis}} > n^\beta$ for some $\beta > 0$, this term tends to zero exponentially. The running time of the partial search can be bounded similarly to (10.2.17) as

$$s \binom{k}{2} \leq sk^2 \leq nk, \quad (10.2.24)$$

where we have used that $s \leq n/k$. □

10.3 Experimental results

Figure 10.1 shows the fraction of times Algorithm 1 successfully finds a cycle of size k in a rank-1 inhomogeneous random graph on 10^7 vertices. Even though for large n Algorithm 1 should find an instance of a cycle of size k in Step 7 of the algorithm with high probability, we see that Algorithm 1 never succeeds in finding one. This is because of the finite-size effects discussed in Section 10.1.2.

Figure 10.2a also plots the fraction of times Algorithm 2 succeeds to find a cycle. We set the parameter $s = 10000$ so that the algorithm fails if the algorithm does not succeed to detect motif H after executing step 13 of Algorithm 2 10000 times. Because s gives the number of attempts to find H , increasing s may increase the success probability of Algorithm 2 at the cost of a higher running time. However, in Figure 10.2b for small values of k , the mean number of times Step 13 is executed when the algorithm succeeds is much lower than 10000, so that increasing s in this experiment probably only has a small effect on the success probability. Algorithm 2 clearly outperforms Algorithm 1 in terms of its success probability. Figure 10.2b also shows that the number of attempts needed to detect a cycle of length k is small for $k \leq 6$. For larger values of k the number of attempts increases. This can again be ascribed to the finite-size effects that cause the set V' to be small, so that large motifs may not be present on vertices in set V' . We also plot the success probability when using different values of the functions f_1 and f_2 . When only the lower bound f_1 on the vertex degrees is used, as in [85], the success probability of the algorithm decreases. This is because the set V' now contains many high-degree vertices that are much more likely to form cliques than cycles or other connected subgraphs on k vertices. This makes $f_2 = \infty$ a very efficient bound for detecting cliques [85]. However, Figure 10.2b shows that more checks are needed before a cycle is detected, and in some cases the cycle is not detected at all. Thus, $f_2 = \infty$ is not efficient for detecting other induced subgraphs than cliques.

Setting $f_1 = 0$ and $f_2 = \infty$ is also less efficient, as Figure 10.2a shows. In this situation, the number of attempts needed to find a cycle of length k is larger than for Algorithm 2 for $k \leq 6$.

10.3.1 Real network data

We now check Algorithm 2 on four real-world networks with power-law degrees: a Wikipedia communication network [137], the Gowalla social network [137], the Baidu online encyclopedia [165] and the Internet on the autonomous systems level [137]. Table 10.1 presents several statistics of these scale-free data sets. Figure 10.3 shows the fraction of runs where Algorithm 2 finds a cycle as an induced subgraph. For the

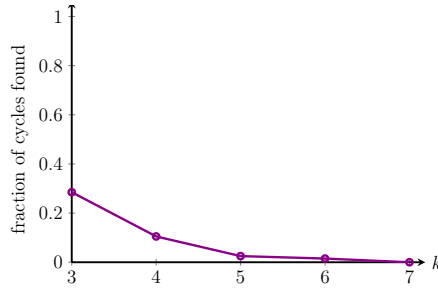


Figure 10.1: The fraction of times Algorithm 1 succeeds to find a cycle of length k on an inhomogeneous random graph with $n = 10^7$, $\tau = 5/2$, averaged over 500 network samples with $f_1 = 1/\log(n)$ and $f_2 = 0.9$.

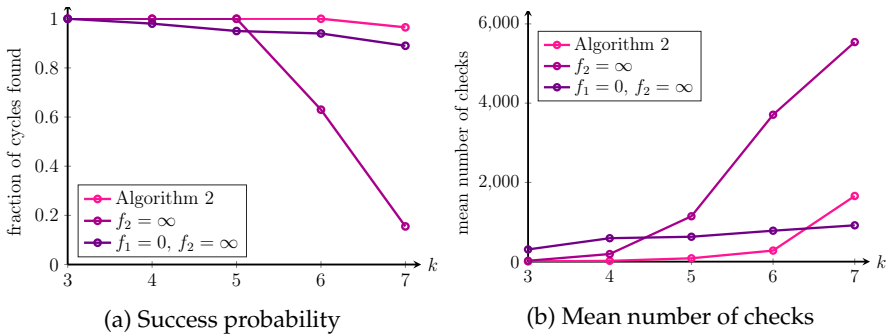


Figure 10.2: Results of Algorithm 2 on an inhomogeneous random graph with $n = 10^7$, $\tau = 5/2$ for detecting cycles of length k . The parameters are chosen as $s = 10000$, $f_1 = 1/\log(n)$, $f_2 = 0.9$. The values are averaged over 500 generated networks.

Wikipedia social network in Figure 10.3a, Algorithm 2 is more efficient than looking for cycles among all vertices in the network. For the Baidu online encyclopedia in Figure 10.3c however, Algorithm 2 performs much worse than looking for cycles among all possible vertices. In the other two network data sets in Figures 10.3b and 10.3d the performance on the reduced vertex set and the original vertex set is almost the same. Figure 10.4 shows that in general, Algorithm 2 indeed seems to finish in fewer steps than when using the full vertex set. However, as Figure 10.4c shows, for larger values of k the algorithm fails almost always.

These results show that while Algorithm 2 is efficient on rank-1 inhomogeneous random graphs, it may not always be efficient on real-world data sets. This is not surprising, because there is no reason why the vertices of degrees proportional to \sqrt{n} should behave like an Erdős-Rényi random graph in real-world networks, like in the inhomogeneous random graph. We therefore investigate whether selecting vertices with degrees in $I_n = [(\mu n)^\gamma / \log(n), (\mu n)^\gamma]$ for some other value of γ in Algorithm 2 leads to a better performance. Figure 10.3 and 10.4 show for every data set one particular value of γ that works well. For the Gowalla, Wikipedia and Autonomous systems network, this leads to a faster algorithm to detect cycles. Only

	n	E	τ
Wikipedia	2,394,385	5,021,410	2.46
Gowalla	196,591	950,327	2.65
Baidu	2,141,300	17,794,839	2.29
AS-Skitter	1,696,415	11,095,298	2.35

Table 10.1: Statistics of the data sets: the number of vertices n , the number of edges E , and the power-law exponent τ fitted by the method of [66].

for the Baidu network other values of γ do not improve upon randomly selecting from all vertices. This indicates that for most networks, cycles do appear mostly on vertices with degrees of specific orders of magnitude, making it possible to sample these cycles faster. Unfortunately, these orders of magnitude may be different for different networks. Across all four networks, the best value of γ seems to be smaller than the value of $1/2$ that is optimal for the rank-1 inhomogeneous random graph.

10.4 Conclusion

We have presented an algorithm which solves the induced subgraph problem on inhomogeneous random graphs with infinite variance power-law degrees in time nk with high probability as n grows large. This algorithm is based on the observation that for fixed k , any subgraph is present on k vertices with degrees slightly smaller than $\sqrt{\mu n}$ with positive probability. Therefore, the algorithm first selects vertices with those degrees, and then uses a random search method to look for the induced subgraph on those vertices.

We show that this algorithm performs well on simulations of rank-1 inhomogeneous random graphs. Its performance on real-world data sets varies for different data sets. This indicates that the degrees that contain the most induced subgraphs of size k in real-world networks may not be close to \sqrt{n} . We then show that on these data sets, it may be more efficient to find induced subgraphs on degrees proportional to n^γ for some other value of γ . The value of γ may be different for different networks.

Our algorithm exploits that induced subgraphs are likely formed among $\sqrt{\mu n}$ -degree vertices. However, we have shown in Chapters 7 and 8 that certain subgraphs may occur more frequently on vertices of other degrees. For example, star-shaped subgraphs on k vertices appear more often on one vertex with degree much higher than $\sqrt{\mu n}$ corresponding to the middle vertex of the star, and $k - 1$ lower-degree vertices corresponding to the leafs of the star. An interesting open question is whether there exist better degree-selection steps for specific subgraphs than the one used in Algorithms 1 and 2.

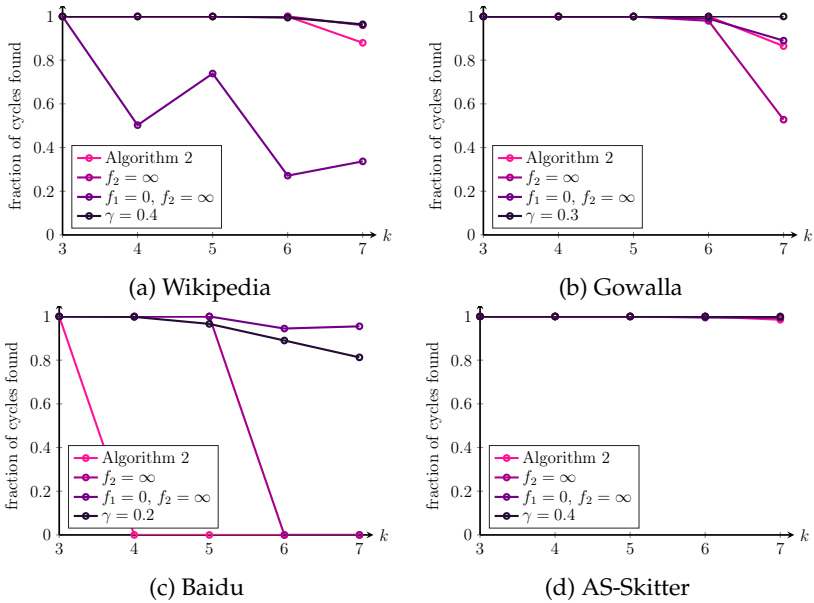


Figure 10.3: The fraction of times Algorithm 2 succeeds to find a cycle of length k on four network data sets, using $s = 10000$, $f_1 = 1/\log(n)$, $f_2 = 0.9$. The black line uses Algorithm 2 on vertices of degrees in $I_n = [(\mu n)^\gamma / \log(n), (\mu n)^\gamma]$. The values are averaged over 500 runs.

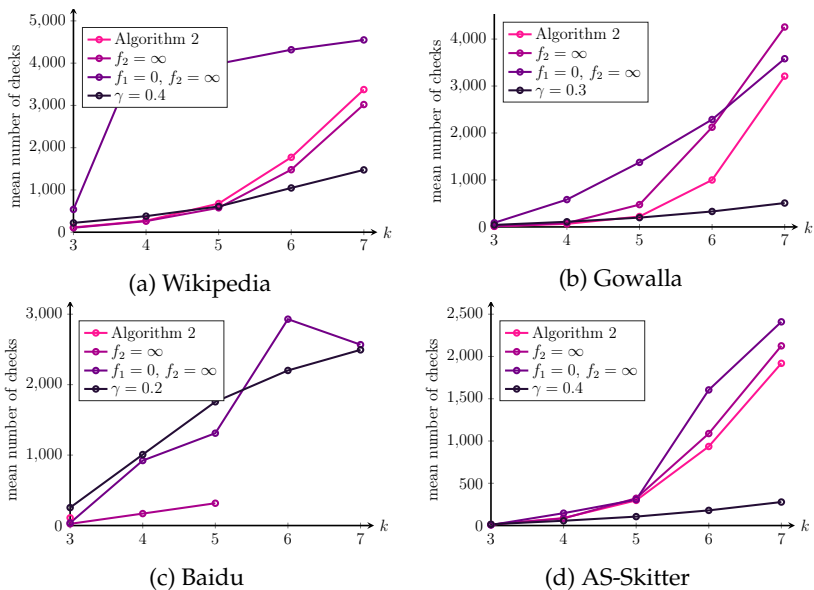


Figure 10.4: The number of times Step 12 of Algorithm 2 is invoked when the algorithm does not fail on four network data sets for detecting cycles of length k , using $s = 10000$, $f_1 = 1/\log(n)$, $f_2 = 0.9$. The black line uses Algorithm 2 on vertices of degrees in $[(\mu n)^\gamma / \log(n), (\mu n)^\gamma]$. The values are averaged over 500 runs.

Part III

Networks with community structure

11 Hierarchical configuration models

Based on:
Hierarchical configuration model
R. van der Hofstad, J.S.H. van Leeuwen and C. Stegehuis
Internet Mathematics, (2017)

In this chapter, we introduce a class of random graphs with a community structure, which we call the *hierarchical configuration model* (HCM). On the inter-community level, the graph is a configuration model, and on the intra-community level, every vertex in the configuration model is replaced by a community: i.e., a small graph. These communities may have any shape, as long as they are connected. For these hierarchical graphs, we find the size of the largest component, the degree distribution, the assortativity and the clustering coefficient. Furthermore, we determine the conditions under which a giant percolation cluster exists, and find its size.

11.1 Introduction and model

A characteristic feature of many real-world complex networks is that the degree distribution obeys a power law. A popular model for such power-law networks is the configuration model described in Section 1.1.1. A major shortcoming of this model, however, is that it is locally tree-like – it contains only a few short cycles and the graph next to most vertices is a tree – while a prominent feature of complex networks is that they often have a community structure [159].

Communities are relatively densely connected and contain relatively many short cycles. Since the configuration model contains only few short cycles, it cannot accurately model networks with community structure. One possibility to add community structure to random graphs is by adding so-called households [13, 14, 68, 206]. In this line of work, on the macroscopic level, the graph is initially a configuration model in which each vertex of the graph can be replaced by a complete graph (referred to as household). Vertices in a household have links to all other household members, which creates a community structure. These household models allow to study networks with a prescribed degree distribution and a tunable clustering coefficient, because the clustering coefficient can be manipulated by the household structure. Hence, the focus in [13, 14, 68, 206] is on locally incorporating short cycles to explain clustering at the global network level. In a similar spirit, a class of random graphs was introduced in [157] in the form of a random network that only contains random edges and triangles. Each vertex is assigned the number of triangles it is in. The triangles are formed by pairing the nodes at random, and regular edges are formed

according to the statistical rules of the configuration model. The model in [157] was extended in [126] to networks with arbitrary distributions of subgraphs.

In this chapter, we introduce the hierarchical configuration model (HCM), a random graph model that can describe networks with an arbitrary community structure. Like in these previous works, our goal is to develop a more realistic yet tractable random network model, by creating conditions under which the tree-like structure is violated within the communities, but remains to hold at a higher network level – the network of communities in our case. There are, however, considerable differences with these earlier works. The model in [126, 157] departs from a specification of all possible subgraphs or motifs, which is the triangle in [157] and all possible subgraphs in [126]. The network is then created by specifying the number of subgraphs attached to each vertex and then sampling randomly from the set of compatible networks. A community can thus exist of many subgraphs, think of a large cluster of triangles, which makes the framework in [126, 157] harder to fit on real-world networks. In fact, in [126] the appropriate selection of subgraphs and their frequencies for practical purposes is mentioned as a challenging open problem. The approaches in [13, 14, 68, 157, 206] are geared towards increasing clustering and fitting a global clustering coefficient, but are less suitable to directly describe community structure. Like [68, 206] we construct a random graph model that at the higher level is a tree-like configuration model, and at the lower level contains subgraphs, but these subgraphs do not need to be complete graphs. Large real-world communities are relatively dense, but not necessarily completely connected. We thus generalize the setting of [14, 68, 206] to arbitrary community structures, to account for heterogeneity in size and internal connectivity.

This generalization makes it possible to apply the HCM to real-world data sets. When the community structure of a real-world network is detected by an algorithm, the HCM is able to produce random graphs that have a similar community structure. Furthermore, due to the general community structure of the HCM, several existing random graph models turn out to be special cases. The advantage of the HCM is that it is quite flexible in its *local* structure, yet it is still analytically tractable due to its mesoscopic locally tree-like structure. In Chapters 12 and 13 we study how this model fits real-world networks, the conclusion being that our model fits surprisingly well. This is an important step to come to more realistic random graph models for real-world networks.

This chapter is organized as follows. In Section 11.1.1 we define the HCM. Section 11.2 presents several analytical results for the HCM, including the condition for a giant component to emerge, the degree distribution and the clustering coefficient. In Section 11.3 we study bond percolation on the HCM. Section 11.4 describes examples of graph models in the literature that fit into our general framework. Then we show in Section 11.5 how some stylized community structures affect percolation. Finally, we present some conclusions in Section 11.6.

11.1.1 Hierarchical configuration model

We now describe the HCM in more detail. Consider a random graph G with n communities $(H_i)_{i \in [n]}$. A community H is represented by $H = (F, \mathbf{d})$, where $F =$

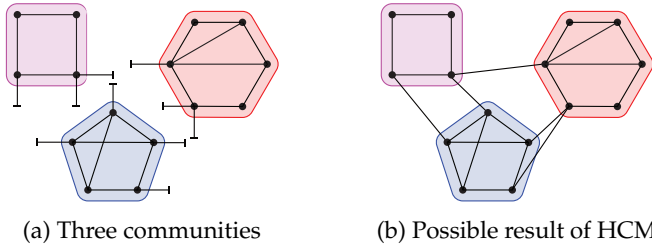


Figure 11.1: Illustration of HCM

(V_F, E_F) is a simple, connected graph and $\mathbf{d} = (d_v^{(b)})_{v \in V_F}$, where $d_v^{(b)}$ is the number of edges from $v \in V_F$ to other communities. Thus \mathbf{d} describes the degrees between the communities. We call $d_v^{(b)}$ the *inter-community degree* of a vertex v in community F . A vertex inside a community also has an *intra-community degree* $d_v^{(c)}$: the number of edges from that vertex to other vertices in the same community. The sum of the intra-community and the inter-community degrees of a vertex is the degree of the vertex, i.e., $d_v = d_v^{(b)} + d_v^{(c)}$. Let $d_H = \sum_{v \in V_F} d_v^{(b)}$ be the total inter-community degree of community H . Then the (HCM) is formed in the following way. We start with n communities. Every vertex v has $d_v^{(b)}$ half-edges attached to it, as shown in Figure 11.1a. These inter-community half-edges are paired uniformly at random. This results in a random graph G with a community structure, as shown in Figure 11.1b. On the macroscopic level, G is a configuration model with degrees $(d_{H_i})_{i \in [n]}$.

We will need to use some assumptions on the parameters of our model. For this, we start by introducing some notation. Let \mathcal{H}_n denote a uniformly chosen community in $[n] = \{1, 2, \dots, n\}$. Furthermore, denote the number of communities of type H in a graph with n communities by $n_H^{(n)}$. Then $n_H^{(n)} / n$ is the fraction of communities that are of type H . Let D_n be the inter-community degree of a uniformly chosen community, i.e., $D_n = d_{\mathcal{H}_n}$. Let the size of community i be denoted by s_i , and the size of a uniformly chosen community in $[n]$ by S_n . Then the total number of vertices in the graph is

$$N = \sum_{i=1}^n s_i = n\mathbb{E}[S_n]. \tag{11.1.1}$$

Let S and D be the limits in distribution of S_n and D_n respectively, as $n \rightarrow \infty$. We assume that the following conditions hold:

Condition 11.1 (Community regularity).

- (i) $P_n(H) = n_H^{(n)} / n \xrightarrow{\mathbb{P}} P(H)$, where $P(H)$ is a probability distribution,
- (ii) $\lim_{n \rightarrow \infty} \mathbb{E}[S_n] = \mathbb{E}[S]$,

for some random variable S with $\mathbb{E}[S] < \infty$.

Condition 11.2 (Intercommunity connectivity).

- (i) $\lim_{n \rightarrow \infty} \mathbb{E}[D_n] = \mathbb{E}[D]$,

(ii) $\mathbb{P}(D = 2) < 1$,

for some random variable D with $\mathbb{E}[D] < \infty$.

Condition 11.1(i) implies that $D_n \xrightarrow{d} D$ and $S_n \xrightarrow{d} S$, so that S and D are the asymptotic community size distribution and community inter-community degree distribution, respectively. Define

$$p_{k,s}^{(n)} = \sum_{H=(F,\mathbf{d});|F|=s,d_H=k} P_n(H), \quad (11.1.2)$$

$$p_{k,s} = \sum_{H=(F,\mathbf{d});|F|=s,d_H=k} P(H), \quad (11.1.3)$$

as the probabilities that a uniformly chosen community has size s and inter-community degree k , for finite n and $n \rightarrow \infty$, respectively. Then Condition 11.1 implies that $p_{k,s}^{(n)} \rightarrow p_{k,s}$ for every (k, s) .

We can think of $P_n(H)$ as the probability that a uniformly chosen community has a certain shape. In a data set we can approximate $P_n(H)$ and use the HCM in the following way. Suppose a community detection algorithm gives the empirical distribution of the community shapes $P_n(H)$. Now we construct a random graph in the way that was described above. The probability that a certain community is of shape H is $P_n(H)$. We condition on the total inter-community degree to be even so that edges between communities can be formed as in a configuration model. This results in a graph with roughly the same degree sequence as the original graph. Additionally, the community structure in the random graph is the same as in the original graph. This construction preserves more of the microscopic features of the original graph than a standard configuration model with the same degree sequence as the original graph. It also shows the necessity of extending the work of [68, 206] to go beyond the assumption that communities are complete graphs, because communities in real-world networks can be non-complete. Using this construction, the HCM can match the community structure in many complex networks. We will study HCM with real-world communities as inputs in Chapters 12 and 13.

11.2 Model properties

For a connected component of G , we can either count the number of communities in the component, or the number of vertices in it. We denote the number of communities in a connected component \mathcal{C} by $v^{(\text{H})}(\mathcal{C})$, and the number of communities with inter-community degree k in the component by $v_k^{(\text{H})}(\mathcal{C})$. The number of vertices in component \mathcal{C} is denoted by $v(\mathcal{C})$. Let $\mathcal{C}_{(1)}$ and $\mathcal{C}_{(2)}$ be the largest and second largest components of G , respectively, so that

$$v(\mathcal{C}_{(1)}) = \max_{u \in [N]} v(\mathcal{C}(u)), \quad (11.2.1)$$

where $\mathcal{C}(u)$ denotes the component of vertex u . Furthermore, define v_D as

$$v_D = \frac{\mathbb{E}[D(D-1)]}{\mathbb{E}[D]}, \quad (11.2.2)$$

where D is the asymptotic community degree of Condition 11.2. Let $p_k = \mathbb{P}(D = k)$ and let $g(x) = \sum_k p_k x^k$ be the probability generating function of D , and $g'(x) = \sum_k k p_k x^{k-1}$ its derivative.

11.2.1 Giant component

In the standard configuration model, a giant component exists w.h.p. if $\nu_D > 1$ [117, 154, 155]. In the HCM a similar statement holds:

Theorem 11.1. *Let G be a HCM satisfying Conditions 11.1 and 11.2. Then,*

(i) If $\nu_D > 1$,

$$\frac{v(\mathcal{C}_{(1)})}{N} \xrightarrow{\mathbb{P}} \frac{\sum_{k,s} s p_{k,s} (1 - \xi^k)}{\mathbb{E}[S]} > 0, \quad (11.2.3)$$

where ξ is the unique solution in $[0, 1)$ of $g'(\xi) = \xi \mathbb{E}[D]$. Furthermore, $v(\mathcal{C}_{(2)})/N \xrightarrow{\mathbb{P}} 0$.

(ii) If $\nu_D \leq 1$, then $v(\mathcal{C}_{(1)})/N \xrightarrow{\mathbb{P}} 0$.

Proof. Suppose $\nu_D > 1$. By [107, Theorem 4.1], if Condition 11.2 holds, $D_n \xrightarrow{\mathbb{P}} D$ and $\nu_D > 1$ in a standard configuration model, then w.h.p. there will be one component with a positive fraction of the vertices as $n \rightarrow \infty$. Furthermore, the number of vertices in the largest component in a standard configuration model $v(\mathcal{C}_{(1)}^{\text{CM}})$ and the number of vertices of degree k in its largest connected component, $v_k(\mathcal{C}_{(1)}^{\text{CM}})$ satisfy

$$v(\mathcal{C}_{(1)}^{\text{CM}})/n \xrightarrow{\mathbb{P}} 1 - g(\xi) > 0, \quad (11.2.4)$$

$$v_k(\mathcal{C}_{(1)}^{\text{CM}})/n \xrightarrow{\mathbb{P}} p_k (1 - \xi^k). \quad (11.2.5)$$

If $\nu_D \leq 1$, then $v(\mathcal{C}_{(1)}^{\text{CM}})/n \xrightarrow{\mathbb{P}} 0$. Therefore, if Conditions 11.1 and 11.2 hold and $\nu_D > 1$ in the HCM, then there is a component with a positive fraction of the communities as $n \rightarrow \infty$. Hence, we need to prove that the largest hierarchical component is indeed a large component with size given by (11.2.3) if $\nu_D > 1$, and that a small hierarchical component is also a small component of G .

Let $\mathcal{C}_{(1)}^{(\text{CM})}$ denote the largest hierarchical component of G : the component with the largest number of communities. We denote the number of communities in the largest hierarchical component with inter-community degree k and size s by $v_{k,s}^{(\text{H})}(\mathcal{C}_{(1)}^{(\text{CM})})$. Since G is a configuration model on the community level, (11.2.4) and (11.2.5) apply on the community level. Furthermore, given a community in the largest hierarchical component of inter-community degree k , its size is *independent* of being in the largest hierarchical component. Moreover, $\sum_k s v_{k,s}^{(\text{H})}(\mathcal{C}_{(1)}^{(\text{CM})})/n \leq \sum_k s p_{k,s}^{(n)}$. Therefore, by Condition 11.1, the fraction of vertices in the largest hierarchical component satisfies

$$\frac{v(\mathcal{C}_{(1)})}{N} = \frac{\sum_{i \in \mathcal{C}_{(1)}^{\text{H}}} s_i}{\sum_i s_i} = \frac{\sum_{k,s} n^{-1} s v_{k,s}^{(\text{H})}(\mathcal{C}_{(1)}^{(\text{CM})})}{n^{-1} \sum_i s_i} \xrightarrow{\mathbb{P}} \frac{\sum_{k,s} s p_{k,s} (1 - \xi^k)}{\mathbb{E}[S]} > 0. \quad (11.2.6)$$

The last inequality follows from Condition 11.1(ii) and the fact that $\xi \in [0, 1)$ and $s \geq 1$. Now we need to prove that the largest hierarchical component indeed is the largest component of G . We show that a hierarchical component of size $o(n)$ is a component of size $o_{\mathbb{P}}(N)$. Take a hierarchical component \mathcal{C} which is not the largest hierarchical component, so that it is of size $o(n)$. Then,

$$\begin{aligned} \frac{v(\mathcal{C})}{N} &= \frac{n^{-1} \sum_{k,s} s v_{k,s}^{(H)}(\mathcal{C})}{\mathbb{E}[S_n]} = \frac{n^{-1} \sum_{s=1}^K \sum_k s v_{k,s}^{(H)}(\mathcal{C})}{\mathbb{E}[S_n]} + \frac{n^{-1} \sum_{s>K} \sum_k s v_{k,s}^{(H)}(\mathcal{C})}{\mathbb{E}[S_n]} \\ &\leq K \frac{n^{-1} \sum_{s=1}^K \sum_k v_{k,s}^{(H)}(\mathcal{C})}{\mathbb{E}[S_n]} + \frac{\mathbb{E}[S_n \mathbb{1}_{\{S_n > K\}}]}{\mathbb{E}[S_n]} \leq K \frac{n^{-1} v^{(H)}(\mathcal{C})}{\mathbb{E}[S_n]} + \frac{\mathbb{E}[S_n \mathbb{1}_{\{S_n > K\}}]}{\mathbb{E}[S_n]}. \end{aligned} \quad (11.2.7)$$

First we take the limit for $n \rightarrow \infty$, and then we let $K \rightarrow \infty$. By [117], $v^{(H)}(\mathcal{C})/n \xrightarrow{\mathbb{P}} 0$, hence the first term tends to zero as $n \rightarrow \infty$. Furthermore, $\mathbb{E}[S_n \mathbb{1}_{\{S_n > K\}}] \rightarrow \mathbb{E}[S \mathbb{1}_{\{S > K\}}]$ as $n \rightarrow \infty$ by Condition 11.1. By Condition 11.1(ii), this tends to zero as $K \rightarrow \infty$. Thus, $v(\mathcal{C})/N \xrightarrow{\mathbb{P}} 0$. Since (11.2.7) is uniform in \mathcal{C} , this proves that the largest hierarchical component is indeed the largest component of G . This also proves part (ii), since by [107, Theorem 4.1], if $\nu_D \leq 1$, $v^{(H)}(\mathcal{C}_{(1)}^{(CM)}) = o_{\mathbb{P}}(n)$, so that $v(\mathcal{C}_{(1)}) = o_{\mathbb{P}}(N)$. Similarly, for $\nu_D > 1$, $v^{(H)}(\mathcal{C}_{(2)}^{(CM)}) = o_{\mathbb{P}}(n)$ by [107, Theorem 4.1]. Thus we may conclude that $v(\mathcal{C}_{(2)})/N \xrightarrow{\mathbb{P}} 0$. \square

We conclude that if Conditions 11.1 and 11.2 hold and $\nu_D > 1$, then a giant component exists in the HCM. Equation (11.2.3) gives the fraction of vertices in the largest component. The fraction of vertices in the giant component may be different from the fraction of communities in the giant hierarchical component. If the sizes and the inter-community degrees of the communities are independent, then the fraction of vertices in the largest component is equal to the fraction of communities in the largest hierarchical component.

Corollary 11.1. *Suppose that in the HCM G satisfying Conditions 11.1 and 11.2, the size of the communities and the inter-community degrees of the communities are independent. Then, if $\nu_D > 1$,*

$$\frac{v(\mathcal{C}_{(1)})}{N} \xrightarrow{\mathbb{P}} 1 - g(\xi), \quad (11.2.8)$$

$$\frac{v^{(H)}(\mathcal{C}_{(1)})}{n} \xrightarrow{\mathbb{P}} 1 - g(\xi), \quad (11.2.9)$$

where ξ is the unique solution in $[0, 1)$ of $g'(\xi) = \xi \mathbb{E}[D]$. Hence the fraction of vertices in the largest component is equal to the fraction of communities in the largest hierarchical component. If the size and the inter-community degrees are dependent, then this does not have to be true.

Proof. The equality in (11.2.9) is given by [107, Theorem 4.1]. The equality in (11.2.8) follows by substituting $p_{k,s} = p_k p_s$ in (11.2.3), so that

$$\frac{v(\mathcal{C}_{(1)})}{N} \xrightarrow{\mathbb{P}} \frac{\sum_s s p_s \sum_k p_k (1 - \xi^k)}{\mathbb{E}[S]} = \frac{\mathbb{E}[S](1 - \sum_k p_k \xi^k)}{\mathbb{E}[S]} = 1 - g(\xi). \quad (11.2.10)$$

To show that (11.2.8) may not hold when the inter-community degrees and the sizes are dependent, consider the HCM with

$$p_{k,s} = \begin{cases} \frac{1}{3} & \text{if } (k,s) = (3,10), \\ \frac{2}{3} & \text{if } (k,s) = (1,1). \end{cases} \quad (11.2.11)$$

Since $\nu_D = \frac{6}{5} > 1$, a giant component exists w.h.p. Furthermore, ζ solves

$$\frac{2}{3} + \zeta^2 = \frac{5}{3}\zeta, \quad (11.2.12)$$

which has $\frac{2}{3}$ as its only solution in $[0, 1)$. Therefore, the fraction of communities in the largest component is given by $1 - g(\frac{2}{3}) = \frac{37}{81}$. To find the fraction of vertices in the largest component, we use (11.2.3), which gives

$$\frac{1}{4}(\frac{2}{3}(1 - \frac{1}{2}) + 10\frac{1}{3}(1 - (\frac{1}{2})^3)) = \frac{13}{16} > \frac{37}{81}. \quad (11.2.13)$$

Thus, the fraction of vertices in the largest component is larger than the fraction of communities in the largest component. \square

If there is a difference between the fraction of communities and the fraction of vertices in the largest component, then this difference is caused by the dependence of the sizes and the inter-community degrees of the communities. A community with a large inter-community degree has a higher probability of being in the largest hierarchical component than a community with a small inter-community degree. In the example in the proof of Corollary 11.1, the communities with large inter-community degrees are large communities. This causes the fraction of vertices in the largest component to be larger than the fraction of communities in the largest hierarchical component.

11.2.2 Degree distribution

In the HCM, the macroscopic configuration model has a fixed degree sequence. The degree distribution of G depends on the sizes and shapes of the communities. Let $n_k^{(H)}$ denote the number of vertices in community H with the sum of their intra-community degree and inter-community degree equal to k . Then the degree distribution of the total graph G is described in Proposition 11.1:

Proposition 11.1. *Let G be a HCM such that Conditions 11.1 and 11.2 hold. The asymptotic probability \hat{p}_k that a randomly chosen vertex inside G has degree k satisfies*

$$\hat{p}_k = \frac{\sum_H P(H)n_k^{(H)}}{\mathbb{E}[S]}, \quad (11.2.14)$$

as $n \rightarrow \infty$.

Proof. Consider a HCM G on n communities. Let $n_H^{(n)}$ be the number of communities in G of type H . The total number of vertices of degree k is the sum of the

number of degree k vertices inside all communities, hence it equals $\sum_H n_H^{(n)} n_k^{(H)}$. Furthermore, $P_n(H) n_k^{(H)} \leq P_n(H) s_H$, so that $\lim_{n \rightarrow \infty} \sum_H P_n(N) n_k^{(H)} = \sum_H P(H) n_k^{(H)}$ by Condition 11.1. This gives

$$\hat{p}_k^{(n)} = \frac{n^{-1} \sum_H n_H n_k^{(H)}}{n^{-1} N} = \frac{\sum_H P_n(H) n_k^{(H)}}{\mathbb{E}[S_n]} \xrightarrow{\mathbb{P}} \hat{p}_k, \quad (11.2.15)$$

as $n \rightarrow \infty$. □

11.2.3 The probability of obtaining a simple graph

In the standard configuration model, the probability of obtaining a simple graph converges to $e^{-\nu/2 - \nu^2/4}$ under the condition that $\mathbb{E}[D^2] < \infty$ [106]. In the HCM, the probability of obtaining a simple graph is largely dependent on the shapes of the communities. Since we have assumed that the communities are simple, only the inter-community edges can create self-loops and multiple edges.

Suppose that each vertex in a community has at most one half-edge to other communities, i.e., $d_v^{(b)} \in \{0, 1\}$. A double edge in the macroscopic configuration model corresponds to a community where two vertices have an edge to the same other community. Since $d_v^{(b)} \in \{0, 1\}$, a double edge in the macroscopic configuration model cannot correspond to a double edge in the HCM. A self-loop in the macroscopic configuration model corresponds to an edge from one vertex v inside a community to another vertex w inside the same community. This self-loop in the macroscopic configuration model corresponds to a double edge in the HCM if an edge from v to w was already present in the community. Thus, when $d_v^{(b)} \in \{0, 1\}$ the probability that the macroscopic configuration model is simple is lower bounded by the probability that no self-loops exist in the macroscopic configuration model,

$$\liminf_{n \rightarrow \infty} \mathbb{P}(G_n \text{ simple}) \geq e^{-\nu_D/2}, \quad (11.2.16)$$

which is larger than the corresponding probability for the configuration model. In the case of complete graph communities, every self-loop of the macroscopic configuration model corresponds to a double edge in the HCM. Therefore, equality holds when all communities are complete graphs.

11.2.4 Assortativity

The assortativity of a graph $G = (V, E)$ can be interpreted as the correlation between the degrees at the end of a randomly chosen edge [158] and is given by

$$r(G) = \frac{2 \sum_{\{i,j\} \in E} d_i d_j - \frac{1}{2L} (\sum_i d_i^2)^2}{\sum_i d_i^3 - \frac{1}{2L} (\sum_i d_i^2)^2}. \quad (11.2.17)$$

Positive assortativity indicates that vertices of high degree are connected to other vertices of high degree, and negative assortativity indicates that high degree vertices are typically connected to vertices of low degree. Assortativity is a frequently used

network statistic, despite its dependence on the network size [141], and can also be interpreted in terms of the number of short walks on the network [149].

The assortativity of the HCM can be computed analytically using (11.2.17). We denote the degree (inter- plus extra-community degree) of a randomly chosen vertex among the N vertices of the graph by \hat{D}_N . Furthermore, we denote by $D_N^{(b)}$ the inter-community degree of a randomly chosen vertex. For a given community H , let $Q(H)$ denote

$$Q(H) = \sum_{\{i,j\} \in E_H} d_i d_j, \tag{11.2.18}$$

and let Q_n be the value of $Q(H)$ for a randomly chosen community. Then we can prove the following result for the expected assortativity:

Proposition 11.2. *Let G be a HCM such that Conditions 11.1 and 11.2 hold. Then the expected assortativity of G satisfies*

$$\mathbb{E}[r(G)] = \frac{2\mathbb{E}[\hat{D}_N D_N^{(b)}]^2 \mathbb{E}[S_n] (\mathbb{E}[D_n] - \frac{1}{n})^{-1} + 2\mathbb{E}[Q_n] \mathbb{E}[S_n]^{-1} - \mathbb{E}[\hat{D}_N^2]^2 \mathbb{E}[\hat{D}_N]^{-1}}{\mathbb{E}[\hat{D}_N^3] - \mathbb{E}[\hat{D}_N^2]^2 \mathbb{E}[\hat{D}_N]^{-1}}. \tag{11.2.19}$$

Proof. We rewrite (11.2.17) as

$$\begin{aligned} r(G) &= \frac{2 \sum_{\{i,j\} \in E} d_i d_j - \frac{1}{2L} (\sum_i d_i^2)^2}{\sum_i d_i^3 - \frac{1}{2L} (\sum_i d_i^2)^2} = \frac{\frac{2}{N} \sum_{\{i,j\} \in E} d_i d_j - \frac{(\frac{1}{N} \sum_i d_i^2)^2}{\frac{1}{N} \sum_i d_i}}{\frac{1}{N} \sum_i d_i^3 - \frac{(\frac{1}{N} \sum_i d_i^2)^2}{\frac{1}{N} \sum_i d_i}} \\ &= \frac{\frac{2}{N} \sum_{\{i,j\} \in E} d_i d_j - \frac{\mathbb{E}[\hat{D}_N^2]^2}{\mathbb{E}[\hat{D}_N]}}{\mathbb{E}[\hat{D}_N^3] - \frac{\mathbb{E}[\hat{D}_N^2]^2}{\mathbb{E}[\hat{D}_N]}}. \end{aligned} \tag{11.2.20}$$

Therefore, the only term of assortativity that depends on the community structure is the first term in the numerator. The edges of a HCM can be split into two sets: the edges that are entirely inside a community, and the edges that are between two different communities, denoted by E_c and E_b respectively. The edges inside communities are fixed given the community shape. Therefore the contribution of the intra-community edges to the first term in the numerator can be written as

$$\frac{1}{N} \sum_{\{i,j\} \in E_c} d_i d_j = \frac{1}{N} \sum_{k=1}^n \sum_{\{i,j\} \in E_{H_k}} d_i d_j = \frac{1}{n \mathbb{E}[S_n]} \sum_{k=1}^n Q(H_k) = \frac{\mathbb{E}[Q_n]}{\mathbb{E}[S_n]}. \tag{11.2.21}$$

Let $D_N^{(b)}$ denote the inter-community degree of a randomly chosen vertex inside a community. The HCM consists of $n \mathbb{E}[D_n]$ half-edges. The probability that a specific half-edge will be paired with another specific half-edge equals $1/(n \mathbb{E}[D_n] - 1)$, since the half-edges are paired at random. Then the expected contribution of the inter-community edges can be written as

$$\mathbb{E}\left[\frac{1}{N} \sum_{\{i,j\} \in E_b} d_i d_j\right] = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^{d_i^{(b)}} \sum_{l=1}^{d_j^{(b)}} \frac{d_i d_j}{n \mathbb{E}[D_n] - 1} = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N \frac{d_i d_i^{(b)} d_j d_j^{(b)}}{n \mathbb{E}[D_n] - 1}$$

$$\begin{aligned}
&= \frac{1}{N(n\mathbb{E}[D_n] - 1)} \left(\sum_{i=1}^N d_i d_i^{(b)} \right)^2 = \frac{\mathbb{E}[S_n]}{\mathbb{E}[D_n] - \frac{1}{n}} \left(\frac{1}{N} \sum_{i=1}^N d_i d_i^{(b)} \right)^2 \\
&= \frac{\mathbb{E}[S_n] \mathbb{E}[\hat{D}_N D_N^{(b)}]^2}{\mathbb{E}[D_n] - \frac{1}{n}}. \tag{11.2.22}
\end{aligned}$$

Combining (11.2.20), (11.2.21) and (11.2.22) proves the proposition. \square

11.2.5 Clustering coefficient

We now investigate the global clustering coefficient defined as

$$C = \frac{3 \times \text{number of triangles}}{\text{number of connected triples}}. \tag{11.2.23}$$

A connected triple is a vertex with edges to two different other vertices. In the standard configuration model, the clustering coefficient tends to zero when $\mathbb{E}[D^2] < \infty$ [161]. Thus, in the HCM, we expect that the clustering is entirely caused by triangles inside communities.

Another measure of clustering is the local clustering coefficient for vertices of degree k . This coefficient can be interpreted as the fraction of neighbors of degree k vertices that are directly connected and is defined as

$$c(k) = \frac{\text{number of pairs of connected neighbors of degree } k \text{ vertices}}{k(k-1)/2 \times \text{number of degree } k \text{ vertices}}. \tag{11.2.24}$$

As in Section 11.2.2, let $n_k^{(H)}$ denote the number of vertices in community H with degree equal to k . Furthermore, let $P_v^{(H)}$ denote the number of pairs of neighbors of a vertex $v \in V_H$ within community H that are also neighbors of each other. We denote the clustering coefficient of community H by $\tilde{c}(H)$. Every vertex v in community H has $d_v^{(c)}(d_v^{(c)} - 1)/2$ pairs of neighbors inside H . Hence, the total number of connected triples inside the community is given by $\sum_{v \in V_H} d_v^{(c)}(d_v^{(c)} - 1)/2$. Then, by (11.2.23),

$$\tilde{c}(H) = \frac{2 \sum_{v \in V_H} P_v^{(H)}}{\sum_{v \in V_H} d_v^{(c)}(d_v^{(c)} - 1)}. \tag{11.2.25}$$

Proposition 11.3 states that the clustering coefficient of the HCM can be written as a combination of the clustering coefficients inside communities. Let \hat{D} denote the asymptotic degree as in Proposition 11.1.

Proposition 11.3. *Let G be a HCM satisfying Conditions 11.1 and 11.2, $\lim_{n \rightarrow \infty} \mathbb{E}[D_n^2] = \mathbb{E}[D^2] < \infty$ and $\lim_{N \rightarrow \infty} \mathbb{E}[\hat{D}_N^2] = \mathbb{E}[\hat{D}^2] < \infty$. Then the clustering coefficient $C^{(n)}$ and average clustering coefficient for vertices of degree k , $c^{(n)}(k)$, satisfy*

$$C^{(n)} \xrightarrow{\mathbb{P}} C := \frac{2 \sum_H \sum_{v \in V_H} P(H) c(H) d_v^{(c)} (d_v^{(c)} - 1)}{\sum_H \sum_{v \in V_H} P(H) d_v (d_v - 1)}, \tag{11.2.26}$$

$$c^{(n)}(k) \xrightarrow{\mathbb{P}} c(k) := \frac{2 \sum_H \sum_{v \in V_H: d_v=k} P(H) P_v^{(H)}}{k(k-1)}. \tag{11.2.27}$$

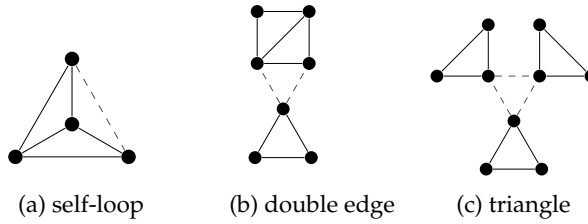


Figure 11.2: Possibilities to form triangles in the HCM that are not entirely inside communities. Edges between communities (dashed) that add clustering correspond to either a self-loop, a double edge or a triangle in the macroscopic configuration model.

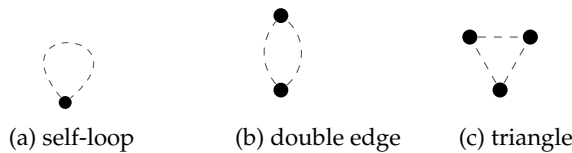


Figure 11.3: Figure 11.2 on macroscopic level. The inter-community edges that add clustering correspond to either a self-loop, a double edge or a triangle of the macroscopic configuration model.

Proof. In the HCM, the number of triples is deterministic. A vertex v with degree d_v has $d_v(d_v - 1)/2$ pairs of neighbors. Thus, the total number of connected triples in G is given by $\sum_H n_H^{(n)} \sum_{v \in V_H} d_v(d_v - 1)/2$, where $n_H^{(n)}$ is the number of type H communities.

Triangles in G can be formed in several ways. First of all, a triangle can be formed by three edges inside the same community. In this case the triangle in G is formed by a triangle in one of its communities H . Another possibility to create a triangle is shown in Figure 11.2a. The black edges show intra-community edges, and the dashed edges are formed by edges of the macroscopic configuration model. Thus, this triangle is formed by two intra-community edges, and one edge of the macroscopic configuration model. Figure 11.3a shows that this inter-community edge is a self-loop of the macroscopic configuration model. One self-loop of the macroscopic configuration model can create multiple triangles; at most $s_i - 2$. Figure 11.2b shows the case where only one edge of the triangle is an intra-community edge. Figure 11.3b shows that the two inter-community edges must form a double edge in the macroscopic configuration model. The last possibility is that all three edges of the triangle are inter-community edges as in Figure 11.2c. This corresponds to a triangle in the macroscopic configuration model (Figure 11.3c).

Hence, either the triangle was present in H already, or it corresponds to a double edge, self-loop or triangle in $\phi(G)$, where $\phi(G)$ denotes the macroscopic configuration model. Let the number of self-loops, double edges and triangles in $\phi(G)$ be denoted by $W^{(n)}$, $M^{(n)}$ and $T^{(n)}$ respectively. Denote the number of triangles entirely formed by intra-community edges of G by $T_{\text{com}}^{(n)}$. The number of triangles in G is bounded

from below by $T_{\text{com}}^{(n)}$. Using (11.2.25), we obtain that

$$3T_{\text{com}}^{(n)} = \sum_H \sum_{v \in V_H} n_H^{(n)} P_v^{(H)} = \sum_H \sum_{v \in V_H} n_H \tilde{c}(H) d_v^{(c)} (d_v^{(c)} - 1). \quad (11.2.28)$$

Since $c(H) \leq 1$, and $\lim_{N \rightarrow \infty} \mathbb{E}[\hat{D}_N^2] = \mathbb{E}[\hat{D}^2]$,

$$\begin{aligned} C^{(n)} &= \frac{3 \times \text{number of triangles in } G}{\text{number of connected triples in } G} \geq \frac{3T_{\text{com}}^{(n)}/n}{\sum_H n_H^{(n)} \sum_{v \in V_H} d_v (d_v - 1) / (2n)} \\ &\xrightarrow{\mathbb{P}} \frac{2 \sum_H \sum_{v \in V_H} P(H) \tilde{c}(H) d_v^{(c)} (d_v^{(c)} - 1)}{\sum_H \sum_{v \in V_H} P(H) d_v (d_v - 1)}. \end{aligned} \quad (11.2.29)$$

The sums in (11.2.29) are finite due to the assumptions $\mathbb{E}[D^2] < \infty$ and $\mathbb{E}[\hat{D}^2] < \infty$.

For the upper bound, we use that every self-loop on the community level adds at most $s_i - 2$ triangles, and every triangle or double edge on the community level add at most one triangle. This yields the inequality

$$\text{number of triangles } G \leq T_{\text{com}}^{(n)} + M^{(n)} + S^{(n)} + \sum_{i=1}^{W^{(n)}} (s_{\mathcal{I}_i} - 2). \quad (11.2.30)$$

Here the sum is over all communities where a self-loop is present, written as $(\mathcal{I}_i)_{i=1}^{W^{(n)}}$. If a community has multiple self-loops, then the community is counted multiple times in the sum. By [13, Theorem 5]

$$(M^{(n)} + T^{(n)}) / n \xrightarrow{\mathbb{P}} 0. \quad (11.2.31)$$

in a configuration model with $\mathbb{E}[D^2] < \infty$.

The last term in (11.2.30) satisfies

$$\begin{aligned} \frac{\sum_{i=1}^{W^{(n)}} (s_{\mathcal{I}_i} - 2)}{n} &= \frac{W^{(n)} \mathbb{E}[S_n - 2 \mid \text{self-loop}]}{n} \\ &\leq \frac{W^{(n)} \max_{i \in [n]} s_i}{n} = \frac{W^{(n)} o(n)}{n} \xrightarrow{\mathbb{P}} 0. \end{aligned} \quad (11.2.32)$$

The last equality follows because $\mathbb{E}[S_n] \rightarrow \mathbb{E}[S] < \infty$, which implies that

$$\lim_{k \rightarrow \infty} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j \in [n]} s_j \mathbb{1}\{s_j > k\} = 0, \quad (11.2.33)$$

so that $\max_i s_i = o(n)$. The convergence follows since the number of self-loops in a configuration model converges to a Poisson distribution with mean ν_D [106, Proposition 7.11], combined with $\mathbb{E}[D^2] < \infty$.

Combining (11.2.31) and (11.2.32) yields

$$C^{(n)} \leq \frac{3T_{\text{com}}^{(n)} + 3(M^{(n)} + T^{(n)}) + 3 \sum_{i=1}^{W^{(n)}} (s_{\mathcal{I}_i} - 2)}{\sum_H n_H^{(n)} \sum_{v \in V_H} d_v (d_v - 1) / 2}$$

$$\mathbb{P} \rightarrow \frac{2 \sum_H P(H) \sum_{v \in V_H} c(H) d_v^{(c)} (d_v^{(c)} - 1)}{\sum_H P(H) \sum_{v \in V_H} d_v (d_v - 1)}. \tag{11.2.34}$$

Together with (11.2.29) this proves (11.2.26).

To prove (11.2.27), a similar argument can be used. The number of connected neighbors of vertices of degree k is bounded from below by $\sum_H n_H^{(n)} \sum_{v \in V_H: d_v=k} P_v^{(H)}$, and from above by

$$\sum_H n_H^{(n)} \sum_{v \in V_H: d_v=k} P_v^{(H)} + M^{(n)} + S^{(n)} + \sum_{i=1}^{W^{(n)}} (s_{\mathcal{I}_i} - 2). \tag{11.2.35}$$

Then, dividing by $k(k-1)n\hat{p}_k^{(n)}$, where $\hat{p}_k^{(n)}$ is the probability of having a vertex of degree k , and taking the limit yields (11.2.27). The assumption that $\mathbb{E}[\hat{D}^2] < \infty$ is not necessary for this clustering coefficient, since $P_v/k(k-1) \leq 1$ for all vertices of degree k . □

11.3 Percolation

We now consider bond percolation on G , where each edge of G is removed independently with probability $1 - \pi$. We are interested in the critical percolation value and the size of the largest percolating cluster. Percolation on the configuration model was studied in [81, 115]. Here we extend these results to the HCM.

Percolating G is the same as first percolating only the intra-community edges, and then percolating the inter-community edges. For percolation inside a community, only the intra-community edges are removed with probability $1 - \pi$. The half-edges attached to a community are not percolated. Let H_π denote the subgraph of H , where each edge of H has been deleted with probability $1 - \pi$. When percolating a community, it may split into different connected components, as illustrated in Figure 11.4b. Let $g(H, v, l, \pi)$ denote the probability that the connected component of H_π containing v has inter-community degree l . If H_π is still connected, then the component containing v still has d_H outgoing edges for all $v \in V_H$. If H_π is disconnected, then this does not hold. If one of the components of H_π has an inter-community edge, each vertex in another component of H_π cannot reach that edge. Therefore, a vertex in this other component is connected to less than d_H inter-community edges.

To compute the size of the largest percolating cluster, we need the following definitions:

$$p'_k := \frac{\sum_H \sum_{v \in V_H} d_v^{(b)} P(H) g(H, v, k, \pi) / k}{\sum_H \sum_{v \in V_H} \sum_l d_v^{(b)} P(H) g(H, v, l, \pi) / l'} \tag{11.3.1}$$

$$h(z) := \sum_{k=1}^{\infty} k p'_k z^{k-1}, \tag{11.3.2}$$

$$\lambda := \sum_{k=0}^{\infty} k p'_k. \tag{11.3.3}$$

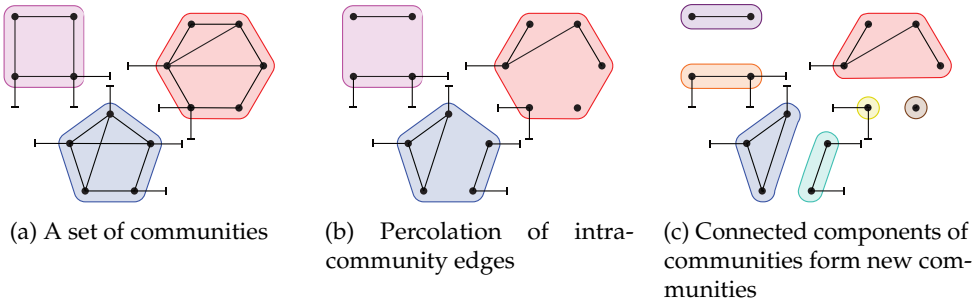


Figure 11.4: Percolating only intra-community edges first results in a HCM with different communities.

The probabilities $(p'_k)_{k \geq 0}$ can be interpreted as the asymptotic probability distribution of the inter-community degrees of the connected parts of communities after percolation inside communities (the communities of Figure 11.4c). Then $h(z)$ and λ are the derivative of the probability generating function and the mean of the inter-community degrees of the components of communities after percolation respectively.

Define D_π^* as the number of inter-community edges after entering a percolated community from a randomly chosen edge. The probability of entering at vertex v in community H , equals $P(H)d_v^{(b)}/\mathbb{E}[D]$. After entering H at vertex v , there are in expectation $\sum_{k=1}^{D_H-1} kg(H, v, k+1, \pi)$ edges to other communities (since one edge was used to enter H). Hence,

$$\mathbb{E}[D_\pi^*] = \frac{1}{\mathbb{E}[D]} \sum_H P(H) \sum_{v \in V_H} d_v^{(b)} \sum_{k=1}^{D_H-1} kg(H, v, k+1, \pi). \quad (11.3.4)$$

After percolating the inter-community edges, a fraction of π of these edges remain. Thus, after percolating all edges, when entering a community, the expected number of outgoing edges excluding the traversed edge is $\pi \mathbb{E}[D_\pi^*]$. We expect the critical value of π to satisfy $\pi \mathbb{E}[D_\pi^*] = 1$, i.e., the expected number of edges to other communities is one, after entering a community from a randomly chosen edge. The next theorem states that this is indeed the critical percolation value:

Theorem 11.2. *Assume G is a HCM satisfying Conditions 11.1 and 11.2. The critical value of the percolation parameter π_c of G is the unique solution of*

$$\pi_c = \frac{1}{\mathbb{E}[D_{\pi_c}^*]}. \quad (11.3.5)$$

Furthermore:

1. For $\pi > \pi_c$, the size of the largest component of the percolated graph satisfies

$$\frac{v(\mathcal{C}_{(1)})}{N} \xrightarrow{\mathbb{P}} \frac{1}{\mathbb{E}[S]} \sum_{k=1}^{\infty} \sum_H \sum_{v \in V_H} P(H) g(H, v, k, \pi) \left(1 - (1 - \sqrt{\pi} + \sqrt{\pi} \xi)^k \right) > 0, \quad (11.3.6)$$

where ξ is the unique solution in $(0, 1)$ of

$$\sqrt{\pi}h(1 - \sqrt{\pi} + \sqrt{\pi}\xi) + (1 - \sqrt{\pi})\lambda = \lambda\xi. \tag{11.3.7}$$

2. For $\pi \leq \pi_c$, $v(\mathcal{C}_{(1)})/N \xrightarrow{\mathbb{P}} 0$.

For the standard configuration model, a special case of HCM with all communities of size one, (11.3.5) simplifies to $\pi_c = \mathbb{E}[D]/\mathbb{E}[D(D - 1)]$, since in that case, for a vertex of degree d , $g(v, v, k, \pi) = \mathbb{1}_{\{k=d\}}$. Furthermore, $\pi_c = 0$ when for any $\pi > 0$, the expected number of edges to other communities is infinite when entering a community via a uniformly chosen edge.

The proof of Theorem 11.2 has a similar structure as the proof of [68, Theorem 1]. The proof consists of three key steps:

- (a) First, each intra-community edge is removed with probability $1 - \pi$. This may split the community into several connected components. We find the distribution of the inter-community degrees of the connected components of the percolated communities, which is given by p'_k , as in (11.3.1). We identify vertices that are in the same connected component of a community. Lemma 11.1 shows that this results in a graph $\phi(G_\pi)$ that is distributed as a configuration model with asymptotic degree probabilities p'_k (recall (11.3.1)).
- (b) We then remove each intra-community edge with probability $1 - \pi$. Results of [115] can now be applied to the configuration model with distribution p'_k to find the critical percolation value and the size of the giant hierarchical component.
- (c) Next, we translate the number of communities in the largest percolated hierarchical component to the number of vertices. Then we show that this is indeed the largest component of the percolated graph.

We now present the details in these steps.

Auxiliary graph. We introduce the auxiliary graph $\phi(\bar{G})$, defined for every subgraph $\bar{G} \subset G$, and obtained by identifying the vertices that belonged to the same community in G , and are connected in \bar{G} [68]. Hence, in $\phi(\bar{G})$ every vertex represents a connected part of a community. Figure 11.5 illustrates $\phi(\bar{G})$. For a HCM G , the graph $\phi(G)$ is a configuration model where communities of G are collapsed into single vertices.

Lemma 11.1. *Let G be a HCM satisfying Conditions 11.1 and 11.2. Let G_π denote the subgraph of G where each intra-community edge is removed with probability $1 - \pi$. Then the graph $\phi(G_\pi)$ is distributed as a configuration model with degree probabilities p'_k given in (11.3.1).*

Proof. We independently delete each intra-community edge with probability $1 - \pi$. We want to find the degree distribution of $\phi(G_\pi)$. Let $M^{(n)}(H, v, k, \pi)$ denote the number of connected components of the percolated versions of community H containing vertex v and having inter-community degree k . Each community of shape H has an

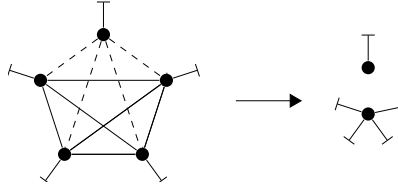


Figure 11.5: Left, a subgraph \bar{H} of a community H , dashed lines are not in \bar{H} , but are present in H , solid lines are present in \bar{H} . The graph $\phi(\bar{H})$ is shown on the right.

equal probability that the component containing v has inter-community degree k given by $g(H, v, k, \pi)$. Furthermore, the probability that a randomly chosen community has shape H is independent of the probability that the inter-community degree is k after intra-community percolation in community H . Therefore, given the number of type H communities $n_H^{(n)}$, we can write $M^{(n)}(H, v, k, \pi) \sim \text{Bin}(n_H^{(n)}, g(H, v, k, \pi))$. Thus, by the weak law of large numbers,

$$\frac{M^{(n)}(H, v, k, \pi)}{n} = \frac{M^{(n)}(H, v, k, \pi)}{n_H^{(n)}} \frac{n_H^{(n)}}{n} \xrightarrow{\mathbb{P}} P(H)g(H, v, k, \pi). \quad (11.3.8)$$

Let $N^{(n)}(H, k, \pi)$ denote the total number of connected components of the percolated versions of H having inter-community degree k . This number can be obtained by counting the number of half-edges of all connected components of percolated graphs with inter-community degree k , and then dividing by k . Each vertex v in such a percolated community contributes $d_v^{(b)}$ to the inter-community degree of the percolated community. Thus,

$$N^{(n)}(H, k, \pi) = \sum_{v \in V_H} d_v^{(b)} M^{(n)}(H, v, k, \pi) / k. \quad (11.3.9)$$

Let \tilde{n} denote the number of vertices in $\phi(G_\pi)$, so that $\tilde{n} = \sum_H \sum_k N^{(n)}(H, k, \pi)$. Similarly, the number vertices of degree k in $\phi(G_\pi)$ is denoted by $\tilde{n}_k = \sum_H N^{(n)}(H, k, \pi)$. Furthermore, $\sum_k N^{(n)}(H, k, \pi) / n \leq P_n(H) s_H$, and therefore by the Dominated Convergence Theorem, Condition 11.1, (11.3.8) and (11.3.9),

$$\tilde{n} / n \xrightarrow{\mathbb{P}} \sum_H \sum_{k=1}^{D_H} \sum_{v \in V_H} d_v^{(b)} P(H) g(H, v, k, \pi) / k. \quad (11.3.10)$$

This also implies that

$$\frac{N^{(n)}(H, k, \pi)}{\tilde{n}} = \frac{N^{(n)}(H, k, \pi) / n}{\tilde{n} / n} \xrightarrow{\mathbb{P}} \frac{\sum_{v \in V_H} d_v^{(b)} P(H) g(H, v, k, \pi) / k}{\sum_H \sum_{v \in V_H} \sum_l d_v^{(b)} P(H) g(H, v, l, \pi) / l}. \quad (11.3.11)$$

Hence, the proportion of vertices in $\phi(G_\pi)$ with degree k tends to

$$\frac{\tilde{n}_k}{\tilde{n}} = \sum_H \frac{N^{(n)}(H, k, \pi)}{\tilde{n}} \xrightarrow{\mathbb{P}} p'_k. \quad (11.3.12)$$

Since the edges between communities in G were paired at random, this means that the graph $\phi(G_\pi)$ is distributed as a configuration model with degree probabilities p'_k . \square

Using Lemma 11.1, we now prove Theorem 11.2:

Proof of Theorem 11.2. Step (a). Lemma 11.1 proves that $\phi(G_\pi)$ is distributed as a configuration model with degree probabilities p'_k .

Step (b). $\phi(G_\pi)$ and $\phi(G)$ have $\sum_k k\tilde{n}_k$ and $\sum_k kn_k$ half-edges, respectively. Since only intra-community edges have been deleted, $\sum_k kn_k = \sum_k k\tilde{n}_k$. By Condition 11.2(i), $\sum_k kn_k/n \rightarrow \mathbb{E}[D]$. Furthermore, by (11.3.10) \tilde{n}/n converges, hence $\sum_k k\tilde{n}_k/\tilde{n}$ converges. Therefore we can apply Theorem 3.9 from [115], which states that after percolation, a configuration model with degree probabilities p'_k has a giant component if

$$\pi \sum_k k(k-1)p'_k > \sum_k kp'_k. \tag{11.3.13}$$

From Theorem 11.1 we know that a giant hierarchical component is also a giant component in G , and a hierarchical component of size $o_{\mathbb{P}}(n)$ is a component of size $o_{\mathbb{P}}(N)$. Hence, the giant component emerges precisely when the giant hierarchical component emerges. Substituting (11.3.12) gives for the critical percolation value π_c that

$$\begin{aligned} \pi_c &= \frac{\sum_k kp'_k}{\sum_k k(k-1)p'_k} = \frac{\sum_H \sum_{v \in V_H} d_v^{(b)} \sum_{k=1}^{D_H} P(H)g(H, v, k, \pi_c)k/k}{\sum_H \sum_{v \in V_H} d_v^{(b)} \sum_{k=1}^{D_H} P(H)g(H, v, k, \pi_c)k(k-1)/k} \\ &= \frac{\sum_H \sum_{v \in V_H} d_v^{(b)} P(H)}{\sum_H \sum_{v \in V_H} d_v^{(b)} \sum_{k=1}^{D_H} P(H)g(H, v, k, \pi_c)(k-1)} \\ &= \frac{\mathbb{E}[D]}{\sum_H \sum_{v \in V_H} \sum_{k=1}^{D_H-1} d_v^{(b)} P(H)g(H, v, k+1, \pi_c)k} \\ &= \frac{1}{\mathbb{E}[D_{\pi_c}^*]}. \end{aligned} \tag{11.3.14}$$

Since $\mathbb{E}[D_{\pi}^*]$ is increasing in π , so that there is only one solution to the above equation. **Step (c).** Now assume that $\pi > \pi_c$. Let $\mathcal{C}_{(1)}^{(\text{CM})}$ denote the largest hierarchical component after percolation. The number of degree r vertices in the largest component of $\phi(G_\pi)$ satisfies $v_r^{(\text{H})}(\mathcal{C}_{(1)}^{(\text{CM})})/\tilde{n} \xrightarrow{\mathbb{P}} \sum_{l \geq r} b_{lr}(\sqrt{\pi})p'_l(1 - \zeta^r)$ [115], where $b_{lr}(\sqrt{\pi}) = \binom{l}{r} \sqrt{\pi}^r (1 - \sqrt{\pi})^{l-r}$ is the probability that a binomial with parameters l and $\sqrt{\pi}$ takes value r , and ζ is as in (11.3.7). To translate the number of percolated communities in the largest component into the number of vertices in the largest component, we need to know the expected number of vertices in the largest component that are in a percolated community with inter-community degree k . The size of a percolated community is independent of being in the largest hierarchical component, but does depend on the inter-community degree of the percolated community. The total number of vertices in connected percolated components with inter-community degree k is given by $\sum_H \sum_{v \in V_H} M^{(n)}(H, v, k, \pi)$, and the total number of percolated

communities of inter-community degree k is given by $\sum_H N^{(n)}(H, k, \pi)$. Furthermore, $\sum_{v \in V_H} M^{(n)}(H, v, k, \pi) / n \leq P_n(H) s_H$. Hence, by Condition 11.2, the expected size of a percolated community, given that it has inter-community degree k , satisfies

$$\begin{aligned} \mathbb{E}[S_\pi \mid \text{inter-community degree } k] &= \frac{\sum_H \sum_{v \in V_H} M^{(n)}(H, v, k, \pi)}{\sum_H N^{(n)}(H, k, \pi)} \\ &= \frac{\sum_H \sum_{v \in V_H} M^{(n)}(H, v, k, \pi) / n}{\sum_H N^{(n)}(H, k, \pi) / n} \\ &\xrightarrow{\mathbb{P}} \frac{\sum_H \sum_{v \in V_H} P(H) g(H, v, k, \pi)}{\sum_H \sum_{v \in V_H} d_v^{(b)} P(H) g(H, v, k, \pi) / k}. \end{aligned} \quad (11.3.15)$$

Since $\sum_{r=0}^{\infty} v_r(\mathcal{C}_{(1)}^{(\text{CM})}) / N \leq 1$, we use the Dominated Convergence Theorem to compute the asymptotic number of vertices in the largest component of $\phi(G_\pi)$ as

$$\begin{aligned} \frac{v(\mathcal{C}_{(1)}^{(\text{CM})})}{N} &= \sum_{r=0}^{\infty} \frac{v_r(\mathcal{C}_{(1)}) / \tilde{n}}{N / n \cdot n / \tilde{n}} \\ &\xrightarrow{\mathbb{P}} \frac{\sum_{l \geq r} b_{lr}(\sqrt{\pi}) p_l' (1 - \zeta^r) \mathbb{E}[S_\pi \mid \text{inter-community degree } l]}{\mathbb{E}[S] / \sum_H \sum_{v \in V_H} \sum_k d_v^{(b)} P(H) g(H, v, k, \pi) / k} \\ &= \frac{\sum_{r=0}^{\infty} \sum_{l \geq r} b_{lr}(\sqrt{\pi}) (1 - \zeta^r) p_l'}{\mathbb{E}[S] / \sum_H \sum_{v \in V_H} \sum_k d_v^{(b)} P(H) g(H, v, k, \pi) / k} \\ &\quad \times \frac{\sum_H \sum_{v \in V_H} P(H) g(H, v, l, \pi)}{\sum_H \sum_{v \in V_H} d_v^{(b)} P(H) g(H, v, l, \pi) / l} \\ &= \frac{\sum_{l=0}^{\infty} \sum_{r=0}^l b_{lr}(\sqrt{\pi}) (1 - \zeta^r) \frac{p_l'}{p_l} \sum_H \sum_{v \in V_H} P(H) g(H, v, l, \pi)}{\mathbb{E}[S]} \\ &= \frac{\sum_{l=0}^{\infty} (1 - (1 - \sqrt{\pi} + \sqrt{\pi} \zeta)^l) \sum_H \sum_{v \in V_H} P(H) g(H, v, l, \pi)}{\mathbb{E}[S]}. \end{aligned} \quad (11.3.16)$$

Any other component of $\phi(G_\pi)$ has size $o_{\mathbb{P}}(\tilde{n})$ by [115, Theorem 3.9]. As shown in the proof of Theorem 11.1, any component of size $o_{\mathbb{P}}(\tilde{n})$ in $\phi(G_\pi)$ is a component of size $o_{\mathbb{P}}(N)$ in the total graph. Hence, w.h.p. $\mathcal{C}_{(1)}^{(\text{CM})}$ is the largest component of the percolated graph G_π .

When $\pi < \pi_c$, the largest component of $\phi(G_\pi)$ satisfies $v^{(H)}(\mathcal{C}_{(1)}^{(\text{CM})}) / \tilde{n} \xrightarrow{\mathbb{P}} 0$ [115]. Again, by the analysis of Theorem 11.1, this component is of size $o_{\mathbb{P}}(N)$ in the original graph. \square

Equation (11.3.7) also has an intuitive explanation. Let Q be the distribution of the inter-community degrees after percolation when following a randomly chosen half-edge. Then we can interpret ζ as the extinction probability of a branching process with offspring distribution Q . Percolating the inter-community edges with probability $1 - \pi$ is the same as deleting each half-edge with probability $1 - \sqrt{\pi}$. Then, with probability $1 - \sqrt{\pi}$ the randomly chosen half-edge is paired to a deleted half-edge, in which case the branching process goes extinct. With probability $\sqrt{\pi}$, the half-edge

leads to a half-edge which still exists after percolation, and leads to a community. The probability generating function of the number of half-edges pointing out of this community before percolating the half-edges is $\frac{1}{\lambda}h(\xi)$. Since the number of half-edges after percolation is binomial given the number of half-edges that were present before percolation, the probability generating function of the number of half-edges attached to a community entered by a randomly chosen half-edge is $\frac{1}{\lambda}h(1 - \sqrt{\pi} + \sqrt{\pi}\xi)$. Combining this yields (11.3.7).

The case $\mathbb{E}[D^2] = \infty$. In the standard configuration model $\pi_c = 0$ precisely when $\mathbb{E}[D^2] = \infty$. In the HCM, this may not be true, since it is possible to construct communities with large inter-community degrees, while all individual vertices have low degree. An example of such a community structure is the HCM where each community is a line graph H_L of L vertices with probability \bar{p}_L , where each vertex has inter-community degree one. Figure 11.6a illustrates H_L for $L = 5$. We assume that \bar{p}_L obeys the power law $\bar{p}_L = cL^{-\alpha}$, with $\alpha \in (2, 3)$. Then $\mathbb{E}[D] < \infty$, but $\mathbb{E}[D^2] = \infty$. Hence, communities may have large inter-community degrees. However, G is a 3-regular graph, so no individual vertex has high degree. From this fact, we can already conclude that $\pi_c \neq 0$. Suppose $\pi_c < \frac{1}{2}$. Then, after percolation every vertex has less than two expected neighbors. Hence, there is no giant component w.h.p. We can also use Theorem 11.2 to show that $\pi_c \neq 0$. We compute the denominator of (11.3.5), and show that it is finite. We have

$$\sum_{v \in V_H} d_v^{(b)} g(H_L, v, k, \pi_c) = \begin{cases} 2k\pi_c^{k-1}(1 - \pi_c) + k\pi_c^{k-1}(1 - \pi_c)^2(L - k - 1) & \text{if } k < L, \\ k\pi_c^{k-1} & \text{if } k = L. \end{cases} \tag{11.3.17}$$

This gives

$$\sum_{k=1}^{L-1} k \sum_{v \in V_H} d_v^{(b)} g(H_L, v, k + 1, \pi_c) = \frac{2\pi_c(\pi_c^L + L(1 - \pi_c)) - 1}{(1 - \pi_c)^2}. \tag{11.3.18}$$

Using that $\bar{p}_l = cl^{-\alpha}$ gives for (11.3.4)

$$\begin{aligned} \mathbb{E}[D_{\pi_c}^*] &= \frac{1}{\mathbb{E}[D]} \sum_H P(H) \sum_{v \in V_H} d_v^{(b)} \sum_{k=1}^{D_H-1} kg(H, v, k + 1, \pi_c) \\ &= \frac{1}{\mathbb{E}[D]} \sum_{L=1}^{\infty} cL^{-\alpha} \frac{2\pi_c(\pi_c^L + L(1 - \pi_c)) - 1}{(1 - \pi_c)^2} \\ &= \frac{1}{\mathbb{E}[D]} \frac{2\pi_c}{(1 - \pi_c)^2} \left(-1 + (1 - \pi_c)\mathbb{E}[D] + \sum_{L=1}^{\infty} c\pi_c^L L^{-\alpha} \right). \end{aligned} \tag{11.3.19}$$

From (11.3.19) we see that $\pi_c = 0$ is not a solution of (11.3.5). Hence, $\pi_c > 0$, even though $\mathbb{E}[D^2] = \infty$.

Infinite second moment of degree. When the second moment of the degree distribution as defined in Proposition 11.1 is infinite, π_c also does not have to be zero. It

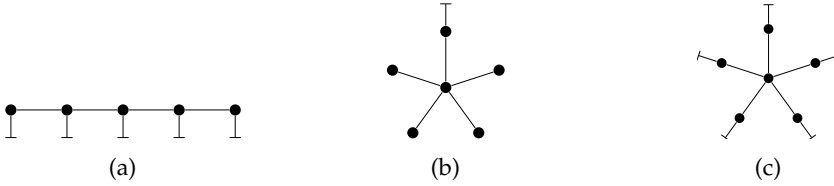


Figure 11.6: Communities with $L = 5$.

is possible to ‘hide’ all vertices of high degree inside communities that have small inter-community degrees. The small inter-community degrees make it difficult to leave the community in percolation. One example of such a community structure is the case in which each community is a star-shaped graph with L endpoints with probability p_L . One vertex in the graph has inter-community degree one, and all the other vertices have inter-community degree zero. Figure 11.6b illustrates the star-shaped graph for $L = 5$. Since each community has only one outgoing edge, there cannot be a giant component in G . We can also see this from Theorem 11.1, since $\mathbb{E}[D] = \mathbb{E}[D^2] = 1$. By Proposition 11.1, the degree distribution equals

$$\hat{p}_k = \begin{cases} (\sum_{L=1}^{\infty} (L-1)p_L + p_1) / \mathbb{E}[S] & \text{if } k = 1, \\ (1 + p_2) \mathbb{E}[S] & \text{if } k = 2, \\ p_k / \mathbb{E}[S] & \text{if } k > 2. \end{cases} \quad (11.3.20)$$

When p_l is a probability distribution with infinite second moment, the second moment of \hat{p}_l is also infinite. Hence, the degree distribution of the HCM G has infinite second moment, while there is no giant component, so that certainly $\pi_c \neq 0$.

A sufficient condition for $\pi_c = 0$.

By (11.3.5),

$$\begin{aligned} \pi_c &= \frac{\mathbb{E}[D]}{\sum_G P(H) \sum_{v \in V_H} d_v^{(b)} \sum_{k=1}^{D_H-1} k g(H, v, k+1, \pi_c)} \\ &\leq \frac{\mathbb{E}[D]}{\sum_G P(H) \sum_{v \in V_H} (d_v^{(b)})^2} \end{aligned} \quad (11.3.21)$$

Hence, $\sum_H \sum_{v \in V_H} P(H) (d_v^{(b)})^2 = \infty$ is a sufficient condition for $\pi_c = 0$. This condition can be interpreted as an infinite second moment of the inter-community degrees of individual vertices. However, it is not a necessary condition. It is possible to construct a community where all individual vertices have a small inter-community degree, but are connected to a vertex with high degree. Consider for example the star community of Figure 11.6c, with one vertex in the middle, linked to L other vertices. The L other vertices have inter-community degree one, and the middle vertex has inter-community degree zero, hence all vertices have small inter-community degree. However, the middle vertex can be of high degree. Let each community be a star-shaped community with L outgoing edges with probability \bar{p}_L . We can calculate that $\pi_c = \sum_L \bar{p}_L L(L-1) \pi_c^2$. Hence, if we choose \bar{p}_L with finite first moment and infinite second moment, $\pi_c = 0$. However, $\sum_H \sum_{v \in V_H} (d_v^{(b)})^2 = \sum_L L \bar{p}_L = \mathbb{E}[D] < \infty$.

11.4 Existing graph models with a community structure

In this section, we show how three existing random graph models with community structure fit within the HCM.

11.4.1 Trapman's household model

Trapman [206] replaces vertices in a configuration model by households in the form of complete graphs, such that the degree distribution of the resulting graph is p_k . To achieve this, each community is a single vertex of degree k with probability $(1 - \gamma)p_k$, or a complete graph of size k with probability $\gamma\bar{p}_k$. Here \bar{p}_k , the probability that a certain clique has degree k , is given by

$$\bar{p}_k = k^{-1}p_k\mathbb{E}[W^{-1}]^{-1}, \quad (11.4.1)$$

where W is a random variable satisfying $\mathbb{P}(W = k) = p_k$. Each vertex of the complete graph has one edge to another community. Figure 11.7 illustrates a household of size 5. This model is a special case of the HCM with

$$H_i = \begin{cases} (K_k, (1, \dots, 1)) & \text{w.p. } \gamma\bar{p}_k, \\ (v, (k)) & \text{w.p. } (1 - \gamma)p_k, \end{cases} \quad (11.4.2)$$

where K_k is a complete graph on k vertices.

We now check when (11.4.2) satisfies Conditions 11.1 and 11.2. The assumption $\mathbb{P}(D = 2) < 1$ is satisfied if and only if $p_2 < 1$. The expected inter-community degree of a community is given by

$$\mathbb{E}[D] = (1 - \gamma) \sum_k kp_k + \gamma \sum_k k\bar{p}_k = (1 - \gamma)\mathbb{E}[W] + \frac{\gamma}{\mathbb{E}[W^{-1}]}. \quad (11.4.3)$$

Hence, $\mathbb{E}[D] < \infty$ if $\mathbb{E}[W] < \infty$ and $\mathbb{E}[W^{-1}] \neq 0$. By Jensen's inequality, $\mathbb{E}[W^{-1}] \geq \mathbb{E}[W]^{-1} > 0$, hence $\mathbb{E}[D] < \infty$ if and only if $\mathbb{E}[W] < \infty$. For every community in this model, its size is smaller than or equal to its inter-community degree, so that also $\mathbb{E}[S] < \infty$ if $\mathbb{E}[D] < \infty$. Thus, Conditions 11.1 and 11.2 hold if $\mathbb{E}[W] < \infty$ and $p_2 < 1$. Under these conditions we can apply the results for the HCM as derived in Sections 11.2 and 11.3.

Suppose that p_k follows a power law with exponent α . Then \bar{p}_k follows a power law with exponent $\alpha - 1$, and the distribution of the inter-community degrees D is a mixture of a power law with exponent α and a power law with exponent $\alpha - 1$ by (11.4.1).

11.4.2 Lelarge and Coupechoux' household model

Another model that takes complete graphs as communities is the model of Coupechoux and Lelarge [68]. This model is very similar to Trapman's model. Again, each community is either a complete graph or a single vertex. In contrast to [206], the probability that a certain community is a clique depends on the degree of the clique.

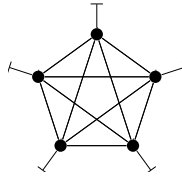


Figure 11.7: A household of size 5

Each vertex of degree k in the macroscopic configuration model is replaced by a complete graph with probability γ_k . This graph can be modeled as a HCM with

$$H_i = \begin{cases} (K_k, (1, \dots, 1)) & \text{w.p. } \gamma_k \bar{p}_k \\ (v, (k)) & \text{w.p. } (1 - \gamma_k) \bar{p}_k \end{cases} \quad (11.4.4)$$

where $(\bar{p}_k)_{k \geq 1}$ is a probability distribution. Since the inter-community degrees of all communities have distribution $\mathbb{P}(D = k) = \bar{p}_k$, Condition 11.2 holds if the probability distribution \bar{p}_k has finite mean and $\bar{p}_2 < 1$. The size of a community is always smaller than or equal to its inter-community degree, so that also $\mathbb{E}[S] < \infty$ if \bar{p}_k has finite mean. Thus, Conditions 11.1 and 11.2 hold if \bar{p}_k has finite mean and $\bar{p}_2 < 1$.

If these conditions on \bar{p}_k hold, then the degree distribution p_k of the resulting graph can be obtained from Proposition 11.1 as

$$p_k = \frac{(k\gamma_k + (1 - \gamma_k))\bar{p}_k}{\sum_{i \geq 0} (i\gamma_i + (1 - \gamma_i))\bar{p}_i}. \quad (11.4.5)$$

Suppose that $\gamma_k \geq \gamma > 0$. Then, in contrast to Trapman's household model in Section 11.4.1, the degree distribution of the edges between communities, \bar{p}_k , follows a power law with exponent $\alpha + 1$ if the degree distribution p_k follows a power law with exponent α .

As an example of such a household model, consider a graph with $p_3 = a$ and $p_6 = 1 - a$ and a tunable clustering coefficient. We take $\gamma_6 = 0$, but increase γ_3 , while the degree distribution remains the same. Thus, the graph consists of only single vertices of degree 6, single vertices of degree 3 and triangle communities. Since we increase γ_3 , the number of triangles increases, so that also the clustering coefficient increases. Figures 11.8a and 11.8b show the size of the giant component under percolation for different values of the clustering coefficient using $a = 0.75$ and $a = 0.95$ respectively. In the case where $a = 0.75$, clustering decreases the value of π_c , whereas if $a = 0.95$, clustering increases the value of π_c . This illustrates that the influence of clustering on bond percolation of a random graph is non-trivial. In two similar random graph models, introducing clustering has a different effect.

11.4.3 Configuration model with triangles

A third random graph model with clustering is the model by Newman [157]. In this model, each vertex v has an edge-degree $d_v^{(1)}$ and a triangle degree $d_v^{(2)}$, denoting the number of triangles that the vertex is part of. Then a random graph is formed by

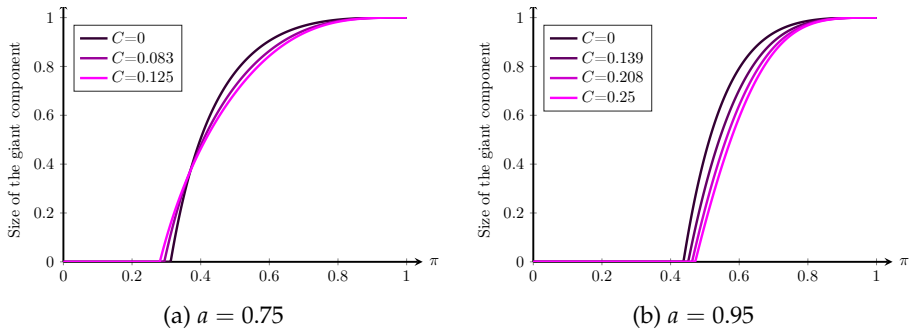


Figure 11.8: The size of the giant component after bond percolation with probability π in a household model with $p_3 = a$ and $p_6 = 1 - a$ for various clustering coefficients C . If $a = 0.75$, adding clustering decreases the critical percolation value, whereas if $a = 0.95$, adding clustering increases the critical percolation value.

pairing edges at random and pairing triangles at random. Even though this model does not explicitly replace vertices in a configuration model by communities, it is also a special case of the HCM if some conditions on the degrees are satisfied. The communities in this model are the connected components consisting only of triangles. Figure 11.9 shows two possible realizations of such communities.

From results derived in [157], we can find the probability generating function $h_r(z)$ of the number of vertices in triangles that can be reached from a uniformly chosen triangle, and the probability generating function $h_{S^*}(z)$ of the size of the triangle component of a randomly chosen vertex, that together satisfy

$$h_r(z) = z g_q(h_r^2(z)), \quad h_{S^*}(z) = z g_p(h_r^2(z)), \tag{11.4.6}$$

where g_q is the probability generating function of the size-biased distribution of the triangle degrees, and g_p the probability generating function of the triangle degree distribution. In the HCM, $h_{S^*}(z)$ can be interpreted as the probability generating function of the size-biased community sizes. Thus, the mean size-biased community size is given by

$$\mathbb{E}[S^*] = 1 + \frac{2\mathbb{E}[D^{(2)}]}{3 - 2\mathbb{E}[D^{(2)*}]}, \tag{11.4.7}$$

where $D^{(2)*}$ is the size-biased distribution of the triangle degrees. Since $\mathbb{E}[S^*] \geq \mathbb{E}[S]$, Condition 11.1(ii) is satisfied if $\mathbb{E}[D^{(2)*}] < \frac{3}{2}$.

The mean inter-community degree of a community is given by

$$\mathbb{E}[D] = \lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n \sum_{v \in G_i} d_v^{(1)}}{n} = \lim_{n \rightarrow \infty} \frac{\sum_{i=1}^N d_i^{(1)} / N}{n/N} = \mathbb{E}[S]\mathbb{E}[D^{(1)}]. \tag{11.4.8}$$

Hence, Conditions 11.1 and 11.2 are satisfied if $\mathbb{E}[D^{(2)*}] < \frac{3}{2}$ and $\mathbb{E}[D^{(1)}] < \infty$. When these conditions are satisfied, the condition for emergence of a giant component is

$$\frac{\mathbb{E}[D^{(1)2}]\mathbb{E}[S] - \mathbb{E}[D^{(1)}]\mathbb{E}[S]}{\mathbb{E}[D^{(1)}]\mathbb{E}[S]} = \frac{\mathbb{E}[D^{(1)2}] - \mathbb{E}[D^{(1)}]}{\mathbb{E}[D^{(1)}]} > 1. \tag{11.4.9}$$

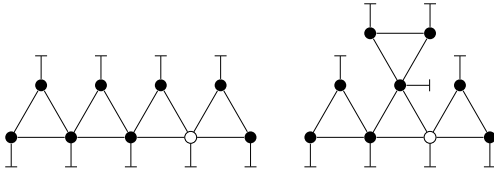


Figure 11.9: Two possible communities with 4 triangles. In the left community, 6 other nodes can be reached from the white node within 2 steps, in the right community 8 nodes.

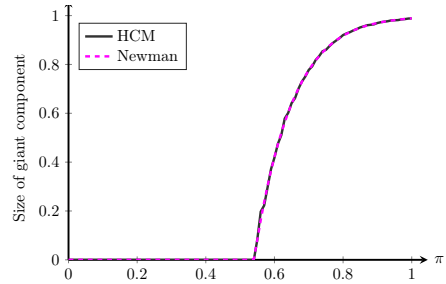


Figure 11.10: The size of the giant percolating cluster calculated by (11.3.6) (HCM) and from results in [157] (Newman) agree.

Therefore, as long as $\mathbb{E}[D^{(2)*}] < \frac{3}{2}$, the emergence of the giant component only depends on the edge degree distribution.

To apply the results of the HCM, we need the probability $P(H)$ that a randomly chosen community is of type H . This probability is not easy to obtain, but it can be approximated using a branching process. The branching process starts at a vertex, and explores the component of triangles. The first generation of the branching process has $Z_0 = 1$. The first offspring, Z_1 is distributed as $2D^{(2)}$. All other offspring, Z_i for $i > 1$ is distributed as $\sum_{j=1}^{Z_{i-1}} 2(D_j^{(2)*} - 1)$. Here $D_j^{(2)*}$ are independent copies, distributed as $D^{(2)*}$. In this branching process approximation, cycles of triangles are ignored. The size-biased probability of having a specific community H can be obtained by summing the probabilities of the possible realizations of the branching process when exploring graph H . This probability is size-biased, since when starting at an arbitrary vertex, the probability of starting in a larger community is higher. This probability then needs to be transformed to the probability of obtaining graph H .

To compute the size of the giant component after percolation from (11.3.6), $g(H, v, k, \pi)$ is needed for every community shape H . This is difficult to obtain, since it largely depends on the shape of the community, and there are infinitely many possible community shapes. Figure 11.9 shows an example of why the shape of a community matters. When percolating the left community, the probability that the red vertex is connected to k other vertices is smaller than for the graph on the right. For this reason, we approximate (11.3.6) numerically using the branching process described above. In [157], Newman gives expressions for the size of the largest percolating cluster. Figure 11.10 compares the size of the giant component computed in that way with a numerical approximation of (11.3.6). We see that indeed the equations from [157] give the same results for the largest percolating cluster as (11.3.6).

11.5 Stylized networks

In this section, we study two stylized examples of community structures. The first example gives a community type that decreases the critical percolation value compa-



Figure 11.11: A line community with $L = 5$

red to a configuration model with the same degree distribution. The second example increases the critical percolation value when compared to a configuration model.

11.5.1 A community structure that decreases π_c

As an example of a community structure that decreases π_c , we consider a HCM where with probability ϕ a community is given by H_1 : a path of L vertices, with a half-edge at each end of the path as illustrated in Figure 11.11. With probability $1 - \phi$ the community is H_2 : a vertex with three half-edges. The degree distribution of this HCM can be found using Proposition 11.1 and is given by

$$p_k = \begin{cases} \frac{L\phi}{L\phi+1-\phi} & \text{if } k = 2, \\ \frac{1-\phi}{L\phi+1-\phi} & \text{if } k = 3, \\ 0 & \text{otherwise.} \end{cases} \tag{11.5.1}$$

In this example $\mathbb{E}[D] = 2\phi + 3(1 - \phi)$. Furthermore, $g(H_1, v, 2, \pi) = \pi^{L-1}$ for all $v \in H_1$. In H_2 there is no percolation inside the community, hence $g(H_2, v, 3, \pi) = 1$. Equation (11.3.5) now gives

$$\pi_c = \frac{3 - \phi}{2\phi\pi_c^{L-1} + 6(1 - \phi)}, \tag{11.5.2}$$

hence $2\phi\pi_c^L + 6(1 - \phi)\pi_c - 3 + \phi = 0$.

Now we let the degree distribution as defined in (11.5.1) remain the same, while changing the length of the path communities L . If in the total graph, we want to have a fraction of a vertices of degree 3, then $a = p_3 = \frac{1-\phi}{1-\phi+L\phi}$. Hence, $\phi = \frac{1-a}{1-a+La}$. In this way, we obtain HCMs with the same degree distribution, but with different values of L . Figure 11.12 shows the size of the largest component as calculated by (11.3.6) for $a = 1/3$. As L increases, π_c decreases. Hence, adding this community structure ‘helps’ the diffusion process. This can be explained by the fact that increasing L decreases the number of line communities. Therefore, more vertex communities will be connected to one another, which decreases the value of π_c . Another interesting observation is that the size of the giant component is non-convex in π . These non-convex shapes can be explained intuitively. As the lines get longer, there are fewer and fewer of them, since the degree distribution remains the same. Hence, if L is large, there will only be a few long lines. These lines have $\pi_c \approx 1$. Since there are only a few lines, almost all vertices of degree 3 will be paired to one another. The critical value for percolation on a configuration model with only vertices of degree 3 is 0.5. Hence, for this HCM with L large we will see the vertices of degree 3 appearing in the giant component as $\pi = 0.5$, and the vertices in the lines as $\pi = 1$.

11.5.2 A community structure that increases π_c

As an example of a community structure that inhibits the diffusion process, consider a configuration model with intermediate vertices as introduced in [110]: a configuration model where every edge is replaced by two edges with a vertex in between them. This is equal to a HCM with star-shaped communities as in Figure 11.6c: one vertex that is connected to L other vertices. Each of the L other vertices has inter-community degree one. The vertex in the middle is not connected to other communities. We consider a HCM where all communities are stars of the same size. Therefore all star-shaped communities have the same number of outgoing edges, and $\mathbb{E}[D] = L$.

The degree distribution of this HCM is given by

$$p_k = \begin{cases} \frac{L}{L+1} & \text{if } k = 2, \\ \frac{1}{L+1} & \text{if } k = L, \\ 0 & \text{otherwise.} \end{cases} \quad (11.5.3)$$

Under percolation, the connected component of a vertex v at the end point of a star can link to other half-edges only if the edge to the middle vertex is present. Then the number of half-edges to which v is connected is binomially distributed, so that $g(H, v, k, \pi) = \pi \binom{L-1}{k-1} \pi^{k-1} (1-\pi)^{L-k}$ for $k \geq 2$. Then (11.3.4) gives

$$\begin{aligned} \mathbb{E}[D_\pi^*] &= \frac{1}{\mathbb{E}[D]} \sum_H \sum_k \sum_{v \in V_H} P(H) d_v^{(b)} k g(H, v, k+1, \pi) \\ &= \pi \sum_{k \geq 1} k \pi^k (1-\pi)^{L-k-1} \binom{L-1}{k} = (L-1) \pi^2. \end{aligned} \quad (11.5.4)$$

Then (11.3.5) yields $\pi_c = (L-1)^{-1/3}$.

Now we consider a configuration model with the same degree distribution (11.5.3). For this configuration model, $\pi_c = \frac{3L}{4L+L^2-3L} = \frac{3}{L+1}$. Figure 11.13 shows the size of the giant component of the HCM compared with a configuration model with the same degree distribution for different values of L . This hierarchical configuration has a higher critical percolation value than its corresponding configuration model. Intuitively, this can be explained from the fact that all vertices with a high degree are 'hidden' behind vertices of degree 2, whereas in the configuration model, vertices of degree L may be connected to one another.

Combined with the previous example, we see that adding communities may lead to a higher critical percolation value or a lower one. Furthermore, the size of the giant component may be smaller or larger after adding communities.

11.6 Conclusions and discussion

In this chapter, we have introduced the HCM, where the macroscopic graph is a configuration model, and on the microscopic level vertices are replaced by communities. We have analytically studied several properties of this random graph model, which led to several interesting insights. For example, the condition for a giant component to emerge in the HCM is completely determined by properties of the macroscopic

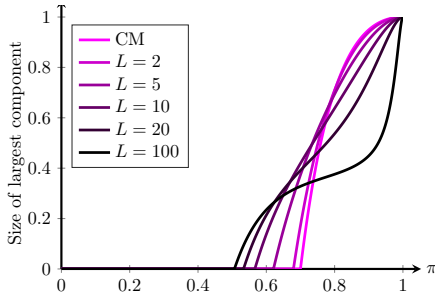


Figure 11.12: Size of giant component against π for line communities with different values of L .

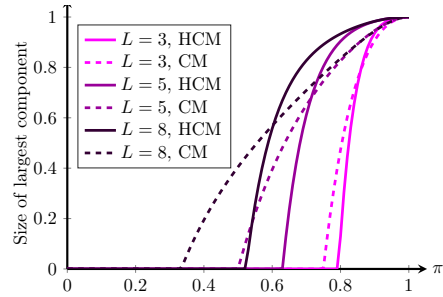


Figure 11.13: Size of giant component against π for star communities with different values of L .

configuration model. However, the size of the giant component also depends on the community sizes. In contrast, the asymptotic clustering coefficient is entirely defined by the clustering inside the communities. For bond percolation on the HCM, the critical percolation value depends on both the inter-community degree distribution as well as the shape of the communities.

Finally, we have shown that several existing models incorporating a community structure can be interpreted as a special case of the HCM, which underlines its generality. Worthwhile extensions of the HCM for future research include directed or weighted counterparts and a version that allows for overlapping communities.

The analysis of percolation on the HCM has shown that the size of the largest percolating cluster and the critical percolation value do not necessarily increase or decrease when adding clustering. It would be interesting to investigate how other characteristics of the graph like degree-degree correlations influence the critical percolation value.

12 Power-law relations in networks with community structures

Based on:

Power-law relations in random networks with communities
C. Stegehuis, R. van der Hofstad and J.S.H. van Leeuwen
Physical Review E 94 p.012302 (2016)

In this chapter, we again study the HCM model introduced in Chapter 11. We view several real-world networks as realizations of HCM. By doing this, we observe two previously undiscovered power-law relations: between the number of edges inside a community and the community sizes, and between the number of edges going out of a community and the community sizes. We also relate the power-law exponent τ of the degree distribution with the power-law exponent of the community-size distribution γ . In the case of extremely dense communities (e.g., complete graphs), this relation takes the simple form $\tau = \gamma - 1$.

12.1 Introduction

In Chapter 11 we studied the mathematical properties of the HCM. In this chapter, we look at several data sets through the lens of the HCM. A crucial property of the HCM is that it can use any proposed community structure as input. That is, the HCM viewed as an algorithm first models the community structure, and then creates the random network model. To be more specific, the community structure can be uncovered by some detection algorithm that, when applied to a real-world network, leads to a collection of plausible communities and their frequencies. By sampling from this collection of communities, the HCM can generate resampled networks with similar structure as the original network.

Analyzing several real-world networks through the lens of the HCM reveals an interesting property of real-world networks: the joint distribution $(p_{k,s})_{k,s \geq 1}$ of the community size s and the number of connections k a community has to other communities. The size of the giant component delicately depends on this joint distribution, which can be determined from network data once the community structure is determined. In fact, after studying $p_{k,s}$ for several real-world networks, we observe a *power-law relation between the size of a community and the number of edges out of a community in many real-world networks*.

In Chapter 11 we have shown that except for this joint distribution, the size of the giant component does not depend on detailed information about the structure

of the communities, but when we perform percolation on the network, more precise structural information does matter. Inspired by this need to include detailed community structure, and thus extend the model description beyond $(p_{k,s})_{k,s \geq 1}$, we observe a second *power-law relation between the denseness of a community and its size* in several real-world networks.

For the present chapter, the most important application of the HCM is to investigate power-law networks. Statistical analyses of network data shows that many networks possess a power-law degree distribution [66, 158, 162, 212]. Traditionally, this is captured by using the CM and assuming that the probability p_k that a node has k neighbors then scales with k as $p_k \sim Ck^{-\tau}$ for some constant C and power-law exponent $\tau > 0$. Many scale-free networks have an exponent τ between 2 and 3 [6, 79], so that the second moment of the degree distribution diverges in the infinite-size network limit. The exponent τ is an important characteristic of the network and determines for example the mean distances in the network [108, 109, 164], or the behavior of various processes on the graph like bond percolation [55], first passage percolation [23] and an intentional attack [67]. Using the HCM instead of the CM, it no longer suffices to describe the degree distribution $(p_k)_{k \geq 1}$, but instead assumptions need to be made about the joint distribution $(p_{k,s})_{k,s \geq 1}$. In the special case of extremely dense communities, this joint perspective gives rise to the following phenomenon (that we formalize in Section 12.2): *If the total degree distribution of a network with extremely dense communities follows a power law with exponent τ , then the power law of the community sizes has exponent $\gamma = \tau + 1$.* In the household model, where each community is a complete graph, we observe that indeed $\gamma = \tau + 1$. However, real-world network data show that communities are not always extremely dense, in which case we find that $\gamma \neq \tau + 1$. This is due to the fact that the edge density of communities turns out to decay with community size.

The outline of this chapter is as follows. In Section 12.2 we consider the special case of the HCM in which the degree distribution as well as the community-size distribution follows a power law. It is here that we discover the power-law shift caused by community structure when the communities are *extremely dense*. In Section 12.3 we apply the HCM to several real-world networks, and we observe two more power-law relations in graphs with communities. In Section 12.4 we present conclusions and future research directions.

12.2 Power-law community sizes

In several real-world networks, the community-size distribution appears to be of power-law form over some significant range [36, 65, 102, 186]. The analysis in this section assumes that both the degree and the community-size distributions obey power laws with exponents τ and γ , respectively. Typical values reported in the literature are $2 \leq \tau \leq 3$ and $1 \leq \gamma \leq 3$ [80]. We then investigate the behavior of $(p_{k,s})_{k,s \geq 1}$ defined in (11.1.3): the probability that a randomly chosen community has size s and inter-community degree k . We start by investigating a simple community structure where all communities are complete graphs, and then proceed to investigate more general dense community shapes.

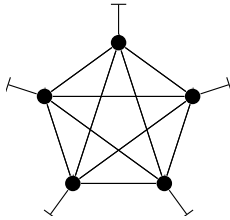


Figure 12.1: A household community with $s = 5$

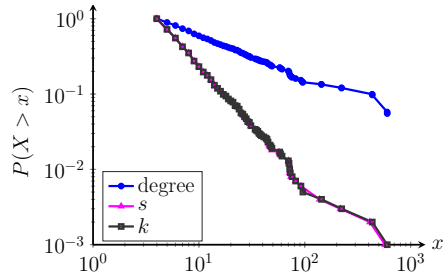


Figure 12.2: The degree distribution of a household model follows a power law with a smaller exponent than the community-size distribution and outside degree distribution

Household communities. An extreme community structure is that of household communities [206], in which all communities are complete graphs. Each vertex inside the community has outside degree one. Figure 12.1 shows an example of a household community with $s = 5$.

In a household community, $k = s$, hence $p_{k,s} = 0$ if $k \neq s$. Suppose the distribution of the community sizes follows a power law with exponent γ , $p_{k,k} = Ck^{-\gamma}$. Then the outside degrees also follow a power-law distribution with exponent γ . Now we derive $(\hat{p}_k)_{k \geq 0}$, the degree distribution of the HCM with this household structure. For a vertex in the household model to have degree k , it must be in a community of size k . Furthermore, there are k of such vertices inside each community, so that

$$\hat{p}_k = \frac{kp_{k,k}}{\sum_{i=1}^{\infty} ip_{i,i}} = \frac{kCk^{-\gamma}}{\mathbb{E}[S]} = C_2k^{-\gamma+1}, \tag{12.2.1}$$

where $\mathbb{E}[S]$ denotes the average community size. Thus, the degree distribution of the graph with household communities again obeys a power law but with exponent $\tau = \gamma - 1$, as observed in [206]. We call this phenomenon a *power-law shift*, because the edges out of a community have a smaller degree distribution than the individual edges (see Figure 12.2).

Extremely dense communities. We will now show that a power-law shift still occurs in communities that are not complete graphs, but still extremely dense. We should, however, remark upfront that the data sets that we will investigate in Section 12.3 do not display this extremely dense community structure. In an extremely dense community, many edges are contained in communities. Let e_{in} denote the number of edges inside a community. We assume that there exists $\varepsilon > 0$ independent of the number of communities n such that for each community H of size s ,

$$e_{\text{in}} \geq \varepsilon s(s - 1). \tag{12.2.2}$$

In this case, every community of size s contains a positive fraction of the edges that are present in a complete graph of the same size. The household model gives $\varepsilon = \frac{1}{2}$.

Since the power-law shift states that the outside degrees of the communities are ‘small’, we need the outside degrees of individual vertices to be small as well. Thus, we assume that there exists a $K < \infty$ such that for all vertices

$$d_v^{(b)} \leq Ks_v, \quad (12.2.3)$$

where s_v denotes the community size of the community containing vertex v . This implies that $k \leq Ks^2$ for every community. Using assumptions (12.2.2) and (12.2.3) we show that a power-law shift occurs.

Suppose that the community-size distribution follows a power law with exponent γ . Denote the cumulative degree distribution by $P_i = \sum_{j \leq i} \hat{p}_j$. Since the maximal inside degree of a vertex in a community of size s is $s - 1$, and by (12.2.2) the average inside degree of a vertex is greater than or equal to $\varepsilon(s - 1)$, at least a fraction of ε vertices in any community have inside degree at least $\varepsilon(s - 1)$. Thus, a vertex inside a community of size $i/\varepsilon + 1$ has probability of at least ε to have inside degree at least $\varepsilon(i/\varepsilon + 1 - 1) = i$. Hence, $1 - P_i$ is bounded from below by ε times the probability of choosing a vertex in a community of size at least $i/\varepsilon + 1$. The probability that a randomly chosen vertex is in a community of size j is given by $\sum_k r_{k,j}$. This yields

$$\begin{aligned} 1 - P_j &\geq \sum_{i \geq j} \sum_k r_{k,i/\varepsilon+1} \varepsilon = \sum_{i \geq j} \left(\frac{i}{\varepsilon} + 1 \right) \sum_k p_{k,i/\varepsilon+1} \frac{1}{\mathbb{E}[S]} \varepsilon \\ &\approx Cj^{-\gamma+2}. \end{aligned} \quad (12.2.4)$$

Furthermore, given the distribution of the community sizes, $1 - P_j$ is maximal when all communities are complete graphs, and every vertex has Ks half-edges attached to it. Then each vertex in a community of size s has degree $s - 1 + Ks$. Hence, to choose a vertex with degree at least j , we have to choose a vertex inside a community of size at least $(j + 1)/(K + 1)$. Then

$$\begin{aligned} 1 - P_j &\leq \sum_{i \geq \frac{j+1}{K+1}} \sum_k r_{k,i} \\ &= \sum_{i \geq \frac{j+1}{K+1}} i \sum_k p_{k,i} \frac{1}{\mathbb{E}[S]} = \frac{c}{\mathbb{E}[S]} \left(\frac{j+1}{K+1} \right)^{-\gamma+2}. \end{aligned} \quad (12.2.5)$$

Combining (12.2.4) and (12.2.5) shows that the degree distribution follows a power law with exponent $\tau = \gamma - 1$. In other words, when the community-size distribution of a network with extremely dense communities follows a power law with exponent γ , the power law of the degrees has exponent $\tau = \gamma - 1$. A similar analysis shows that for power-law distributions with cutoffs the exponents of the degree distribution and the community-size distribution are also related as $\tau = \gamma - 1$.

Under a more strict assumption on the inter-community degrees

$$d_v^{(b)} \leq K, \quad (12.2.6)$$

we can also relate the power-law exponent of the intra-community degrees to the degree distribution. Assumption (12.2.6) implies $k \leq sK$, and therefore if the community-size distribution follows a power law with exponent $\gamma = \tau + 1$, then the distribution

of the community outside degrees cannot have a power-law distribution with exponent smaller than γ . Suppose we want to construct a graph where the degree distribution follows a power law with exponent $\tau \in (2, 3)$. One possibility to construct such a graph is to use the CM. However, the CM with $\tau \in (2, 3)$ has vanishing probability to create a simple graph. Another way to construct a graph with this degree distribution is to use the HCM with extremely dense communities of power-law size with exponent $\tau + 1$. The outside degrees of the communities then follow a power law with exponent at least $\tau + 1 > 3$. Since the outside degrees are paired according to the CM, the probability that the resulting graph is simple, will be larger than zero in the limit of infinite graph size. Thus, the HCM is able to construct a simple graph with degree power-law exponent $\tau \in (2, 3)$.

Another interesting application of this power-law shift is in the critical percolation value. It is well known that the critical percolation value $\pi_c = 0$ for a CM with $\tau \in (2, 3)$ [154]. When we use Theorem 11.2 on highly connected communities, close to the critical value π_c , these communities will still be connected after percolation, and $g(H, v, l, \phi) \approx 1$ for $l = k$. Then Theorem 11.2 reduces to $\pi_c = \mathbb{E}[D] / (\mathbb{E}[D(D-1)])$, as in the standard CM. Thus, if communities are highly connected compared to the inter-community edges, the critical percolation value is entirely determined by the inter-community edges. Since the inter-community degrees have exponent larger than 3, the HCM is able to construct random graphs with $\tau \in (2, 3)$ and $\pi_c > 0$. This shows that the HCM with extremely dense communities is in another universality class than the CM.

The role of hubs. A power-law degree distribution implies the existence of hubs: nodes with a very high degree. We now show that under assumptions (12.2.2) and (12.2.3) this only occurs in networks that are not very realistic. Since every vertex is inside a community in the HCM, the hubs also need to be assigned to some community. In these communities, hubs can have several roles, as observed in [103]. There are two possibilities, as shown in Figure 12.3. When most neighbors of the hub are also inside the community as in Figure 12.3a, then the hub is in a very large community. Assumption (12.2.2) states that most neighbors of the hub should also be connected to one another, and thus also have a high degree. However, in real-world networks this might not be realistic. For example, when one person in a social network has many friends, this does not mean that most of these friends are friends with one another. Hence, putting most neighbors of a hub inside the same community can create communities that are not dense. The other possibility (see Figure 12.3a) is to have only a small fraction of the neighbors inside a community. However, now the outside degree of the hub is large, which may contradict assumption (12.2.3) when the hub is in a small community. Therefore, it is not realistic to assume that networks have both hubs and extremely dense communities. Although it is theoretically possible to have networks with a dense community structure and power-law degrees, this only happens if the nodes with very high degrees are in big communities with one another, and have only a few links to nodes of lower degrees.

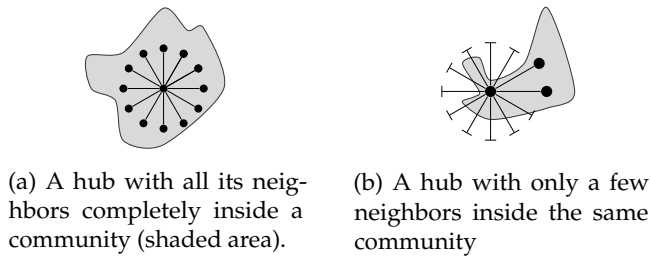


Figure 12.3: Illustration of the role of hubs.

	AMAZON	GOWALLA	WordNet	GOOGLE
S (data)	1,000	1,000	0,994	0,977
S (HCM)	1,000	1,000	0,994	0,978
S (CM)	0,999	0,993	0,999	0,997
γ	2,89	-	3,23	2,44
τ	3,31	2,48	2,46	-
α	0,15	0,31	0,21	0,21
β	1,14	1,18	1,28	1,24
C_{macro}	0,084	0,044	0,124	0,088
density	$1,6 \cdot 10^{-5}$	$4,9 \cdot 10^{-5}$	$6,2 \cdot 10^{-5}$	$1,1 \cdot 10^{-5}$

Table 12.1: Several characteristics of four different data sets

12.3 Real-world networks

In this section, we apply the HCM to four different data sets: an AMAZON co-purchasing network [225], the GOWALLA social network [60], a network of relations between English words [150] and a GOOGLE web graph [139]. To extract the community structure of the networks, we use the Infomap community detection method [190], a community detection method that performs well on several benchmarks [135]. Table 12.1 shows that Theorem 11.1 identifies the size of the giant component almost perfectly, in contrast to the value calculated by the CM. Furthermore, we see that the clustering coefficient of the macroscopic CM C_{macro} is quite low in three out of the four data sets; only the clustering coefficient of WordNet is higher. This implies that our assumption that the connections between communities are tree-like, even though it is not completely satisfied in real-world networks, still often is a reasonable approximation. In particular, HCM still allows for more deviation from the locally tree-like structure than the classical CM model.

Figure 12.4 illustrates the distributions of the degrees, the community sizes, and the community outside degrees for these data sets. While the outside degree distribution in Figures 12.4a and 12.4c clearly do not follow power laws, the degrees and the community sizes seem to follow power-law distributions, possibly with cutoffs. Tables 12.2 and 12.3 of Appendix 12.A show the goodness-of-fits of the power-law distributions for the degree distribution and the community-size distribution. In Appendix B we also test alternative distributions, using the methods of Clauset et

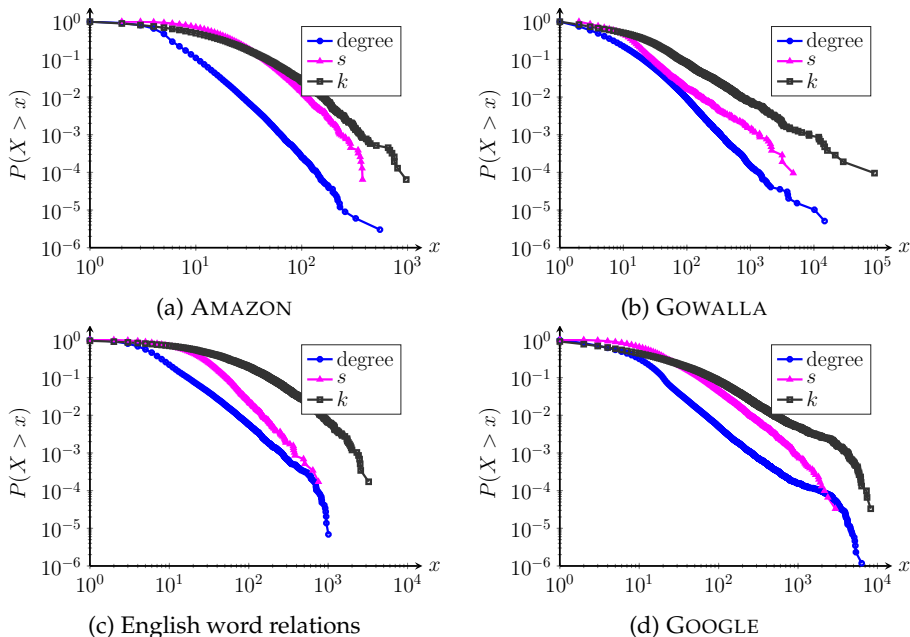


Figure 12.4: Tail probabilities of the degrees, community sizes s and the inter-community degrees k in real-world networks.

al. [66]. We see that most of the data sets can be approximated by either power laws, or power laws with an exponential cutoff. Table 12.1 presents the estimated power-law exponents of the degrees, τ , and the community sizes, γ . In the cases where the power law with exponential cutoff seems the more appropriate distribution, the exponent of this power-law distribution is shown. The degree distribution of the Google data set and the community-size distribution of the Gowalla data set do not show statistical evidence for a power-law distribution or a power-law distribution with an exponential cutoff. Therefore, Table 12.1 does not show exponents for these data sets. We see that a power-law shift is less pronounced than in the stylized household model, if existing at all. This indicates that the communities in the data sets do not have the intuitive extremely dense structure. Thus, we test assumption (12.2.2). The maximum number of edges inside a community is obtained if the community is a complete graph, in which case $e_{in} = \frac{1}{2}s(s-1)$. Dividing (12.2.2) by $s(s-1)/2$ gives $\frac{e_{in}}{s(s-1)/2} \geq 2\epsilon$. This fraction measures how dense a community is. Figure 12.5 plots the community size s against the average value of $2e_{in}/(s^2 - s)$, and the last row of Table 12.1 presents this measure of denseness for the entire network. For all networks, the density of a community is not independent of its size s . Larger communities are less dense than smaller communities. Therefore, the large communities do not satisfy the intuitive picture of an extremely densely connected subset. Still, Table 12.1 shows that the density within communities is much higher than the density of the entire network. This is a similar observation as in [139], where the authors discover that most real-world networks have a strongly connected core, which con-

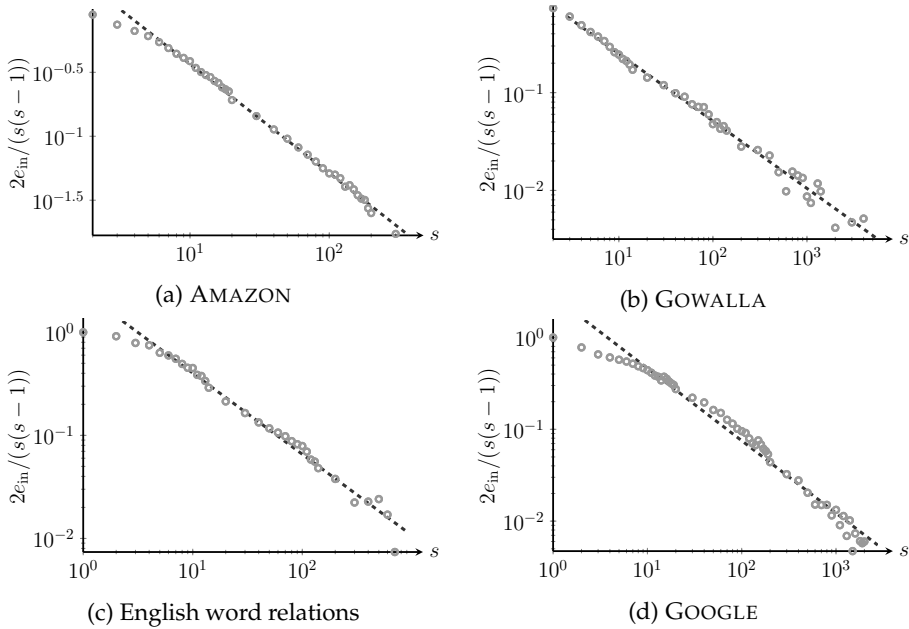


Figure 12.5: The relation between the denseness of a community $2e_{\text{in}}/(s^2 - s)$ and the community size s can be approximated by a power law.

sists of several interconnected communities that are hard to distinguish. The core is connected to the periphery, some isolated, densely connected small communities. This structure could explain the dependence of the density of the communities on s . The large communities that are not very dense are part of the core, whereas the small communities are the more isolated parts of the network. Another interesting property of the community structures in Figure 12.5 is the power-law relation between the community sizes and their densities, $e_{\text{in}} \approx cs^{\alpha+1}$. The AMAZON, GOWALLA and WordNet network have a clear power-law relation, whereas the GOOGLE network deviates more from its power-law approximation. Table 12.1 shows the values of α for these networks, using a least squares regression. In assumption (12.2.2), we assume that $\alpha = 1$. However, Table 12.1 shows that the example data sets have $\alpha < 1$. For this reason, we replace (12.2.2) by

$$e_{\text{in}} \geq \varepsilon s(s-1)^\alpha. \quad (12.3.1)$$

Now (12.2.5) still holds, but (12.2.4) needs to be modified. The average inside degree of a vertex now is $\varepsilon(s-1)^\alpha$. Since the maximum inside degree is $s-1$, there are at least $\varepsilon(s-1)^{\alpha-1}$ vertices of degree at least $\varepsilon(s-1)^\alpha$. A similar analysis as (12.2.4) yields

$$1 - P_j \geq \sum_{i \geq j} \sum_k r_{k,(i/\varepsilon)^{(1/\alpha)+1}} \varepsilon (i/\varepsilon)^{(\alpha-1)/\alpha}$$

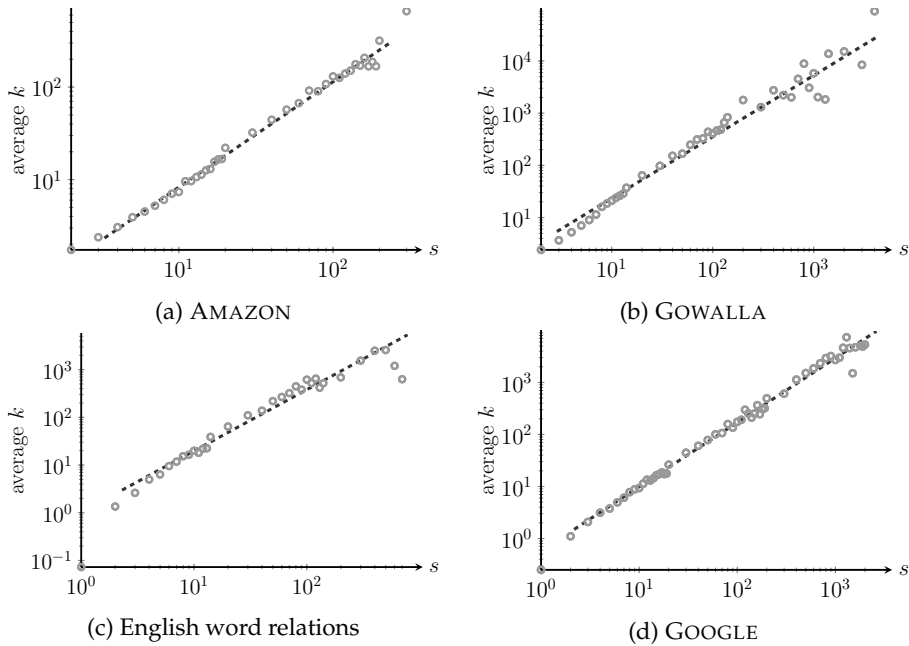


Figure 12.6: The outside degree of a community k follows a power-law relation with the community size s .

$$\begin{aligned}
 &= \frac{\varepsilon}{\mathbb{E}[S]} \sum_{i \geq j} \left(\left(\frac{i}{\varepsilon} \right)^{(1/\alpha)} + 1 \right)^{-\gamma+1} (i/\varepsilon)^{(\alpha-1)/\alpha} \\
 &\approx Cj^{-\gamma/\alpha+2}.
 \end{aligned} \tag{12.3.2}$$

Together with (12.2.5), this shows that the exponent τ of the degree distribution satisfies $\tau \in [\gamma - 1, \frac{\gamma}{\alpha} - 1]$. Table 12.1 shows several values of τ , γ and $\gamma/\alpha - 1$. We see that indeed $\tau \in [\gamma - 1, \frac{\gamma}{\alpha} - 1]$ in the example data sets. However, the interval may be quite wide when α is small.

We next test assumption (12.2.3). Interestingly, Figure 12.6 shows a power-law relationship between k and s , of the form $k \approx s^\beta$. If $k \leq Ks^2$ would hold, then $\beta \leq 2$, whereas the more strict assumption (12.2.6) would imply $\beta \leq 1$. Table 12.1 shows that the example data sets all have $1 < \beta < 2$. Therefore, the more strict assumption (12.2.6) does not hold, but (12.2.3) does hold. Thus, large communities have very large outside degrees.

12.4 Conclusions and outlook

In this chapter, we have analyzed the hierarchical configuration model (HCM) introduced in Chapter 11 for networks with a power-law degree distribution. For the caricature case of extremely dense communities, we show that a power-law degree distribution with exponent τ implies a power-law distribution for the community

sizes with exponent $\gamma = \tau + 1$. Real-world networks, however, rarely possess an extremely dense community structure [139].

Studying the HCM allows us to observe two previously unobserved power-law relations in several real-world networks. The relation between the number of edges inside a community e_{in} and the community sizes s follows a power law of the form $e_{\text{in}} \propto s^{1+\alpha}$. The second power-law relation is between the number of edges going out of a community k and the community sizes: $k \propto s^\beta$. The data sets that were studied in this chapter satisfy $1 < \beta < 2$ and $\alpha < 1$. Combined, the two power-law relations improve our understanding of the community structure in the data sets. Large communities are not extremely densely connected, and have a large number of edges going out of the community per vertex. Smaller communities are dense, and vertices in the community only have a few edges going out of the community. Although the communities are still denser than the entire network, our intuitive picture of extremely densely connected communities only holds for the small communities in a network, the larger communities do not fit into this picture. The observation that large communities are not extremely dense may be a consequence of not allowing for *overlapping* communities. In case of several overlapping communities, community detection algorithms may merge these communities into one large community. As a consequence, this large community will be far from extremely dense. In the case of overlapping communities, many networks still display a power-law community-size distribution [173]. It would be interesting to investigate the relation between the exponent of the degree distribution and the community-size distribution when communities are allowed to overlap. Further research could also study how the denseness of the communities and the number of edges out of the communities are related to the community sizes in the case of overlapping communities.

Both power-law relations are observational, and therefore depend on the Infomap community detection algorithm. It is also possible to use other community detection algorithms to investigate these power-law relations. We found that when using the Louvain community detection algorithm [29], the power-law relations still hold. The estimates for the exponents α and β however did change. This can be explained by the fact that the Louvain method finds larger communities in general, which are therefore less dense.

The power-law exponent of the degree distribution τ is known to influence the behavior of various processes on random graphs, for example percolation or epidemic models. Furthermore, mean distances in random graphs are different for $\tau \in (2, 3)$, or $\tau > 3$ [108, 109, 164]. Our results do not shed light on whether for networks with a community structure, the behavior of these processes is explained by τ , or rather by the community degrees. This remains open for further research. The results on the power-law shift suggest that this may depend on the density of the communities, which is characterized by the exponent α .

12.A Comparing the power-law fit to other distributions

Tables 12.2 and 12.3 show the goodness of fit criterion (g.o.f.) developed in [66] for the power laws of the degree distribution and the community-size distribution. If the goodness of fit is larger than 0.10, then the power-law assumption seems plausible.

We compare the fit of the power laws with the fit of a lognormal distribution and a power-law distribution with an exponential cutoff by computing likelihood ratio tests as in [66]. A negative likelihood ratio suggests a better fit than the pure power law, whereas a positive value of the likelihood ratio implies that a power law is a better fit. For the degree distributions, we conclude that the Gowalla data set follows a power law, the Amazon and WordNet degree distributions can be approximated by power laws with cutoffs. The degree distribution of the Google data set does not show evidence for a power law, a lognormal distribution or a power-law distribution with exponential cutoff. For the community-size distribution, we conclude that the WordNet data set can be approximated by a power law. The Amazon and Google community-size distributions are better approximated by power laws with exponential cutoffs. The community-size distribution of the Gowalla set does not show evidence for a lognormal distribution, a power law or a power law with exponential cutoff.

	power law	lognormal		cutoff	
	g.o.f.	LR	p	LR	p
Amazon	0,02	-2,76	0,01	-14,37	0,00
Gowalla	0,80	2,24	0,02	0,00	0,98
WordNet	0,00	4,38	0,00	-61,65	0,00
Google	0,00	10,32	0,00	-1,27	0,26

Table 12.2: Fitting the degree distribution. The first column presents the goodness of fit criterion (g.o.f.) from [66]. The next columns present the likelihood ratios (LR) for the degree distributions: lognormal distribution and a power-law distribution with exponential cutoff versus a power-law distribution. The p -values for significance of the likelihood ratio tests, as in [66] are also shown. Statistically significant values are shown in bold.

	power law	lognormal		cutoff	
	g.o.f.	LR	p	LR	p
Amazon	0,25	-1,98	0,05	-4,74	0,03
Gowalla	0,02	4,95	0,00	-1,33	0,25
WordNet	0,60	0,04	0,00	-0,39	0,53
Google	0,06	-1,35	0,18	-5,59	0,02

Table 12.3: Fitting the community-size distribution. Goodness of fit and likelihood ratios as in Table 12.2 for the community-size distribution. Statistically significant values are again bold.

13 Epidemics on networks with community structures

Based on:
Epidemic spreading on complex networks with community structures
C. Stegehuis, R. van der Hofstad and J.S.H. van Leeuwaarden
Scientific Reports 6 p. 29748 (2016)

In this chapter, we study the performance of two random graph models that create a network with similar community structure as a given network on real-world network data. One model, the HCM introduced in Chapter 11, preserves the exact community structure of the original network, while the other model only preserves the set of communities and the vertex degrees. Testing these models on real-world network data shows that community structure is an important determinant of the behavior of percolation processes on networks, such as information diffusion or virus spreading: the community structure can both *enforce* as well as *inhibit* diffusion processes. In Chapter 11, we already showed that for bond percolation and stylized community shapes both effects may be present, depending on the community structure. In this chapter, we show that both effects can also occur in real-world community structures and for other types of epidemic processes. Our models further show that it is the mesoscopic set of communities that matters. The exact internal structures of communities barely influence the behavior of percolation processes across networks. This insensitivity is likely due to the relative denseness of the communities.

13.1 Introduction

The behavior of dynamic processes such as percolation or epidemic models on networks are of significant interest, since for example they model the spreading of information or a virus across a network [18, 32, 76, 178]. Understanding models for percolation may enhance insight in how an epidemic can be stopped by immunization, or how a message can go viral by choosing the right initial infectives. An important question is how the structure of the network affects the dynamics of the epidemic [163]. A vast amount of research focuses on scale-free networks that possess a power-law degree distribution [66, 106, 158, 162, 212]. The exponent τ of the degree distribution was found to play a central role in various percolation processes [22, 23, 55, 67, 177]. Other authors have focused on the influence of clustering on the spread of epidemics [93, 95, 197, 198, 206].

Real-world networks, however, are not completely characterized by their microscopic and macroscopic properties. Many real-world networks display a community

structure [92], where groups of vertices are densely connected, while edges between different groups are more scarce. Since communities are small compared to the entire network, but seem to scale with the network size, they are typically of mesoscopic scale [80, 181]. The problem of detecting the community structure of a network has received a lot of attention [80, 138]. The exact way in which communities influence the properties of a network is a different problem. For example, the community structure of a network influences the way a cooperation process behaves on real-world networks [143], and using community structure improves the prediction of which messages will go viral across a network [215]. Several stylized random graph models with a community structure have shown that communities influence the process of an epidemic across a network [13, 46, 94, 113, 142, 194, 221, 224], but the extent to which community structure affects epidemics on real-world networks is largely unexplained. Our main goal is to enhance our understanding of the intricate relation between community structures and the spread of epidemics, and in particular to identify the properties of community structures that have the largest influence.

We study two random graph models that generate networks with a similar community structure as any given network: the HCM model introduced in Chapter 11 and a second model that also randomizes the internal community structure. In Chapter 11 we have studied the behavior of bond percolation on the HCM model analytically. In this chapter, we study the behavior of several other types of epidemics by simulation, and compare the results to the behavior of epidemics on real-world networks. We find that these models capture the behavior of epidemics or percolation on real-world networks accurately, and that the mesoscopic community structure is vital for understanding epidemic spreading. We find that the sets of communities are of crucial importance, while quite surprisingly, the precise structure of the intra-community connections hardly influences the percolation process. Furthermore, we find that community structure can both enforce as well as inhibit percolation.

13.2 Models

We now describe our two random graph models in detail. For a given real-world network, both models randomize the edges of the network, while keeping large parts of the community imprint. Suppose that we are given the set of communities of a particular real-world network. Then the first model, the hierarchical configuration model (HCM) as introduced in Chapter 11, keeps all edges inside the communities, while rewiring the inter-community edges. Indeed, all inter-community edges are replaced by two half-edges, one at each end of an inter-community edge. Then, one by one, these half-edges are paired at random. Thus, in HCM, the precise community structure of the network is the same as in the original network, but the inter-community connections are random. The second model (HCM*), introduced as the modular random graph in [193], replaces both the inter-community edges and the intra-community edges by pairs of half-edges. Then again, the half-edges are paired at random. An additional constraint is that all inter-community half-edges must be paired to one another, and all half-edges corresponding to the same community must be paired to one another (see Figure 13.1). Thus, a network generated by HCM* is completely random, except for the set of communities and the degree distributions

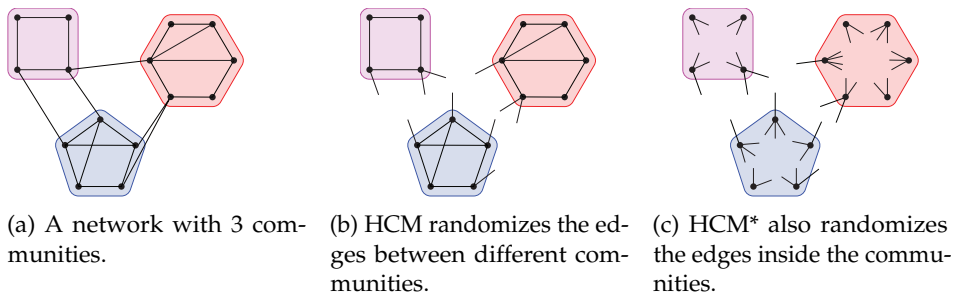


Figure 13.1: HCM and HCM* illustrated.

inside and outside the communities.

More precisely, given a real-world network and the collection of its communities, obtained e.g., using a community detection algorithm, we construct HCM and HCM* in the following way. First, rewire the edges between different communities, using the switching algorithm. Select two inter-community edges uniformly at random, $\{u, v\}$ and $\{w, x\}$. Now delete these edges and replace them by $\{u, x\}$, $\{w, v\}$ if this results in a simple graph. Otherwise keep the original edges $\{u, v\}$ and $\{w, x\}$. Empirically, it has been shown that this randomizes the inter-community edges uniformly if this procedure is repeated at least $100E$ times, where E is the number of inter-community edges [151]. This creates HCM. This construction is not precisely the same as its construction in Chapter 11, since this construction requires HCM to be simple, whereas the construction in Chapter 11 may result in a multigraph. Therefore, the edges of HCM are not quite connected as in the CM, since self-loops and multiple edges are forbidden. To create HCM*, the edges within the communities are also randomized after rewiring the inter-community edges, again using the switching algorithm. This is repeated for all communities.

HCM and HCM* are extensions of the configuration model (CM), described in Section 1.1.1. CM only preserves the microscopic degree distribution of the real-world networks, while HCM* also preserves the mesoscopic community structure. HCM instead, preserves the entire community structure. Therefore, if we sort the random graph models in decreasing randomness, we first have CM, then HCM*, and then HCM. When comparing the behavior of an epidemic process on these random graphs to the original network, we see how much of the behavior of epidemics on real-world networks can be explained by its degree distribution (CM), its rough community structure (HCM*), or the exact community shapes (HCM). The aim of this chapter is to investigate to which extent microscopic and mesoscopic network properties determine the spread of epidemics.

In Chapter 11 we have shown that the fixed community shapes combined with the randomized inter-community connections make HCM analytically tractable. However, keeping all intra-community edges fixed makes HCM prone to overfitting. HCM* does not have this problem and is more suitable to generate a random network with a community structure, since all edges within communities are randomized. Randomizing the intra-community edges makes HCM* harder to analyze analytically than HCM. Some analytical results of HCM such as the largest component size,

however, can be extended to results of HCM*.

13.3 Results

We analyze six different real-world networks: the internet on the Autonomous Systems level [137], an email network of the company Enron [129, 137], the PGP web of trust [36], a collaboration network in High energy physics, extracted from the arXiv [137], a FACEBOOK friendship network [214] and an interaction network between proteins in yeast [53]. Table 13.1 shows several statistics of these data sets and their community structures. We extract the communities of these networks with the Infomap community detection algorithm [190], and use these communities as input for the HCM and HCM* model, to create networks with a similar community structure as the original networks. Table 13.1 shows that the communities are of mesoscopic size: while the communities are small compared to the entire network, and have a small expected size, all networks still contain a few large communities.

	N	$\mathbb{E}[s]$	s_{\max}	δ_{netw}	δ_{com}	δ_{com}^w
AS	11,174	21	910	$3.75 \cdot 10^{-4}$	0.38	0.10
Enron	36,692	15	1,722	$2.73 \cdot 10^{-4}$	0.73	0.22
HEP	9,877	10	181	$5.33 \cdot 10^{-4}$	0.59	0.32
PGP	10,680	12	160	$4.26 \cdot 10^{-4}$	0.41	0.24
FB	63,731	29	2,247	$4.02 \cdot 10^{-4}$	0.41	0.14
yeast	2,361	9	97	$2.57 \cdot 10^{-3}$	0.55	0.25

Table 13.1: Statistics of the data sets. N is the number of vertices in the network, $\mathbb{E}[s]$ the average community size, s_{\max} the maximal community size. The denseness of the network δ_{netw} is defined as the number of edges divided by the number of edges in a complete graph of the same size. δ_{com} equals the average denseness of the communities, and δ_{com}^w the average denseness of the communities weighted by their sizes.

Giant component size. An important property of a network is its connectedness, expressed by the fraction of vertices in the largest component. For HCM, the size of the largest component has been derived analytically in Theorem 11.1. This size is independent of the precise community shapes, and therefore is the same for HCM and HCM*, as long as the communities of HCM* remain connected. The size of the largest component of real-world networks can be well predicted using the analytical estimates of HCM, which only uses the joint distribution of community sizes and the number of edges going out of the communities (Table 13.2). These estimates yield a considerable improvement compared to CM, which is generally a few percent off.

The fact that the analytical estimate for the size of HCM* accurately estimates the giant component size in real-world network data suggests that communities generated by HCM* generally remain connected. Table 13.3 presents the fraction of disconnected communities f_{dis} that HCM* generates. Table 13.3 also presents N_{dis} , the average number of vertices that are not connected to the largest component of

	S (data)	S (HCM)	S (HCM*)	S (CM)
AS	1.000	1.000	1.000	0.960
Enron	0.918	0.918	0.918	0.990
HEP	0.875	0.875	0.875	0.990
PGP	1.000	1.000	1.000	0.960
FB	0.995	0.995	0.995	0.999
yeast	0.941	0.941	0.941	0.948

Table 13.2: The size S of the giant component in the data sets compared to the analytical estimates of HCM (Theorem 11.1) and CM [117].

the community after rewiring, given that the community is disconnected. We see that the fraction of disconnected communities is different for the different networks. For the networks with a more dense community structure, the probability that a community becomes disconnected after rewiring is low, while for for example the AS network this probability is higher. In all cases, the number of vertices that are disconnected from the largest component is low, indicating that the community stays largely connected. Thus, Theorem 11.1 can also be used to approximate the size of the largest component of HCM*.

Community structure of HCM and HCM*. By keeping the sets of communities fixed, we expect both HCM and HCM* to generate networks with a similar community structure as the original data set. We first extract the communities $(\mathcal{C}_i)_{i \geq 1}$ from the data sets. We use these communities to generate HCM and HCM*. To test how similar the community structures of the generated networks and the original networks are, we once more perform a community detection algorithm on the networks generated by HCM and HCM*. Let the communities detected by this algorithm be denoted by $(\mathcal{C}_i^{\text{HCM}})_{i \geq 1}$. We then define the similarity w_{HCM} of the two community structures as

$$w_{\text{HCM}} = \frac{1}{N} \sum_i \frac{|\mathcal{C}_i \cap \mathcal{C}_i^{\text{HCM}}|}{|\mathcal{C}_i|}, \quad (13.3.1)$$

where \mathcal{C}_i and $\mathcal{C}_i^{\text{HCM}}$ are the sets of vertices that are in the same community as vertex i in the original network, and the HCM network respectively. We define the similarity for the community structure generated by HCM* similarly as w_{HCM^*} . Table 13.3 presents this similarity measure for all networks. We see that for most networks, the degree of overlap is large, but for the AS network, the overlap between the original network community sets and the networks generated by HCM or HCM* is smaller. This may be explained by the fact that the AS network has less dense communities, so that rewiring the edges between communities can easily shift vertices from one community to the other.

Overfitting. Table 13.4 shows the fraction of edges of the data sets that are inside communities. HCM fixes all these edges, so one could argue that HCM overfits the data by keeping this fraction of edges fixed. For this reason, we also consider HCM*. Figure 13.2 shows the fraction of rewired edges in HCM* in communities of

	f_{dis}	N_{dis}	w_{HCM}	w_{HCM^*}
AS	0.24	3.00	0.68	0.65
Enron	0.02	3.62	0.94	0.92
HEP	0.04	2.23	0.96	0.94
PGP	0.17	2.54	0.97	0.91
FB	0.17	2.61	0.93	0.92
yeast	0.11	2.29	0.85	0.81

Table 13.3: The fraction of disconnected HCM* communities f_{dis} , the average number of vertices inside a disconnected community not connecting to the largest part of the community N_{dis} and the overlap of community structure of generated networks and original networks w_{HCM} and w_{HCM^*} as defined in (13.3.1).

size s . This is the fraction of edges that are different from the edges in the original community after the rewiring procedure inside communities. In general, a large fraction of edges is different after randomizing the intra-community edges. The cases where only a few edges were rewired correspond to small communities, where only a small amount of simple random graphs with the same degree distribution exist, or larger communities that are complete graphs, or star-shaped (where only one simple graph with that degree distribution exists). This shows that HCM* creates substantially different graphs than HCM, and is less prone to overfitting the data than HCM.

AS	Enron	Hep	PGP	FB	yeast
0.58	0.58	0.70	0.83	0.54	0.52

Table 13.4: The fraction of edges inside communities in the data sets.

Epidemics on HCM and HCM*. The long-term properties of an epidemic outbreak can be mapped into a suitable bond percolation problem. In this framework, the probability p that a link exists is related to the probability of transmission of the disease from an infected vertex to a connected susceptible vertex. The latter corresponds to removing edges in a network with probability $1 - p$ and keeping the edges with probability p independently across edges (other types of epidemics are discussed in Appendix 13.A). A quantity of interest is the size of the largest component as a function of p , which we described analytically for HCM in Chapter 11. However, this size depends on the community shapes, and therefore bond percolation on HCM does not necessarily give the same results as percolation on HCM*. Inspired by the insensitivity of the giant component to the exact community shapes, we establish whether the community shapes significantly influence the size of the giant percolating cluster by simulation, by showing how bond percolation affects the connectivity of the original networks, compared to CM, HCM and HCM* (Figure 13.3).

We see that the behavior of the real-world networks under bond percolation is captured accurately by both HCM and HCM*, in contrast to CM. In Appendix 13.A, we

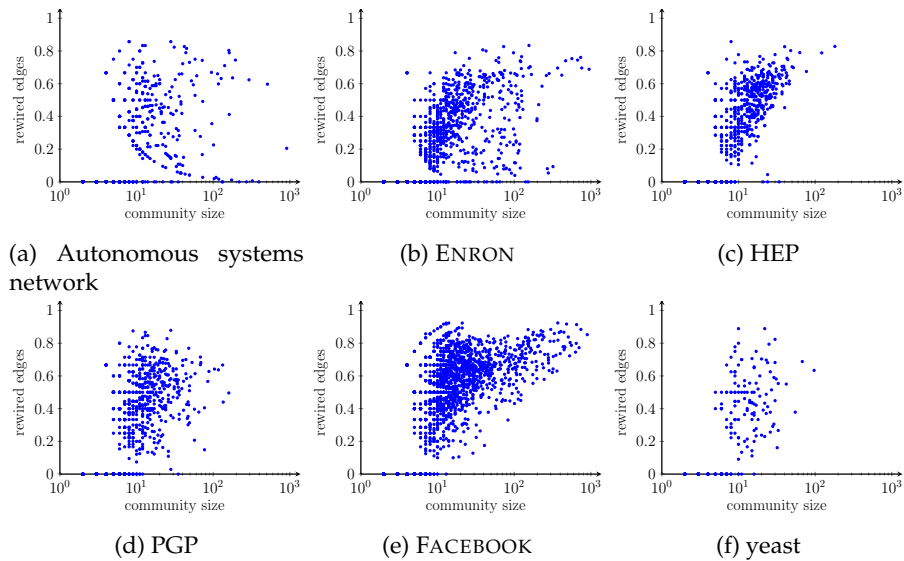


Figure 13.2: The fraction of rewired edges inside communities for HCM*. Every dot corresponds to a community. The fraction of rewired edges is the fraction of edges in the community that are present after randomizing the intra-community edges, but were not present before randomizing.

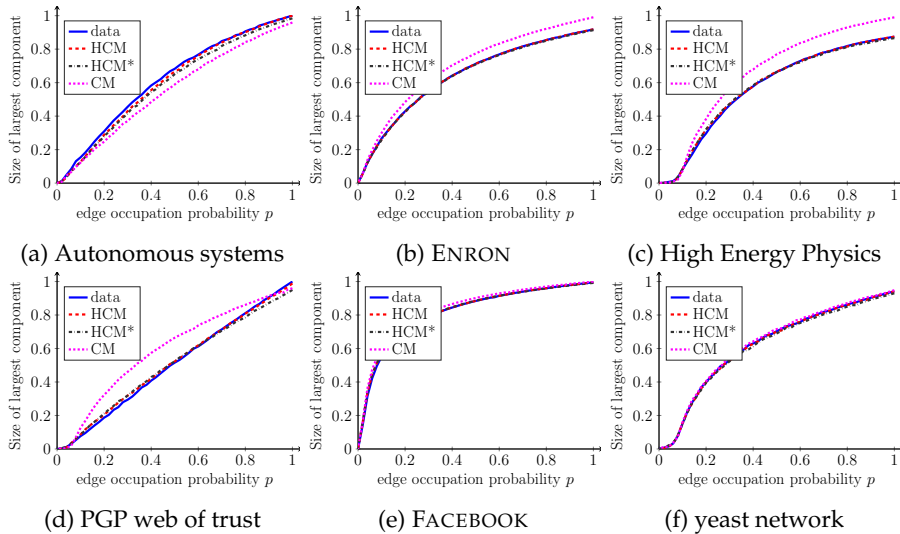


Figure 13.3: HCM, HCM* and CM under bond percolation compared to real-world networks. Independently, each edge is deleted with probability $1 - p$. The size of the largest component after deleting the edges is the average of 500 generated graphs.

see that HCM and HCM* also perform well for other types of percolation processes and an SIR epidemic. These results reveal and confirm the key role of the mesoscopic community structure in percolation processes. Furthermore, the fact that the predictions of HCM and HCM* are *both* close to the behavior of the original network under percolation indicates that the shapes of the communities only have a minor influence on the percolation process. The surprising finding that the exact internal community structure barely influences the epidemic processes may be explained by the denseness of the communities. Table 13.1 shows that the communities are very dense compared to the entire network. Since community detection algorithms look for dense subsets in large complex networks, applying HCM or HCM* to real-world networks typically yields sets of dense communities. The Autonomous Systems network has communities that are much less dense than in most other networks [136], but even in that network the communities are much denser than the entire network. Therefore, in the case of bond percolation for example, the communities of mesoscopic size are supercritical, and the communities will be almost connected after percolation. Thus, an epidemic entering a community of mesoscopic size will reach most other community members. It is more difficult for the epidemic to reach other communities, which makes the inter-community edges the important factor for the spread of an epidemic. When generating a HCM* network, the communities stay of the same denseness, and therefore it is still relatively easy for the epidemic to spread inside the communities, regardless of their exact shapes.

The only process where HCM and HCM* are not always close to the process on the original graph, is a targeted attack (Figure 13.6), even though both models still outperform CM. Furthermore, some networks show a difference between the predictions of HCM and HCM*. Therefore, the exact community structures may have some influence on a targeted attack on a real-world network. Another interesting observation is that where most networks are highly sensitive to a targeted attack, the FACEBOOK network has a community structure that makes it more resistant against a targeted attack than a configuration model. This particular feature of the FACEBOOK network can be explained by the fact that in the FACEBOOK network, most vertices of high degree are in the same community. Therefore, deleting high-degree vertices has a smaller effect than in a corresponding CM model.

The results of the yeast network show that in some situations CM performs equally well as HCM or HCM*. Thus, in some cases the mesoscopic properties of a network do not influence percolation processes. In the case of the yeast network, this can be explained by its almost tree-like structure; there is no noticeable community structure. Thus, by adding the community structure in HCM or HCM*, no structural information is added. This suggests that CM, HCM and HCM* combined can also show whether the community structure given by a community detection algorithm is meaningful. When the behavior of various processes on CM, HCM and HCM* are similar, this may imply that there is no real community structure in the network.

The ENRON, High Energy Physics and PGP networks have communities that inhibit percolation or an SIR epidemic compared to a configuration model with the same degree distribution. This is similar to the observation that communities can act as traps for an epidemic process across a network [168]. In contrast, the communities in the Autonomous Systems graph enforce the percolation process, which may be attributed to its star-like community structure. Since HCM* preserves the degrees

of the vertices inside their own community, HCM* creates a graph that captures this star-like structure.

An important conclusion is that these findings confirm that both HCM and HCM* are realistic models for real-world networks.

Microscopic and macroscopic properties of HCM and HCM* Where [170] creates a reshuffling of a given network using several microscopic properties of every vertex, HCM and HCM* use mesoscopic properties instead. An advantage of using HCM or HCM* is that both models are easy to generate. Since HCM* is more random than HCM, it is a better choice for generating a random network. In HCM*, the rewiring of intra-community edges makes the community structure a uniform simple graph with the prescribed degrees. Specifically, if the interest is to generate a random graph such that percolation on that graph behaves in a similar way as in the original network, then our results show that HCM* is a suitable choice. However, HCM* does not capture the microscopic properties of the original network as effectively as HCM. HCM*, for example, does not generate networks with similar clustering as in the original network [193]. Therefore, when the goal is to create a network with similar clustering as the original network, using HCM* may be less suitable. Indeed Table 13.5 shows that in most cases HCM generates a network with a clustering coefficient that is closer to the value of the original network. An exception is the Autonomous Systems network, where HCM* is closer to the real value of the clustering. An explanation for this is that the communities in the Autonomous Systems network have virtually no clustering; all clustering is between different communities. HCM also has no clustering inside the communities, but the pairing between different communities destroys the clustering between different communities, and therefore HCM creates a network with a lower clustering coefficient. HCM* also destroys the clustering between different communities, but by rewiring the edges inside communities, creates some clustering inside the communities. Therefore, the value of the clustering of HCM* is closer to the value of the original network than the one of HCM.

Table 13.6 shows that HCM and HCM* generate networks that match the assortativity of the original network closer than a configuration model. However, the assortativity generated by HCM does not always match its theoretical value of Proposition 11.2. An explanation for this is that HCM generates simple graphs, while the theoretical estimate does not take this into account. Since both ends of a self-loop have the same degree, having non-simple graphs increases the assortativity. Furthermore, these self-loops typically occur at nodes of large degree, increasing the assortativity even further, so that the theoretical assortativity is higher than the observed assortativity.

The fact that HCM* does not capture the clustering coefficient and the assortativity well, but does capture the spread of an epidemic across a network, again confirms that the mesoscopic properties are of vital importance for the spread of an epidemic across a network. Even though microscopic features such as clustering are destroyed in HCM*, the mesoscopic properties are sufficient to know how an epidemic spreads, making HCM* a suitable random graph model when considering the mesoscopic structure of networks.

Figure 13.4 presents the graph distances for the different data sets. In some

	data	HCM	HCM*	CM
AS	0.30	0.16	0.20	0.09
Enron	0.50	0.35	0.22	0.03
HEP	0.47	0.40	0.24	0.00
PGP	0.26	0.24	0.19	0.00
FB	0.22	0.15	0.08	0.00
yeast	0.13	0.12	0.12	0.01

Table 13.5: Average clustering for the original data set, HCM, HCM* and CM. The presented values are averages of 100 generated graphs.

	data	HCM	HCM*	CM	HCM (theory)
AS	-0.19	-0.16	-0.16	-0.14	0.00
Enron	-0.11	-0.06	-0.05	-0.05	-0.02
HEP	0.27	0.25	0.23	0.00	0.25
PGP	0.24	0.26	0.26	-0.01	0.26
FB	0.18	0.11	0.10	0.00	0.11
yeast	-0.10	-0.03	0.00	-0.01	-0.02

Table 13.6: Assortativity of HCM, HCM* and CM compared to the real network and theoretical HCM value of Proposition 11.2. The values of HCM, HCM* and CM are averages over 500 generated graphs.

instances, HCM and HCM* capture the graph distances better than CM. However, for example for the yeast network, the distances in CM are already close to the distances in the original data set.

13.4 Conclusion

Community structures in real-world networks have a profound impact on percolation or epidemic spreading, which is central to our understanding of dynamical processes in complex networks. In this chapter, we have investigated the behavior of several epidemic processes in the HCM model introduced in Chapter 11 and its randomized counterpart HCM*. Both HCM and HCM* turn out to be highly suitable to capture epidemic spreading on real-world networks. We have shown this by mapping the models to various real-world networks, and by investigating a range of epidemic processes including bond percolation, bootstrap percolation and a SIR epidemic. Our experiments show that while it is essential to take the community structure into account, the precise internal structure of communities is far less important for describing an epidemic outbreak. This insensitivity is likely due to the relative denseness of the communities. When communities are sparse, their internal structures are expected to have a more decisive effect on epidemic spreading.

13.A Other epidemic processes

In Figure 13.3 we compared the behavior of bond percolation on real-world network data to bond percolation on HCM, HCM* and the CM. In Figures 13.5- 13.9 we now present similar results for four other types of epidemic processes. Here we describe these processes in detail.

Site percolation. In site percolation, every vertex, and all edges adjacent to it, are deleted with probability $1 - p$, independently for every vertex. As in bond percolation, we are interested in the fraction of vertices in the largest component after this deletion process.

Targeted attack. In a targeted attack, a fraction of p of the vertices and the edges adjacent to them are removed, starting with the highest degree vertex, then the second highest degree vertex and so on. Again, the quantity of interest is the fraction of vertices in the giant component after deleting the edges.

Bootstrap percolation. In bootstrap percolation with threshold t initially a certain fraction of vertices is infective. The initially infected vertices are selected at random. Then, every vertex with at least t infected neighbors also becomes infected. This process continues until no new vertices become infected anymore. In the results, we consider bootstrap percolation with threshold $t = 2$. The quantity of interest is the fraction of infected vertices when the process has stopped.

SIR epidemic. In an SIR epidemic, vertices are either susceptible, infected or recovered. One vertex is selected uniformly at random to be the initial infective. Then, every infected vertex infects his susceptible neighbors independently at rate β . Every infected vertex recovers at rate γ . As in [194], we set $\gamma = 1$ and $\beta = 3\mathbb{E}[D] / \gamma$, where $\mathbb{E}[D]$ is the average degree of the network. We are interested in how the fraction of infected and recovered vertices evolves over time. Since every vertex is either susceptible, infected or recovered, the fraction of susceptible vertices is then also known.

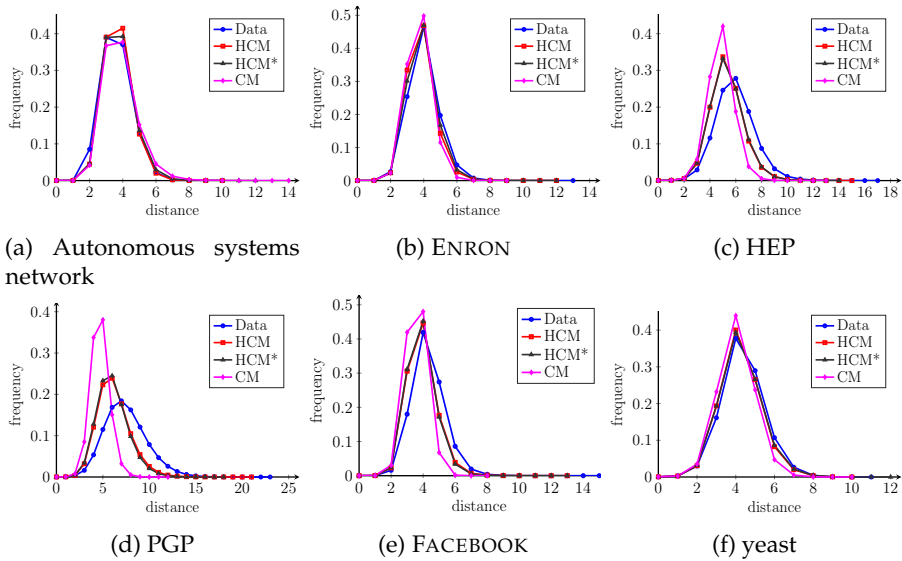


Figure 13.4: Distances in the original network, HCM, HCM* and CM. Distances are approximated by sampling 5,000 nodes from the graphs, and calculating all distances between pairs of nodes in the sampled set. The values for HCM, HCM* and CM are the average over 100 generated graphs.

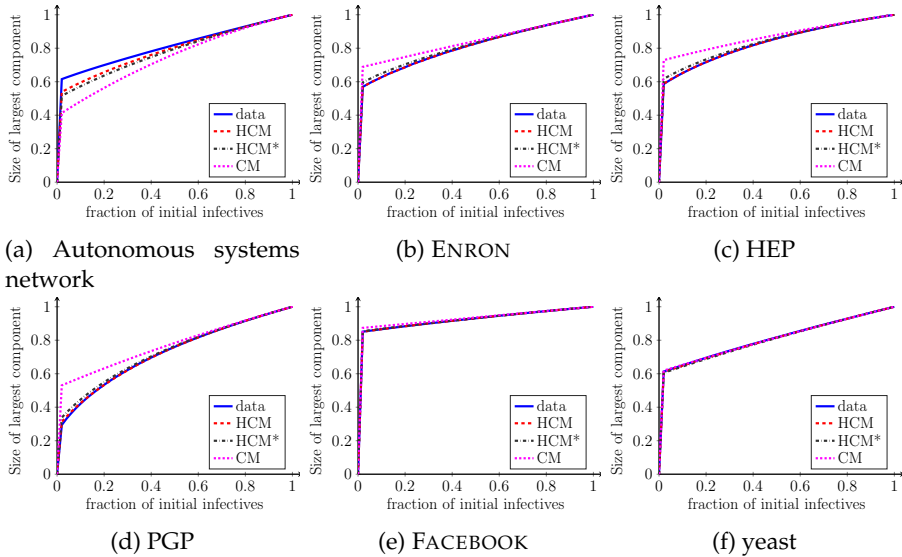


Figure 13.5: HCM, HCM* and CM under bootstrap percolation compared to real-world networks. Initially, a certain fraction of the vertices is infected at random. Then, a vertex becomes infected when at least 2 of its neighbors are infected. The final fraction of infected vertices is the average of 500 generated graphs.

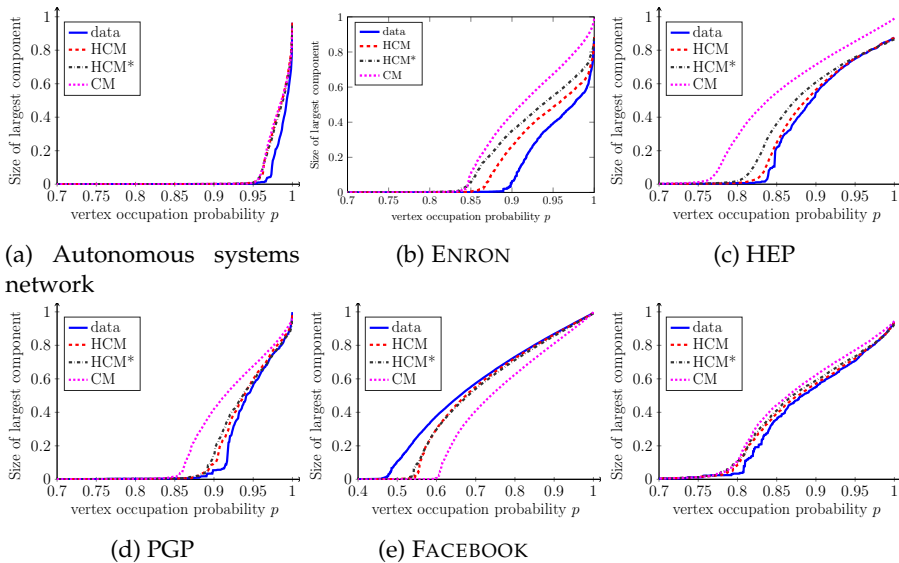


Figure 13.6: HCM, HCM* and CM under a targeted attack, compared to real-world networks. The fraction of $1 - p$ vertices of highest degree are removed. The size of the largest component after the vertices are removed is the average of 500 generated graphs.

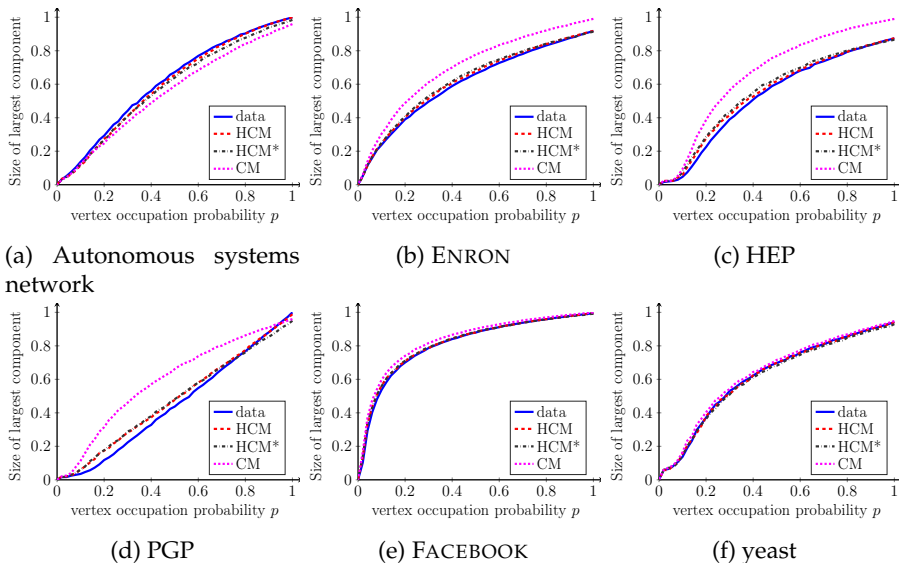


Figure 13.7: HCM, HCM* and CM under site percolation compared to real-world networks. Independently, every vertex is removed from the network with probability $1 - p$. The size of the largest component after the vertices are removed is the average of 500 generated graphs.

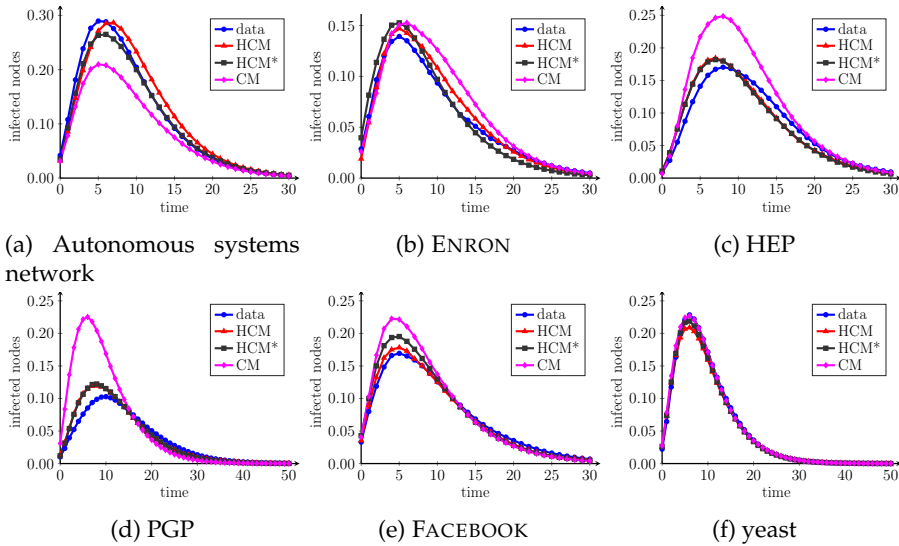


Figure 13.8: The number of infected individuals in an SIR epidemic in HCM, HCM* and CM compared to real-world networks. The presented results are the average of 500 generated graphs, with recovery rate $\gamma = 1$ and infection rate $\beta = 3\mathbb{E}[D] / \gamma$, where $\mathbb{E}[D]$ is the mean degree.

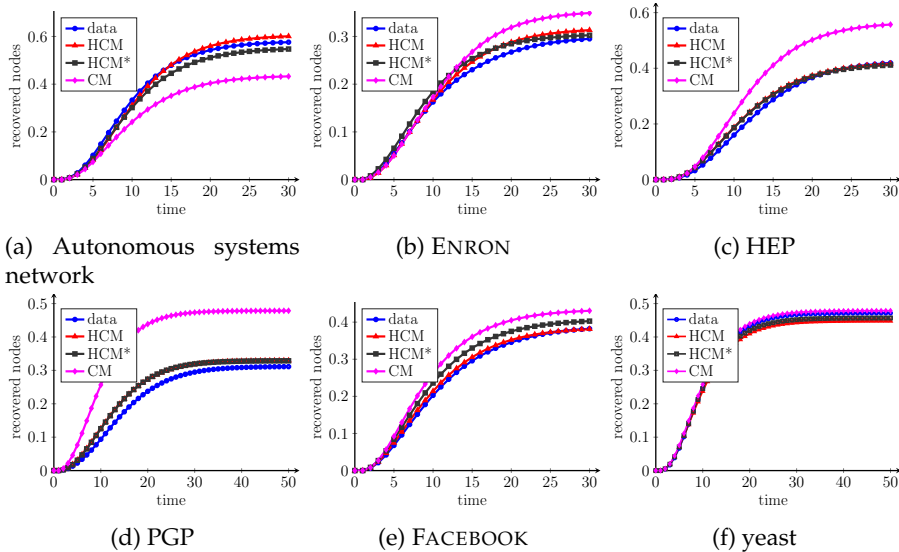


Figure 13.9: The number of recovered individuals in an SIR epidemic in HCM, HCM* and CM compared to real-world networks. The presented results are the average of 500 generated graphs, with the recovery rate $\gamma = 1$ and the infection rate $\beta = 3\mathbb{E}[D] / \gamma$, where $\mathbb{E}[D]$ is the mean degree.

14 Mesoscopic scales in hierarchical configuration models

Based on:
Mesoscopic scales in hierarchical configuration models
R. van der Hofstad, J.S.H. van Leeuwen and C. Stegehuis
Stochastic Processes and their Applications (2018)

In this chapter, we again study the hierarchical configuration model (HCM) introduced in Chapter 11, to better understand the mesoscopic scales of the communities. In Chapter 11 we investigated the component sizes when the HCM is in the supercritical or in the subcritical regime and we derived the component sizes after supercritical and subcritical bond percolation. We expand upon these results by investigating the scaling of the component sizes at criticality and by studying critical bond percolation. We find the conditions on the community sizes under which the critical component sizes of HCM behave similar to the component sizes in a critical configuration model. We show that the ordered components of a critical HCM on N vertices are $O(N^{2/3})$. More specifically, the rescaled component sizes converge to the excursions of a Brownian motion with parabolic drift.

14.1 Introduction

In networks, it is common to distinguish between two levels, referred to as “microscopic” and “macroscopic”. Vertex degrees and edges between vertices provide a microscopic description, whereas most network functionalities require a macroscopic picture. Random graph models are typically defined at the microscopic level, in terms of degree distributions and edge probabilities, leading to a collection of local probabilistic rules. This provides a mathematical handle to characterize the macroscopic network functionality related to global characteristics such as connectivity, vulnerability and information spreading.

Intermediate or “mesoscopic” levels are less commonly considered in random graph models and network theory at large, and apply to substructures between the vertex and network levels. Mesoscopic levels are however becoming increasingly in focus, for example because of community structures or hidden underlying hierarchies, common features of many real-world networks. It is not easy to define what is precisely meant with mesoscopic, apart from the obvious definition of something between microscopic and macroscopic. This chapter deals with large-network limits, in which the network size N (number of vertices) will tend to infinity. The mesoscopic

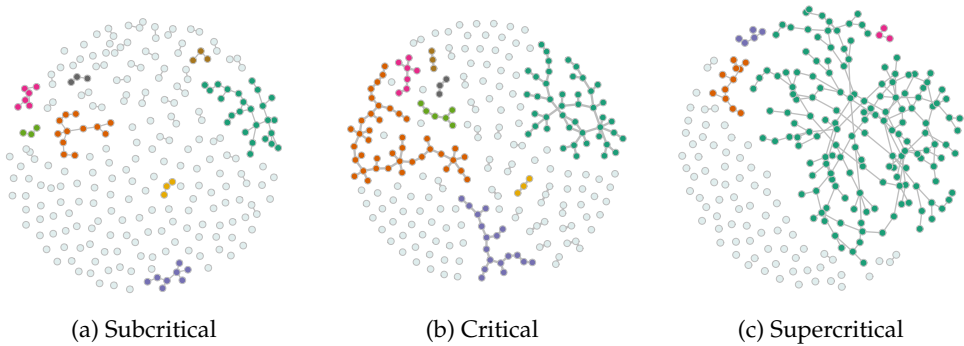


Figure 14.1: Phase transition for the component sizes in a configuration model

scale then naturally refers to structures of size N^α , where it remains to be determined what values of α need to be considered.

We will associate the mesoscopic level with the community structure, defined as the collection of subgraphs with dense connections within themselves and sparser ones between them. Once the number and sizes of the network communities are identified, not only the community sizes are mesoscopic characteristics, but also the connectivity between communities and their internal organization.

To better understand the mesoscopic scale, we study the HCM introduced in Chapter 11 in the critical regime, when the random graph is on the verge of having a giant connected component, by studying a depth-first exploration process. This critical regime has been explored for wide classes of random graph models, including the CM. Indeed, most random graph models undergo a transition in connectivity, a so-called *phase transition*, as illustrated in Figure 14.1. The component sizes of random graphs at criticality were first investigated for Erdős-Rényi random graphs [7, 25], and more recently for inhomogeneous random graphs [23, 209] and for the CM [71, 72, 123, 156, 189]. All these models were found to follow qualitatively similar scaling limits, and hence can be considered to be members of the same universality class.

Taking the HCM as the null model for studying critical connectivity, we can investigate the influence of the community structure. A relevant question is under what conditions the HCM will show the same scaling limit as in the classical random graph models and hence is a member of the same universality class. An alternative formulation of the same question is to ask what the natural order of the mesoscopic scale should be, to influence or even alter the critical graph behavior. Our analysis shows that $\alpha = 2/3$ is a strong indicator for the extent to which mesoscopic scales change the global network picture. When communities are of size $N^{2/3}$ or smaller, the mesoscopic scales are small (yet not negligible), and the critical structures that arise are comparable to the structures encountered in the classical CM. When communities are potentially larger than $N^{2/3}$, the communities themselves start to alter the critical structures, and have the potential to entirely change the macroscopic picture.

We then proceed to study percolation on the HCM. We will exploit the fact that any supercritical HCM can be made critical by choosing a suitable percolation parameter. Therefore, we also study the scaling limits of the component sizes of a HCM under

critical bond percolation. We show that under percolation, the community structure potentially not only affects the component sizes, but also the width of the critical window.

Our main results for the critical components, both before and after percolation, crucially depend on the mesoscopic scale of the community structure. We obtain the precise conditions (Conditions 14.1 and 14.2) under which the mesoscopic scale does not become dominant. These conditions describe the maximum order of the community sizes that can be sustained in order not to distort the picture generated by the CM. In other words, when the community sizes remain relatively small, the results proven for the CM remain valid, despite the fact that the locally tree-like assumption is violated. And equally important, the same conditions indicate when the community sizes become large enough for the mesoscopic scale to take over. In that case, the CM is not an appropriate model.

14.1.1 Model conditions

We study the HCM model as defined in Section 11.1.1. We assume that the following conditions hold:

Condition 14.1 (Community regularity).

- (i) $P_n(H) = n_H^{(n)} / n \rightarrow P(H)$, where $(P(H))_H$ is a probability distribution on labeled graphs H of arbitrary size.
- (ii) $\lim_{n \rightarrow \infty} \mathbb{E}[S_n] = \mathbb{E}[S] < \infty$.
- (iii) $\mathbb{E}[D_n S_n] \rightarrow \mathbb{E}[DS] < \infty$.
- (iv) $s_{max} = \max_{i \in [n]} s_i \ll \frac{n^{2/3}}{\log(n)}$.

Condition 14.2 (Inter-community connectivity).

- (i) $\lim_{n \rightarrow \infty} \mathbb{E}[D_n^3] = \mathbb{E}[D^3] < \infty$.
- (ii) $\mathbb{P}(D = 0) < 1, \mathbb{P}(D = 1) \in (0, 1)$.
- (iii) $\nu_{D_n} := \frac{\mathbb{E}[D_n(D_n - 1)]}{\mathbb{E}[D_n]} = 1 + \lambda n^{-1/3} + o(n^{-1/3})$, for some $\lambda \in \mathbb{R}$.

Condition 14.1(i), 14.1(ii), and 14.2(ii) are the same conditions as the conditions on the HCM model introduced in Section 11.1.1. Condition 14.1(iii), 14.1(iv) and 14.2(iii) are the extra conditions that are necessary to investigate the behavior of the HCM in the critical window.

Remark 14.1. Because the inter-community half-edges are paired uniformly at random, it is possible that two half-edges attached to the same community are paired to each other, so that the inter-community half-edges form an edge inside some community H . However, since this corresponds to a self-loop in the macroscopic configuration model, Condition 14.2(i) ensures that such edges are rare as N becomes large [106, Section 7.3].

14.1.2 Results on critical component sizes

For a connected component of G , we can either count the number of communities in the component, or the number of vertices in it. We denote the number of communities in a component \mathcal{C} by $v^{(\text{H})}(\mathcal{C})$, and the number of communities with inter-community degree k by $v_k^{(\text{H})}(\mathcal{C})$. The number of vertices in component \mathcal{C} is denoted by $v(\mathcal{C})$. Define

$$v_D = \frac{\mathbb{E}[D(D-1)]}{\mathbb{E}[D]}, \quad (14.1.1)$$

where D is the asymptotic community degree in Condition 14.2. Let $p_k = \mathbb{P}(D = k)$.

Let $B_{\lambda,\eta}^{\mu}(t)$ denote Brownian motion with a parabolic drift [1, 99]: $B_{\lambda,\eta}^{\mu}(t) = \frac{\sqrt{\eta}}{\mu}B(t) + \lambda t - \frac{\eta t^2}{2\mu^3}$, where $B(t)$ is a standard Brownian motion. Let $W^\lambda(t)$ be the reflected process of $B_{\lambda,\eta}^{\mu}(t)$, i.e.,

$$W^\lambda(t) = B_{\lambda,\eta}^{\mu}(t) - \min_{0 \leq s \leq t} B_{\lambda,\eta}^{\mu}(s), \quad (14.1.2)$$

and let γ^λ denote the vector of ordered excursion lengths of W^λ . Choose $\mu = \mathbb{E}[D]$, $\eta = \mathbb{E}[D^3]\mathbb{E}[D] - \mathbb{E}[D^2]^2$. Let $\mathcal{C}_{(j)}$ denote the j th largest component of a HCM, and $\mathcal{C}_{(j)}^{(\text{CM})}$ the j th largest component of the underlying CM (i.e., $\mathcal{C}_{(j)}^{(\text{CM})}$ is the j th largest component measured in terms of the number of communities). Since the underlying CM satisfies Condition 14.1, by [71],

$$n^{-2/3} \left(v^{(\text{H})}(\mathcal{C}_{(j)}^{(\text{CM})}) \right)_{j \geq 1} \rightarrow \gamma^\lambda. \quad (14.1.3)$$

Thus, the number of communities in the components of a HCM follows the same scaling limit as the configuration model, since the communities are connected as in a configuration model.

The following theorem shows that the scaled component sizes of a HCM converge to a constant times γ^λ as well:

Theorem 14.1. *Fix $\lambda \in \mathbb{R}$. For a hierarchical configuration model satisfying Conditions 14.1 and 14.2,*

$$N^{-2/3} (v(\mathcal{C}_{(j)}))_{j \geq 1} \xrightarrow{d} \mathbb{E}[S]^{-2/3} \frac{\mathbb{E}[DS]}{\mathbb{E}[S]} \gamma^\lambda, \quad (14.1.4)$$

with respect to the product topology.

In the Erdős-Rényi random graph, the inhomogeneous random graph as well as in the CM, the scaled critical component sizes converge in the ℓ_{\downarrow}^2 topology [7, 23, 71], defined as

$$\ell_{\downarrow}^2 := \left\{ \mathbf{x} = (x_1, x_2, x_3, \dots) : x_1 \geq x_2 \geq x_3 \geq \dots \text{ and } \sum_{i=1}^{\infty} x_i^2 < \infty \right\}, \quad (14.1.5)$$

with the 2-norm as metric. Theorem 14.1 only proves convergence of the scaled component sizes of a HCM in the product topology. Convergence in the product

topology ensures that for any fixed k , the k largest components converge to the limit of Theorem 14.1. Convergence in the ℓ_{\downarrow}^2 topology ensures that the entire infinite-dimensional vector of component sizes converges. Furthermore, for convergence in the ℓ_{\downarrow}^2 -topology, the sum of squared rescaled component sizes is required to be finite. Thus, convergence in the ℓ_{\downarrow}^2 -topology is a stronger notion than convergence in the product topology. In the CM, the conditions for convergence in the product topology and convergence in the ℓ_{\downarrow}^2 -topology are the same. In the HCM however, the conditions for convergence in the product topology turn out to be different than those in the ℓ_{\downarrow}^2 -topology:

Theorem 14.2. *Suppose G is a hierarchical configuration model satisfying Conditions 14.1 and 14.2. Then the convergence of Theorem 14.1 also holds with respect to the ℓ_{\downarrow}^2 -topology if and only if G satisfies $\mathbb{E}[S_n^2] = o(n^{1/3})$.*

Remark 14.2. This theorem shows that there exist graphs where the critical component sizes converge in the product topology, but not in the ℓ_{\downarrow}^2 -topology. As mentioned before, this does not happen in other random graph models such as the Erdős-Rényi random graph, the inhomogeneous random graph and the CM [7, 23, 71]. Furthermore, we can find the exact condition under which the component sizes only converge in the product topology. This theorem also shows that the conditions for convergence in the product topology and the ℓ_{\downarrow}^2 -topology are indeed equivalent in the CM: the CM is a special case of the HCM with size one communities. Therefore, $\mathbb{E}[S_n] = 1$ for the CM, and the component sizes always converge in the ℓ_{\downarrow}^2 -topology if they converge in the product topology. It is surprising that the condition on the ℓ_{\downarrow}^2 convergence only depends on the community sizes, but not on the joint distribution of the community sizes and the inter-community degrees.

Remark 14.3. The results of this chapter can also be applied to a CM or a HCM with vertex attributes. In this setting, every vertex of a CM has a positive, real-valued vertex attribute S , where S satisfies Conditions 14.1 and 14.2. These vertex attributes may for example denote the capacity or the weight of a vertex. If we are interested in the sum of the vertex attributes in each connected component, then these sums scale as in Theorem 14.1. In an even more general setting, the random graph has a community structure, and every vertex within the communities again has a vertex attribute. Then we are back in the HCM setting, only now the community size S does not denote the number of vertices in a community, but the sum over the vertex attributes in a certain community. Therefore, also in a random graph with community structure and vertex attributes, the critical sum over vertex attributes satisfies Theorem 14.1 under appropriate conditions as in Condition 14.1.

14.1.3 Results on critical percolation on the HCM

We now consider bond percolation on the hierarchical configuration model, where every edge is removed independently with probability $1 - \pi$. In the CM, the random graph that remains after percolation can be described in terms of a CM with a different degree sequence [43, 81, 115]. Furthermore, if the CM is supercritical, it is possible to choose π such that the resulting graph is distributed as a critical CM. Similarly, Section 11.3 showed that after percolating a HCM, we again obtain a HCM, but with a

different community distribution since the communities are percolated. By adjusting the parameter π we can make sure that the HCM is critical after percolation. Thus, given any supercritical HCM, it is possible to create a critical HCM, by setting π correctly.

In the hierarchical configuration model, it is convenient to percolate first only the edges inside communities. This percolation results in a HCM with percolated communities. These percolated communities may be disconnected. However, if we define the connected components of the percolated communities as new communities, we have a new HCM. Let $S_n^{(\pi)}$ and $D_n^{(\pi)}$ denote the size and degree of uniformly chosen communities after percolation only inside the communities with probability π , and $S^{(\pi)}$ and $D^{(\pi)}$ their infinite-size limits. After this, we percolate only the inter-community edges. This percolation is similar to percolation on the CM, since the inter-community edges are paired as in the CM.

We assume the following:

Condition 14.3 (Critical percolation window).

1. S_n and D_n satisfy Conditions 14.1 and 14.2(i) and (ii), and

$$\nu_{D_n^{(\pi_n)}}^{(n)} := \frac{\mathbb{E}[D_n^{(\pi_n)}(D_n^{(\pi_n)} - 1)]}{\mathbb{E}[D_n^{(\pi_n)}]} \rightarrow \frac{\mathbb{E}[D^{(\pi)}(D^{(\pi)} - 1)]}{\mathbb{E}[D^{(\pi)}]} > 1. \quad (14.1.6)$$

2. For some $\lambda \in \mathbb{R}$,

$$\pi_n = \pi_n(\lambda) := \frac{1}{\nu_{D^{(\pi_n(\lambda))}}^{(n)}} \left(1 + \frac{\lambda}{n^{1/3}} \right). \quad (14.1.7)$$

Here π is the solution to $\pi = 1/\nu_{D^{(\pi)}}$.

Remark 14.4. It can be shown that $\nu_{D^{(\pi)}}^{(n)}$ is increasing in π . Thus, (14.1.7) has a *unique* solution for every $\lambda \in \mathbb{R}$ when n is large enough.

Equation (14.1.6) makes sure that after percolation of the intra-community edges, the new HCM is supercritical, otherwise there is no hope of making the graph critical by removing more edges. After percolating inside communities with parameter π , $\nu_{D_n^{(\pi)}}^{(n)}$ is the value of ν of the new macroscopic CM.

Let $\check{D}^{(\pi)}$ denote the exploded version of $D^{(\pi)}$, that is, every half-edge of a percolated community is kept with probability $\sqrt{\pi}$, and with probability $1 - \sqrt{\pi}$, it explodes, in that it creates a new community of the same shape with only one half-edge attached to it. By [71, Thm. 3], the component sizes of a percolated CM have similar scaling limits as the original configuration model, but with D replaced by its exploded version. For the HCM, a similar statement holds. Let $\tilde{\gamma}^\lambda$ denote the ordered excursions of the reflected Brownian motion $B_{\eta, \mu}^\lambda$ with $\mu = \mathbb{E}[\check{D}^{(\pi)}]$ and $\eta = \mathbb{E}[(\check{D}^{(\pi)})^3] \mathbb{E}[\check{D}^{(\pi)}] - \mathbb{E}[(\check{D}^{(\pi)})^2]^2$. We denote the percolated components of a hierarchical configuration model by \mathcal{C}' . Then the following theorem shows that the percolated component sizes follow a similar scaling limit as the critical component sizes of Theorem 14.1:

Theorem 14.3. *Under Condition 14.3, in the percolated hierarchical configuration model,*

$$N^{-2/3}(v(\mathcal{C}'_{(j)}))_{j \geq 1} \rightarrow \mathbb{E}[\tilde{S}^{(\pi)}]^{-2/3} \frac{\mathbb{E}[\tilde{D}^{(\pi)} \tilde{S}^{(\pi)}]}{\mathbb{E}[\tilde{D}^{(\pi)}]} \sqrt{\pi} \tilde{\gamma} \lambda, \tag{14.1.8}$$

in the product topology.

Remark 14.5. By [71], the critical window of a CM satisfies

$$\pi_n(\lambda) = \frac{1}{\nu^{(n)}}(1 + \lambda/n^{1/3}) = \pi_n(0)(1 + \lambda/n^{1/3}). \tag{14.1.9}$$

Therefore, the critical window (14.1.7) in the HCM is similar to the critical window in the CM, with the difference that in the HCM, we first perform an extra step of percolation inside the communities. For this reason, the critical window of the HCM (14.1.7) is in an implicit form, as it depends on both the percolation inside communities, captured in $\nu^{(n)}_{D^{(\pi_n)}}$, and it depends on the inter-community percolation. In Section 14.3.2, we show that the critical window in the HCM can be written as

$$\pi_n(\lambda) = \pi_n(0) \left(1 + \frac{c^* \lambda}{n^{1/3}} \right), \tag{14.1.10}$$

for some constant $c^* \leq 1$ when $\mathbb{E}[D_n^2 S_n] \rightarrow \mathbb{E}[D^2 S] < \infty$. In this case, the critical window for the HCM is very similar to the critical window in the CM. The constant c^* captures how much smaller the critical window becomes when adding a community structure. The more vulnerable the communities are to percolation, the smaller the constant c^* .

Remark 14.6. In Theorem 14.3, we assume that the intra-community edges are equally vulnerable for percolation as the inter-community edges. It is also possible to study percolation where the edges inside communities are percolated with some parameter π_{in} , and the inter-community edges are percolated with probability π_{out} . Several different combinations of the parameters π_{in} and π_{out} then correspond to critical percolation. When percolating inside the communities with parameter π_{in} results in a supercritical graph, it is always possible to find $\pi_{out} = 1/\nu_{D^{(\pi_{in})}}(1 + \lambda/n^{1/3})$ such that after percolation the resulting graph is in the critical window with parameter λ . The critical value of the HCM is then defined as the value such that $\pi_{in}(\lambda) = \pi_{out}(\lambda)$. Figure 14.2 illustrates this for several values of λ for star-shaped communities. The higher the intra-community parameter π_{in} , the lower the inter-community percolation parameter π_{out} needs to be to make the resulting graph critical.

Remark 14.7. Theorem 14.3 shows the convergence of the percolated component sizes in the product topology. As in Theorem 14.2, by assuming that $\mathbb{E}[S_n^2] = o(n^{1/3})$, we can also show convergence in the ℓ_{\downarrow}^2 -topology. However, in the case of percolation, $\mathbb{E}[S_n^2] = o(n^{1/3})$ is not a necessary condition for convergence in the ℓ_{\downarrow}^2 -topology. After percolating first only edges inside communities, we apply Theorem 14.1 or 14.2 to the percolated clusters. In the example of line communities, communities that have the shape of a line, we can make the lines large enough such that $\mathbb{E}[S_n^2] > \epsilon n^{1/3}$. However, if we then percolate inside these line communities, $\mathbb{E}[(S_n^{(\pi)})^2] = o(n^{1/3})$.

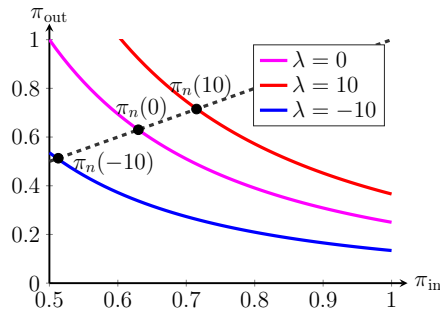


Figure 14.2: The value of π_{out} corresponding to a given π_{in} for star-shaped communities with five end points (as in Figure 14.4), with $n = 10^5$. The intersection with the line $y = x$ gives the critical value $\pi_n(\lambda)$.

Thus, after percolating inside the communities we can use Theorem 14.2 to show that the percolated component sizes converge in the ℓ_1^2 -topology, even though $\mathbb{E}[S_n^2] \geq \varepsilon n^{1/3}$.

14.1.4 Discussion

What makes communities "small"? Communities form a mesoscopic structure of a graph. If the communities become very large, then the critical component sizes will be determined by the sizes of the largest communities. In this chapter, we find the conditions under which the influence of the mesoscopic structure on the critical component sizes is small when the inter-community degrees have a finite third moment. In this situation, the mesoscopic structures of the configuration model are small enough for the model to be in the same universality class as the configuration model. Theorems 14.1 and 14.2 show that there are different scales on which communities can be "small". Theorem 14.1 shows that the communities are small on the mesoscopic scale when Condition 14.1 holds. In particular, the maximum size of a "small" community is $n^{2/3}/\log(n)$. This is much smaller than the total number of vertices in the graph, but it still tends to infinity when $n \rightarrow \infty$, which shows the mesoscopic nature of the communities. Under these conditions, the order of the components is determined by the order of the components in the underlying CM. Since the convergence of Theorem 14.1 holds in the product topology, this only holds for the first k components, for k fixed. Theorem 14.2 shows that an additional condition is necessary for the communities to be small on a macroscopic scale. If this condition does not hold, then the first k components are still determined by the macroscopic CM for any fixed k , but some large components that are small in the macroscopic sense will be discovered eventually.

Similarity to the configuration model. The scaling limit of Theorem 14.1 is similar to the scaling limit of the configuration model; it only differs by a constant. In fact, the CM is a special case of the HCM: the case where all communities have size one. One could argue that therefore the HCM is in the same universality class

as the CM. However, there are still some differences. The variable D in the HCM is the inter-community degree. Then, if D has finite third moment, the scaling limit of a HCM is similar to the scaling limit of a CM with finite third moment of the degrees. However, Chapter 12 showed that it is possible to construct a HCM with finite third moment of D , but infinite third moment of the degree distribution. One example of this is a hierarchical configuration model where all communities are households [14]: complete graphs, where all vertices of the complete graphs have inter-community degree one. In this household model, every community of inter-community degree k , contains also k vertices of degree k . Therefore, the inter-community degree distribution may have finite third moment, while the degree distribution has an infinite third moment. In the CM, the scaling limit under an infinite third moment is very different from the one with finite third moment [72]. However, using this household model, it is possible to construct a random graph with an infinite third moment of the degree distribution, but a similar scaling limit as the CM under the finite third-moment assumption. Similarly, it is possible to create a community structure such that the inter-community degrees have an infinite third moment, but the degree distribution has a finite third moment. Therefore, adding a community structure to a graph while keeping the degree distribution fixed may change the scaling limits significantly.

Surplus edges. The number of surplus edges of a connected graph G is defined as $SP(G) := (\# \text{ edges of } G) - |G| + 1$ and indicates how far G deviates from being a tree. In the CM, the rescaled component sizes and the number surplus edges in the components converge jointly. A surplus edge in the macroscopic CM stays a surplus edge of the HCM, since all communities are connected. In the intuitive picture of densely connected communities, the communities have many surplus edges. In the HCM, we give each vertex in the macroscopic CM a weight: the size of the corresponding community. Then, in Theorem 14.1 and 14.2, we are interested in the weighted size of the components. Counting surplus edges is very similar, now we also give each vertex in the macroscopic CM a weight: the number of surplus edges in the corresponding community. We are again interested in the weighted component sizes, which counts the total number of surplus edges inside communities. The surplus edges between different communities are the surplus edges of the macroscopic CM. The number of such edges rescaled by $N^{2/3}$ goes to zero by [71]. Therefore, if the surplus edges satisfy the same conditions as the community sizes in Condition 14.2(ii), they scale similarly to the component sizes. Thus, if SP_n denotes the number of surplus edges inside a uniformly chosen community, and $\mathbb{E}[SP_n] \rightarrow \mathbb{E}[SP] < \infty$, $\mathbb{E}[SP_n \cdot D_n] \rightarrow \mathbb{E}[SP \cdot D] < \infty$ and $SP_{\max} \ll \frac{n^{2/3}}{\log(n)}$, then

$$N^{-2/3}(v(\mathcal{C}_{(j)}), SP(\mathcal{C}_{(j)}))_{j \geq 1} \xrightarrow{d} \mathbb{E}[S]^{-2/3} \gamma^\lambda \left(\frac{\mathbb{E}[DS]}{\mathbb{E}[S]}, \frac{\mathbb{E}[SP \cdot D]}{\mathbb{E}[S]} \right), \quad (14.1.11)$$

in the product topology, where $SP(\mathcal{C})$ denotes the number of surplus edges in component \mathcal{C} .

Unsurprisingly, this scaling of the surplus edges is very different from the scaling of the surplus edges in the CM, which is locally tree-like, even though the scaling of the component sizes is very similar to the one in the CM.

Let $\mathbf{N}^\lambda = (N^\lambda(s))_{s \geq 0}$ denote a counting process of marks with intensity $W^\lambda(s)/\mathbb{E}[D]$ conditional on $(W^\lambda(u))_{u \leq s}$, with $W^\lambda(s)$ as in (14.1.2). Furthermore, let $N(\gamma)$ denote the number of marks in the interval γ . Let $\text{SP}^{(\text{H})}(\mathcal{C})$ denote the number of surplus edges of component \mathcal{C} that are inter-community edges. These surplus edges are the surplus edges of the macroscopic CM. By [71], these surplus edges converge jointly with the component sizes to $N(\gamma^\lambda)$. Therefore, in the HCM,

$$(N^{-2/3}v(\mathcal{C}_{(j)}), \text{SP}^{(\text{H})}(\mathcal{C}_{(j)}))_{j \geq 1} \xrightarrow{d} \left(\mathbb{E}[S]^{-2/3} \gamma^\lambda \frac{\mathbb{E}[DS]}{\mathbb{E}[S]}, N(\gamma^\lambda) \right). \quad (14.1.12)$$

Infinite third moment. We investigate the scaled component sizes of a random graph with communities, where the inter-community degrees have a finite third moment. A natural question therefore is what happens if we drop the finite-third moment assumption. The scaled component sizes of random graphs with infinite third moment, but finite second moment, have been investigated for several models. If the degrees follow a power-law with exponent $\tau \in (3, 4)$, then the component sizes of an inhomogeneous random graph as well as a CM scale as $n^{(\tau-2)/(\tau-1)}$ [22, 72]. In the HCM this may also be the correct scaling, but clearly then we need to replace Condition 14.1(iv) by $s_{\max} = o(n^{(\tau-2)/(\tau-1)})$, since otherwise the largest community will dominate the component sizes. This indicates that the heavier the power-law tail, the smaller the maximal community size can be for the HCM to be in the same universality class as the CM. What exact assumptions on the community-size distribution are needed to obtain the same scaling limit as in the CM remains open for further research.

Optimality of conditions. Condition 14.2 is necessary for the macroscopic CM to have components of size $O(n^{2/3})$. Clearly it is also necessary that the maximum community size is $o(n^{2/3})$, since otherwise the largest community could dominate the component sizes. For example, it would be possible to create communities of size larger than $n^{2/3}$ with inter-community degree zero. Then, these components are the smallest components of the macroscopic CM, but may be the largest components in the HCM. Condition 14.1(iv) has an extra factor $1/\log(n)$, which we need to prove that the component sizes are not dominated by the community sizes. Probably this condition is not optimal, we believe the optimal condition to be $s_{\max} = o(n^{2/3})$. Furthermore, Conditions 14.1(ii) and (iii) are necessary for taking the limit in (14.1.4).

Outline. The remainder of this chapter is organized as follows. In Section 14.2 we prove Theorem 14.1 by studying a depth-first exploration process. This proof relies heavily on the fact that the macroscopic CM follows a similar scaling limit [71]. Then we prove Theorem 14.2 in Section 14.2.3. In Section 14.3, we study percolation on the HCM. First we prove Theorem 14.3, and after that we show in Section 14.3.2 that the critical window of a HCM is similar to the critical window of the CM. We conclude Section 14.3 with some examples of communities.

14.2 Proof of the scaling of the critical HCM

In this section, we prove Theorems 14.1 and 14.2. We start by describing an exploration process that finds the component sizes in Section 14.2.1. We use this exploration process to show that the components that are found before time $Tn^{2/3}$ for some large T converge to the right scaling limit. After that, we prove in Section 14.2.2 that the probability that a large component is found after that time is small, which completes the proof of Theorem 14.1. Then we prove Theorem 14.2 in Section 14.2.3.

14.2.1 Exploration process

To find the component sizes, we use the same depth-first exploration process for the CM as in [71, Algorithm 1]. However, instead of exploring vertices, we now explore communities. This means that we only explore the macroscopic CM. In each step of the exploration process, we discover an entire community, and we explore further using only the inter-community connections. Therefore, the only difference between our exploration process and the standard exploration process for the CM, is that we count the number of vertices in each community that we discover. At each step t , an inter-community half-edge can be in the ordered set of active half-edges \mathcal{A}_t , in the set of sleeping half-edges \mathcal{S}_t , or none of these two sets. Furthermore, every vertex of the HCM is alive or dead. When a vertex is dead, it is in the set of dead vertices \mathcal{D}_t . The set \mathcal{C}_t that appears in the algorithm keeps track of the half-edges that connect to another half-edge of the same community. The set \mathcal{B}_t keeps track of the number of half-edges that connect to a community that has already been discovered in a previous set of the algorithm.

Algorithm 14.1. For $t = 0$, all inter-community half-edges are in \mathcal{S}_0 , and both \mathcal{D}_0 and \mathcal{A}_0 are empty. While $\mathcal{A}_t \neq \emptyset$ or $\mathcal{S}_k \neq \emptyset$ we set $k = k + 1$ and perform the following steps:

1. If $\mathcal{A}_t \neq \emptyset$, then take the smallest inter-community half-edge a from \mathcal{A}_t .
2. Take the half-edge b that is paired to a . By construction of the algorithm, the community H to which b is attached, is not discovered yet. Declare H to be discovered. Let b_{H_1}, \dots, b_{H_r} be the other half-edges attached to community H , and let \mathcal{V}_H denote the set of vertices of community H . Let b_{H_1}, \dots, b_{H_r} be smaller than all other elements of \mathcal{A}_t , and order them as $b_{H_1} > b_{H_2} > \dots > b_{H_r}$. Let $\mathcal{C}_k \subset \{b_{H_1}, \dots, b_{H_r}\}$ denote all half-edges attached to community H that attach to another half-edge adjacent to H . Furthermore, let $\mathcal{B}_t \subset \mathcal{A}_t \cup \{b_{H_1}, \dots, b_{H_r}\}$ denote the collection of half-edges in \mathcal{A}_t that have been paired to one of the b_{H_i} 's, together with the b_{H_i} s they have been paired to. Then, set $\mathcal{A}_{k+1} = \mathcal{A}_k \cup \{b_{H_1}, \dots, b_{H_r}\} \setminus (\mathcal{B}_t \cup \mathcal{C}_t)$, $\mathcal{S}_{t+1} = \mathcal{S}_k \setminus \{b, b_{H_1}, \dots, b_{H_r}\}$ and $\mathcal{D}_{t+1} = \mathcal{D}_k \cup \mathcal{V}_H$.
3. If $\mathcal{A}_t = \emptyset$, then we pick a half-edge a from \mathcal{S}_t uniformly at random. Let H be the community to which a is attached, and a_{H_1}, \dots, a_{H_r} the other half-edges attached to community H . Again, order them as $a_{H_1} > a_{H_2} > \dots > a_{H_r} > a$. Let \mathcal{V}_H denote the vertices of community H . Declare H to be discovered. Let \mathcal{C}_t again denote the collections of half-edges of H that attach to another half-edge incident to H . Then set $\mathcal{A}_{k+1} = \{a, a_{H_1}, \dots, a_{H_r}\} \setminus \mathcal{C}_t$, $\mathcal{S}_{t+1} = \mathcal{S}_t \setminus \{a, a_{H_1}, \dots, a_{H_r}\}$ and $\mathcal{D}_{t+1} = \mathcal{D}_t \cup \mathcal{V}_H$.

Algorithm 14.1 discovers one community at each step. When the inter-community edges create a cycle, double edge or self-loop on the community level, the corresponding half-edges are in \mathcal{B}_t or \mathcal{C}_t , and they are thrown away. Therefore, at each step, an unexplored community is discovered. Since the communities are found by selecting a half-edge uniformly at random, the communities are explored in a size-biased manner with respect to the number of edges going out of the community. The dead vertices correspond to all vertices inside communities that have already been discovered. We define the additive functional

$$Z_n(t) = |\mathcal{D}_t| \quad (14.2.1)$$

as the number of vertices that have been discovered up to time k . The exploration process finds an undiscovered community of the HCM at each step, and therefore

$$Z_n(t) = \sum_{i=1}^k s_{(i)}, \quad (14.2.2)$$

where $s_{(i)}$ denotes the size of the i th discovered community.

Let $d_{(i)}$ be the inter-community degree of the i th explored community. Define $Q_n(k)$ as

$$\begin{aligned} Q_n(0) &= 0, \\ Q_n(t) &= \sum_{i=1}^t (d_{(i)} - 2 - 2c_{(i)}), \end{aligned} \quad (14.2.3)$$

where $c_{(i)}$ denotes the number of cycles/self-loops or double edges that are found when discovering the i th community. Let \mathcal{C}_j denote the j th component that is found by the exploration process, and define

$$\tau_j = \inf\{t : Q_n(t) = -2j\}. \quad (14.2.4)$$

Then, $\tau_j - \tau_{j-1}$ is the number of communities in component \mathcal{C}_j [71], so that

$$v^{(H)}(\mathcal{C}_j) = \tau_j - \tau_{j-1}. \quad (14.2.5)$$

Furthermore, the size of \mathcal{C}_j equals

$$v(\mathcal{C}_j) = Z_n(\tau_j) - Z_n(\tau_{j-1}). \quad (14.2.6)$$

By [71], the rescaled process $Q_n(t)$ converges to the reflected version of a Brownian motion with negative parabolic drift. To derive the sizes of the components in the hierarchical configuration model, we now study the convergence of the process $Z_n(t)$:

Lemma 14.1. *For any $u \geq 0$,*

$$\sup_{t \leq u} \left| n^{-2/3} Z_n(\lfloor tn^{2/3} \rfloor) - \mathbb{E}[DS] / \mathbb{E}[D] t \right| \xrightarrow{\mathbb{P}} 0. \quad (14.2.7)$$

Proof. We use [71, Proposition 29], choosing $\alpha = 2/3$ and $f_n(i) = s_i$. This yields

$$\sup_{t \leq u} \left| n^{-2/3} \sum_{i=1}^{\lfloor tn^{2/3} \rfloor} s_{(i)} - \frac{t \mathbb{E}[DS]}{\mathbb{E}[D]} \right| = O_{\mathbb{P}}(n^{-1/3} \sqrt{us_{\max}} \vee n^{-1/3} u^2 d_{\max}). \quad (14.2.8)$$

Using that $d_{\max} = o(n^{1/3})$ and $s_{\max} = o(n^{2/3})$ by Condition 14.1(iv) gives the result. \square

Lemma 14.2. *For any $u \geq 0$,*

$$\left(n^{-1/3}Q_n(tn^{2/3}), n^{-2/3}Z_n(tn^{2/3}) \right)_{t \leq u} \xrightarrow{d} \left(W^\lambda(t), \mathbb{E}[DS]/\mathbb{E}[S]t \right)_{t \leq u} \quad (14.2.9)$$

in the $J_1 \times J_1$ topology, which is the product topology of the Skorokhod J_1 topology.

Proof. Since $t \mapsto t\mathbb{E}[DS]/\mathbb{E}[S]$ is deterministic, $\left(n^{-1/3}Q_n(tn^{2/3}) \right) \xrightarrow{d} (B^\lambda(t))$ by [71, Thm. 8] and $\left(n^{-2/3}Z_n(tn^{2/3}) \right) \xrightarrow{\mathbb{P}} (t\mathbb{E}[DS]/\mathbb{E}[S])$ in the J_1 topology, an analogy of Slutsky's theorem for processes proves the lemma. \square

By [71], the excursion lengths of \bar{Q}_n converge to the excursions of $W^\lambda(t)$ defined in (14.1.2), where

$$\bar{Q}_n(t) = n^{-1/3}Q_n(tn^{2/3}). \quad (14.2.10)$$

Since the excursions of Q_n encode the number of communities in the components, and Z_n encodes the sum of the corresponding community sizes, Lemma 14.2 shows that the components that have been discovered before time $Tn^{2/3}$ satisfy (14.1.4).

14.2.2 Sizes of components that are discovered late and convergence in product topology

By Lemma 14.2, the component sizes that have been discovered up to time $Tn^{2/3}$ converge to a constant times the excursion lengths of a reflected Brownian motion with parabolic drift. To prove that the ordered components of the HCM converge to the ordered excursion lengths of this process, we need to show that the probability of encountering a large component after time $Tn^{2/3}$, is small, so that all large components are discovered early in the process. From [72, Lemma 14], we know that for every $\eta > 0$

$$\lim_{T \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P} \left(v^{(H)}(\mathcal{C}_{\max}^{\geq T}) > \eta n^{2/3} \right) = 0, \quad (14.2.11)$$

where $\mathcal{C}_{\max}^{\geq T}$ is the largest component found after time $Tn^{2/3}$. Therefore, we only need to show that the probability that there exists a component with less than $\eta n^{2/3}$ communities such that its size is larger than $\delta n^{2/3}$ is small when $\eta \ll \delta$. We prove that the probability that a uniformly chosen component satisfies this property is exponentially small in n . Therefore, the probability that such a component exists is also small.

We first explore the HCM according to Algorithm 14.1 until the first time after $Tn^{2/3}$ that a component has been explored. Then, we remove all components that have been found so far. We denote the resulting graph by $G^{\geq T}$. For any realization of the graph $G^{\geq T}$, the probability $p_{k,s}^{\geq T}(n)$ that a uniformly chosen community in $G^{\geq T}$ has degree k and size s can be bounded as

$$p_{k,s}^{\geq T}(n) \leq \frac{n_{k,s}}{n - Tn^{2/3}} = p_{k,s}^{(n)}(1 + o(n^{-1/3})), \quad (14.2.12)$$

where $n_{k,s}$ denotes the number of communities of size s and inter-community degree k . Therefore, for any realization of $G^{\geq T}$, the expected size of a community in $G^{\geq T}$, $\mathbb{E}[S^{\geq T}] < \infty$, and similarly $\mathbb{E}[D^{\geq T}] < \infty$ and $\mathbb{E}[D^{\geq T}S^{\geq T}] < \infty$. Now, we start exploring $G^{\geq T}$ as in Algorithm 14.1. We want to show that with high probability $G^{\geq T}$ does not contain components larger than $\delta n^{2/3}$. By [72], the CM with high probability does not contain any components of size $\eta n^{2/3}$ that are discovered after time $Tn^{2/3}$. Therefore, with high probability $G^{\geq T}$ does not contain components with more than $\eta n^{2/3}$ communities.

Lemma 14.3 shows that the probability that the first explored $\eta n^{2/3}$ communities using Algorithm 14.1 contain more than $\delta n^{2/3}$ vertices is small. This lemma holds for all HCMs satisfying Conditions 14.1 and 14.2. In the proof of Theorem 14.1 we will apply this lemma to show that the HCM $G^{\geq T}$ with high probability only has components with at most $\eta n^{2/3}$ communities.

Lemma 14.3. *For any $\eta, \delta > 0$ satisfying $\delta > 2\eta\mathbb{E}[DS] / \mathbb{E}[D]$,*

$$\mathbb{P}\left(\sum_{i=1}^{\lfloor \eta n^{2/3} \rfloor} s_{(i)} > \delta n^{2/3}\right) \leq e^{-\zeta n^{2/3}/s_{\max}}, \quad (14.2.13)$$

for some $\zeta > 0$.

Proof. Let T_i be independent exponential random variables with rate d_{H_i}/ℓ_n . Furthermore, let $M(t) = \#\{j : T_j \leq t\}$. Then,

$$\mathbb{E}[M(t)] = \sum_{i \in [n]} \left(1 - e^{-td_{H_i}/\ell_n}\right) \leq t, \quad (14.2.14)$$

using that $1 - e^{-x} \leq x$. Similarly, using that $1 - e^{-x} \geq x - x^2/2$,

$$\mathbb{E}[M(t)] \geq \sum_{i \in [n]} \left(\frac{td_{H_i}}{\ell_n} - \frac{t^2 d_{H_i}^2}{\ell_n^2}\right) = t - \frac{t^2 \mathbb{E}[D_n^2]}{n \mathbb{E}[D_n]}. \quad (14.2.15)$$

Furthermore, for n large enough, if $t = o(n)$,

$$\begin{aligned} \mathbb{P}(M(2t) < t) &\leq \mathbb{P}\left(M(2t) < \frac{3}{2}t \left(1 - \frac{t \mathbb{E}[D_n]}{n \mathbb{E}[D_n^2]}\right)\right) \\ &\leq \mathbb{P}\left(M(2t) < \frac{3}{4} \mathbb{E}[M(2t)]\right) = \mathbb{P}(e^{-uM(2t)} > e^{-\frac{3}{4}u \mathbb{E}[M(2t)]}) \\ &\leq e^{\frac{3}{4}u \mathbb{E}[M(2t)]} \mathbb{E}\left[e^{-uM(2t)}\right], \end{aligned} \quad (14.2.16)$$

for any $u > 0$, where the last inequality uses the Markov inequality. Let $q_i = 1 - e^{-2td_{H_i}/\ell_n}$. Since $M(2t)$ is a sum of independent indicator variables,

$$\mathbb{E}\left[e^{-uM(2t)}\right] = \prod_{i \in [n]} (1 + q_i(e^{-u} - 1)) \leq \prod_{i \in [n]} e^{q_i(e^{-u} - 1)} = e^{\mathbb{E}[M(2t)](e^{-u} - 1)}, \quad (14.2.17)$$

where the inequality uses that $1 + x \leq e^x$, with $x = q_i(e^{-u} - 1)$. Plugging this into (14.2.16) and setting $u = -\log(\frac{3}{4})$ yields

$$\mathbb{P}(M(2t) < t) \leq e^{\mathbb{E}[M(2t)](\frac{3}{4}u + e^{-u} - 1)} = e^{\mathbb{E}[M(2t)](-\frac{3}{4}\log(\frac{3}{4}) - \frac{1}{4})}. \tag{14.2.18}$$

Then, using that $-(1 - x) \log(1 - x) \leq x - \frac{x^2}{2}$,

$$\mathbb{P}(M(2t) < t) \leq e^{-\frac{\mathbb{E}[M(2t)]}{32}} \leq e^{-Ct(1 - \frac{t}{n})}, \tag{14.2.19}$$

for some $C > 0$, where we have used (14.2.15).

Consider $Y_t = \sum_{i=1}^{M(t)} s_{(i)} - t \frac{\mathbb{E}[D_n S_n]}{\mathbb{E}[D_n]}$. Let \mathcal{F}_t denote the sigma-field generated by the information revealed up to time t . Then,

$$\begin{aligned} \mathbb{E}[Y_t | \mathcal{F}_{t-1}] &= Y_{t-1} + \sum_{i \in [n]} s_i \mathbb{P}(T_i \in [t-1, t] | T_i > t-1) - \frac{\mathbb{E}[D_n S_n]}{\mathbb{E}[D_n]} \\ &= Y_{t-1} + \sum_{i \in [n]} s_i (1 - e^{-d_{H_i}/\ell_n}) - \frac{\mathbb{E}[D_n S_n]}{\mathbb{E}[D_n]} \leq Y_{t-1}. \end{aligned} \tag{14.2.20}$$

Therefore, Y_t is a supermartingale. Furthermore,

$$\begin{aligned} \text{Var}(Y_t | \mathcal{F}_{t-1}) &= \text{Var}\left(\sum_{i \in [n]} s_i \mathbb{1}_{\{T_i \in [t-1, t]\}}\right) = \sum_{i \in [n]} s_i^2 (1 - e^{-d_{H_i}/\ell_n}) e^{-d_{H_i}/\ell_n} \\ &\leq \sum_{i \in [n]} s_i^2 d_{H_i} / \ell_n \leq s_{\max} \frac{\mathbb{E}[D_n S_n]}{\mathbb{E}[D_n]}. \end{aligned} \tag{14.2.21}$$

Thus, we can apply [63, Thm. 7.3], which states that for a supermartingale X with $\text{Var}(X_t | \mathcal{F}_{t-1}) \leq \sigma_t^2$ and $X_t - \mathbb{E}[X_t | \mathcal{F}_{t-1}] \leq K$ for all $t \in [n]$,

$$\mathbb{P}(X_n \geq X_0 + u) \leq \exp\left(-\frac{u^2}{\sum_{i=1}^n \sigma_i^2 + Ku/3}\right). \tag{14.2.22}$$

Applying this to $Y_{\lfloor 2\eta n^{2/3} \rfloor}$ with $Y_0 = 0$, $\sigma_i^2 = s_{\max}$ and $K = s_{\max}$, we obtain

$$\mathbb{P}(Y_{\lfloor 2\eta n^{2/3} \rfloor} > u) \leq \exp\left(-\frac{u^2}{2\eta n^{2/3} s_{\max} + s_{\max} u/3}\right). \tag{14.2.23}$$

Because by assumption $\zeta = \delta - 2\eta \mathbb{E}[DS] / \mathbb{E}[D] > 0$,

$$\begin{aligned} \mathbb{P}\left(\sum_{i=1}^{M(\lfloor 2\eta n^{2/3} \rfloor)} s_{(i)} > \delta n^{2/3}\right) &= \mathbb{P}\left(Y_{\lfloor 2\eta n^{2/3} \rfloor} > \delta n^{2/3} - 2\eta n^{2/3} \frac{\mathbb{E}[D_n S_n]}{\mathbb{E}[D_n]}\right) \\ &= \mathbb{P}\left(Y_{\lfloor 2\eta n^{2/3} \rfloor} > \zeta n^{2/3}\right) \\ &\leq \exp\left(-\frac{\zeta^2 n^{4/3}}{2\eta n^{2/3} s_{\max} + s_{\max} \zeta n^{2/3} / 3}\right) \leq e^{-\zeta n^{2/3} / s_{\max}}, \end{aligned} \tag{14.2.24}$$

for some $\zeta > 0$. Thus,

$$\begin{aligned} \mathbb{P}\left(\sum_{i=1}^{\lfloor \eta n^{2/3} \rfloor} s_{(i)} > \delta n^{2/3}\right) &\leq \mathbb{P}\left(\sum_{i=1}^{M(\lfloor 2\eta n^{2/3} \rfloor)} s_{(i)} > \delta n^{2/3}\right) + \mathbb{P}(M(\lfloor 2\eta n^{2/3} \rfloor) < \eta n^{2/3}) \\ &\leq e^{-\zeta n^{2/3}/s_{\max}} + e^{-Cn^{2/3}(1-n^{-1/3})}, \end{aligned} \quad (14.2.25)$$

which proves the lemma. \square

By applying the previous lemma to the hierarchical configuration model $G^{\geq T}$, we can now show that the probability that a component of size $\delta n^{2/3}$ is found after time $Tn^{2/3}$ is small for T large enough:

Lemma 14.4. *Let $\mathcal{C}_{\max}^{\geq T}$ denote the largest component of a hierarchical configuration model satisfying Conditions 14.1 and 14.2, of which the first vertex is explored after time $Tn^{2/3}$. Then, for all $\delta > 0$,*

$$\lim_{T \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}(v(\mathcal{C}_{\max}^{\geq T}) \geq \delta n^{2/3}) = 0. \quad (14.2.26)$$

Proof. We condition on the size of the components of the underlying configuration model. Choose $\eta > 0$ satisfying $\delta > 2\eta \mathbb{E}[DS] / \mathbb{E}[D]$. Then,

$$\begin{aligned} &\mathbb{P}(v(\mathcal{C}_{\max}^{\geq T}) > \delta n^{2/3}) \\ &= \mathbb{P}(v(\mathcal{C}_{\max}^{\geq T}) > \delta n^{2/3} \mid v^{(H)}(\mathcal{C}_{\max}^{\geq T}) \leq \eta n^{2/3}) \mathbb{P}(v^{(H)}(\mathcal{C}_{\max}^{\geq T}) \leq \eta n^{2/3}) \\ &\quad + \mathbb{P}(v(\mathcal{C}_{\max}^{\geq T}) > \delta n^{2/3} \mid v^{(H)}(\mathcal{C}_{\max}^{\geq T}) > \eta n^{2/3}) \mathbb{P}(v^{(H)}(\mathcal{C}_{\max}^{\geq T}) > \eta n^{2/3}) \\ &\leq \mathbb{P}(v(\mathcal{C}_{\max}^{\geq T}) > \delta n^{2/3} \mid v^{(H)}(\mathcal{C}_{\max}^{\geq T}) \leq \eta n^{2/3}) + \mathbb{P}(v^{(H)}(\mathcal{C}_{\max}^{\geq T}) > \eta n^{2/3}). \end{aligned} \quad (14.2.27)$$

By [71], for any $\eta > 0$,

$$\lim_{T \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}(v^{(H)}(\mathcal{C}_{\max}^{\geq T}) > \eta n^{2/3}) = 0, \quad (14.2.28)$$

so that the second term in (14.2.27) vanishes.

Now we study the first term in (14.2.27). Given any component $\mathcal{C}^{\geq T}$, we start exploring at a vertex of that component, until time $\eta n^{2/3}$. By Lemma 14.3, the probability that more than $\delta n^{2/3}$ vertices have been found at time $\eta n^{2/3}$ is quite small. Furthermore, we know that $\mathcal{C}^{\geq T}$ has been fully explored, since $v^{(H)}(\mathcal{C}_{\max}^{\geq T}) < \eta n^{2/3}$.

Then, by the union bound and by Lemma 14.3,

$$\begin{aligned}
 \mathbb{P}(v(\mathcal{C}_{\max}^{\geq T}) > \delta n^{2/3} \mid v^{(H)}(\mathcal{C}_{\max}^{\geq T}) \leq \eta n^{2/3}) & \\
 & \leq \sum_{j=1}^n \mathbb{P}(v(\mathcal{C}_j^{\geq T}) > \delta n^{2/3} \mid v^{(H)}(\mathcal{C}_{\max}^{\geq T}) \leq \eta n^{2/3}) \\
 & = \frac{\sum_{j=1}^n \mathbb{P}(v(\mathcal{C}_j^{\geq T}) > \delta n^{2/3}, v^{(H)}(\mathcal{C}_{\max}^{\geq T}) \leq \eta n^{2/3})}{\mathbb{P}(v^{(H)}(\mathcal{C}_{\max}^{\geq T}) \leq \eta n^{2/3})} \\
 & \leq \frac{\sum_{j=1}^n \mathbb{P}(v(\mathcal{C}_j^{\geq T}) > \delta n^{2/3}, v^{(H)}(\mathcal{C}_j^{\geq T}) \leq \eta n^{2/3})}{\mathbb{P}(v^{(H)}(\mathcal{C}_{\max}^{\geq T}) \leq \eta n^{2/3})} \\
 & \leq \frac{\sum_{j=1}^n \mathbb{P}(v(\mathcal{C}_j^{\geq T}) > \delta n^{2/3} \mid v^{(H)}(\mathcal{C}_j^{\geq T}) \leq \eta n^{2/3})}{\mathbb{P}(v^{(H)}(\mathcal{C}_{\max}^{\geq T}) \leq \eta n^{2/3})} \\
 & \leq \frac{ne^{-\zeta n^{2/3}/s_{\max}}}{\mathbb{P}(v^{(H)}(\mathcal{C}_{\max}^{\geq T}) \leq \eta n^{2/3})}, \tag{14.2.29}
 \end{aligned}$$

for some $\zeta > 0$. Since $s_{\max} \ll n^{2/3} / \log(n)$, using (14.2.28) and taking limits proves the lemma. \square

Proof of Theorem 14.1. By [71], the ordered excursions of the process $\bar{Q}_n(t)$ defined in (14.2.3) converge to γ^λ , the ordered excursions of a reflected Brownian motion with parabolic drift. Then, by Lemma 14.2 and (14.2.6), the ordered component sizes of the HCM that have been discovered up to time $Tn^{2/3}$ for some $T > 0$ converge to $\mathbb{E}[DS] / \mathbb{E}[D] \gamma^\lambda$. Combining this with Lemma 14.4 then shows that

$$n^{-2/3}(v(\mathcal{C}_{(j)}))_{j \geq 1} \rightarrow \frac{\mathbb{E}[DS]}{\mathbb{E}[S]} \gamma^\lambda, \tag{14.2.30}$$

in the product topology. Then, using that $N/n = \mathbb{E}[S_n]$ completes the proof of Theorem 14.1. \square

14.2.3 Convergence in ℓ_{\downarrow}^2 topology: Proof of Theorem 14.2

To prove Theorem 14.2, we show that the probability that a uniformly chosen vertex is in a large component is small, by using the Markov inequality. Thus, we need to bound the expected component size of a uniformly chosen vertex in a HCM. To do this, we bound the expected component size of a uniformly chosen community of size s and inter-community degree k in Lemma 14.6. To prove Lemma 14.6, we first count the number of paths in the macroscopic configuration model in Lemma 14.5: the number of paths from community to community, ignoring the internal community structures. Let $P_L^{(H_0)}$ be the set of all macroscopic paths of length L in a HCM, starting from community H_0 . Furthermore, define $PW_L^{(H_0)}$ as the number of macroscopic paths of length L , starting in H_0 , weighted by the size of the last community, i.e.,

$$PW_L^{(H_0)} = \sum_{P_L^{(H_0)}} s_{\text{last community}}. \tag{14.2.31}$$

Lemma 14.5. For any $L < \frac{1}{4}n$ and for some $K > 0$,

$$\mathbb{E} \left[PW_L^{(H_0)} \right] \leq K \frac{\mathbb{E} [D_n S_n]}{\mathbb{E} [D_n]} d_{H_0} v_{D_n}^{L-1}, \quad (14.2.32)$$

with v_{D_n} as defined in Condition 14.2(iii).

Proof. This proof is quite similar to the proof of [117, Lemma 5.1]. If $L = 1$, then the equation states that $\mathbb{E} \left[PW_1^{(H_0)} \right] \leq K \frac{\mathbb{E} [D_n S_n]}{\mathbb{E} [D_n]} d_{H_0}$, which is true, since there are at most d_{H_0} paths from H_0 , and the expected weight of the community at the end of the path is $\mathbb{E} [D_n S_n] / \mathbb{E} [D_n]$.

For $L \geq 2$, the path consists of communities H_0, H_1, \dots, H_L . This path consists of two half-edges at communities H_1, \dots, H_{L-1} , and one half-edge at the start and at the end of the path. The probability that these half-edges are paired as a path is $(\ell_n - 1)^{-1}(\ell_n - 3)^{-1} \dots (\ell_n - 2L + 1)^{-1}$. Therefore,

$$\mathbb{E} \left[PW_L^{(H_0)} \right] = \frac{d_{H_0} \sum_{i_1, \dots, i_L \in [n]}^* \prod_{j=1}^{L-1} d_{H_{i_j}} (d_{H_{i_j}} - 1) d_{H_{i_L}} s_{i_L}}{\prod_{j=1}^L (\ell_n - 2j + 1)}, \quad (14.2.33)$$

where \sum^* denotes the sum over distinct indices, since all communities in the path must be distinct. If we only sum over $i_L \neq \{0, i_1, \dots, i_{L-1}\}$, we obtain

$$\begin{aligned} \sum_{i_L \neq \{H_0, i_1, \dots, i_{L-1}\}} d_{H_{i_L}} s_{i_L} &= \sum_{i \in [n]} d_{H_i} s_i - d_{H_0} s_{H_0} - \sum_{j=1}^{L-1} d_{H_{i_j}} s_{i_j} \\ &\leq n \mathbb{E} [D_n S_n] - 2(L-1) - 1 \\ &\leq \ell_n \frac{\mathbb{E} [D_n S_n]}{\mathbb{E} [D_n]} - 2L - 1 \leq \frac{\mathbb{E} [D_n S_n]}{\mathbb{E} [D_n]} (\ell_n - 2L - 1), \end{aligned} \quad (14.2.34)$$

where we have used that $d_{H_{i_j}} \geq 2$ for $j = 1, \dots, L-1$ and that $s_i \geq 1$ for all i . Therefore,

$$\begin{aligned} \mathbb{E} \left[PW_L^{(H_0)} \right] &\leq \frac{\mathbb{E} [D_n S_n]}{\mathbb{E} [D_n]} \frac{d_{H_0} \sum_{i_1, \dots, i_{L-1}}^* \prod_{j=1}^{L-1} d_{H_{i_j}} (d_{H_{i_j}} - 1)}{\prod_{j=1}^{L-1} (\ell_n - 2j + 1)} \\ &\leq \frac{\mathbb{E} [D_n S_n]}{\mathbb{E} [D_n]} (n \mathbb{E} [D_n])^{-L+1} \frac{d_{H_0} \sum_{i_1, \dots, i_{L-1}}^* \prod_{j=1}^{L-1} d_{H_{i_j}} (d_{H_{i_j}} - 1)}{\prod_{j=1}^{L-1} (1 - 2j/\ell_n)}. \end{aligned} \quad (14.2.35)$$

By arguments of [117, Lemma 5.1]

$$\sum_{i_1, \dots, i_{L-1}}^* \prod_{j=1}^{L-1} d_{H_{i_j}} (d_{H_{i_j}} - 1) \leq (n \mathbb{E} [D_n (D_n - 1)])^{L-1} \prod_{j=0}^{L-2} \left(1 - \frac{j}{r} \right), \quad (14.2.36)$$

where r denotes the number of communities with inter-community degree larger than or equal to 2. Since $r \leq \frac{1}{2}\ell_n$,

$$\begin{aligned} \mathbb{E} \left[PW_L^{(H_0)} \right] &\leq \frac{\mathbb{E} [D_n S_n]}{\mathbb{E} [D_n]} \left(\frac{\mathbb{E} [D_n (D_n - 1)]}{\mathbb{E} [D_n]} \right)^{L-1} \frac{d_{H_0} \prod_{j=0}^{L-2} \left(1 - \frac{j}{r} \right)}{\prod_{j=1}^{L-1} \left(1 - \frac{2j}{\ell_n} \right)} \\ &\leq \frac{\mathbb{E} [D_n S_n]}{\mathbb{E} [D_n]} \nu_{D_n}^{L-1} d_{H_0} \frac{\prod_{j=0}^{L-2} \left(1 - \frac{2j}{\ell_n} \right)}{\prod_{j=1}^{L-1} \left(1 - \frac{2j}{\ell_n} \right)} \\ &\leq \frac{\mathbb{E} [D_n S_n]}{\mathbb{E} [D_n]} \nu_{D_n}^{L-1} d_{H_0} \left(1 - \frac{2L-2}{n\mathbb{E} [D_n]} \right)^{-1} \\ &\leq \frac{\mathbb{E} [D_n S_n]}{\mathbb{E} [D_n]} \nu_{D_n}^{L-1} d_{H_0} \left(1 - \frac{1}{2\mathbb{E} [D_n]} \right)^{-1}, \end{aligned} \tag{14.2.37}$$

where we have used that $L < \frac{1}{4}n$. This proves the claim, since $\mathbb{E} [D_n] > 1$. \square

Using Lemma 14.5, we can bound the expected component size in a HCM. We are interested in the expected component size of a randomly chosen community of size s and inter-community degree k , $v(\mathcal{C}_{(k,s)})$.

Lemma 14.6. *For some $C > 0$,*

$$\mathbb{E} \left[v(\mathcal{C}_{(k,s)}) \right] \leq s + C \frac{k}{1 - \nu_n} + o(1). \tag{14.2.38}$$

Proof. We split the expectation into two different parts,

$$\begin{aligned} \mathbb{E} \left[v(\mathcal{C}_{(k,s)}) \right] &= \mathbb{E} \left[v(\mathcal{C}_{(k,s)}) \mathbb{1}_{\{v^{(H)}(\mathcal{C}_{\max}) \leq \frac{1}{4}n\}} \right] \\ &\quad + \mathbb{E} \left[v(\mathcal{C}_{(k,s)}) \mathbb{1}_{\{v^{(H)}(\mathcal{C}_{\max}) > \frac{1}{4}n\}} \right]. \end{aligned} \tag{14.2.39}$$

We bound the first part similar to the argument in [120, Lemma 4.6]. For every community H' in the same component as community H_0 , there is at least one path between H_0 and H' . Furthermore, H' adds $s_{H'}$ vertices to the component. Therefore $\mathcal{C}(H_0)$, the connected component containing community H_0 , satisfies

$$v(\mathcal{C}(H_0)) \leq \sum_L PW_L^{(H_0)} \tag{14.2.40}$$

This yields

$$\mathbb{E} \left[v(\mathcal{C}_{(k,s)}) \right] \leq \sum_L \mathbb{E} \left[PW_L^{H(k,s)} \right], \tag{14.2.41}$$

where $H_{(k,s)}$ is a community of size s and inter-community degree k . The sum of the first term in (14.2.39) only goes up to $L = \frac{1}{4}n$, since the maximal path size is smaller than the maximal component size. Thus, by Lemma 14.5,

$$\mathbb{E} \left[v(\mathcal{C}_{(k,s)}) \mathbb{1}_{\{v^{(H)}(\mathcal{C}_{\max}) \leq \frac{1}{4}n\}} \right] \leq s + \sum_{L=1}^{\frac{1}{4}n} \mathbb{E} \left[PW_L^{H(k,s)} \right] \leq s + \frac{\mathbb{E} [D_n S_n]}{\mathbb{E} [D_n]} Kk \sum_{L=1}^{\infty} \nu_n^{L-1}$$

$$= s + \frac{\mathbb{E}[D_n S_n]}{\mathbb{E}[D_n]} \frac{kK}{1 - \nu_n}. \quad (14.2.42)$$

For the second term, we use that the maximal component size is bounded from above by the total number of vertices $N = \mathbb{E}[S_n]n$. Then we need to bound the probability that the maximal hierarchical component has size at least $\frac{1}{4}n$. This is the probability that the size of the largest component in a regular configuration model is larger than $\frac{1}{4}n$. We can use the same arguments as in [71, Lemma 14] to show that

$$\mathbb{P}(v^{(H)}(\mathcal{C}_{\max}) > n/4) \leq \frac{16\mathbb{E}[D_n]}{n(1 - \nu_n)} + o(n^{-1}). \quad (14.2.43)$$

This gives

$$\mathbb{E}\left[v(\mathcal{C}_{(k,s)}) \mathbb{1}_{\{v^{(H)}(\mathcal{C}_{\max}) > \frac{1}{4}n\}}\right] \leq \frac{16N\mathbb{E}[D_n]}{n(1 - \nu_n)} + o(1) \leq \frac{16k\mathbb{E}[S_n]\mathbb{E}[D_n]}{1 - \nu_n} + o(1), \quad (14.2.44)$$

for $k > 0$. Combining (14.2.43) and (14.2.44) then yields

$$\mathbb{E}\left[v(\mathcal{C}_{(k,s)})\right] \leq s + C \frac{k}{1 - \nu_n} + o(1), \quad (14.2.45)$$

for some $C > 0$. □

Proof of Theorem 14.2. We first show that $\mathbb{E}[S_n^2] = o(n^{1/3})$ is a necessary condition for convergence in the ℓ_{\downarrow}^2 -topology. Thus, we assume that $\mathbb{E}[S_n^2] \geq \varepsilon n^{1/3}$ for some $\varepsilon > 0$. Let j_T denote the maximal index such that $\mathcal{C}_{(1)}, \dots, \mathcal{C}_{(j_T)}$ have all been explored before time $Tn^{2/3}$. Then, for any $\delta > 0$,

$$\lim_{T \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}\left(\sum_{j \geq j_T} v(\mathcal{C}_{(j)})^2 \geq \delta n^{4/3}\right) = 0 \quad (14.2.46)$$

needs to hold for convergence in the ℓ_{\downarrow}^2 topology. Then

$$\begin{aligned} \sum_{j \geq j_k} v(\mathcal{C}_{(j)})^2 &\geq \sum_{j \geq 1} v(\mathcal{C}_j^{\geq T})^2 \geq \sum_{i \geq Tn^{2/3}} s_{(i)}^2 \\ &= \sum_s n_s^{\geq T} s^2 = \sum_s n_s s^2 - \sum_s n_s^{\leq T} s^2, \end{aligned} \quad (14.2.47)$$

where $n_s^{\geq T}$ and $n_s^{\leq T}$ denote the number of communities of size s , discovered after or before time $Tn^{2/3}$ respectively.

We can use a martingale argument similar to [71, Proposition 29], to show that

$$\sup_{u \leq t} \left| n^{-2/3} \sum_{i=1}^{\lfloor un^{2/3} \rfloor} s_{(i)}^2 - \frac{\sum_{k,s} ks^2 n_{k,s}}{\ell_n} u \right| = o_{\mathbb{P}}(n^{2/3}). \quad (14.2.48)$$

Therefore

$$\sum_s n_s^{\leq T} s^2 = Tn^{2/3} \frac{\sum_{k,s} ks^2 n_{k,s}}{\ell_n} + o_{\mathbb{P}}(n^{4/3}) = T \frac{o(n)}{\ell_n} \sum_s s^2 n_s + o_{\mathbb{P}}(n^{4/3}), \quad (14.2.49)$$

where we have used that $d_{\max} = o(n^{1/3})$. Therefore, we obtain

$$\sum_s s^2 n_s^{\geq T} = \sum_s s^2 n_s (1 - T o(1)) + o_{\mathbb{P}}(n^{4/3}) \geq \varepsilon n^{4/3} (1 - T o(1) + o_{\mathbb{P}}(1)). \quad (14.2.50)$$

Taking the limit first for $n \rightarrow \infty$, and then for $K \rightarrow \infty$ shows that

$$\lim_{T \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P} \left(\sum_{j \geq j_T} v(\mathcal{C}_j)^2 > \delta n^{4/3} \right) > 0 \quad (14.2.51)$$

for $\delta < \varepsilon$, hence the component sizes do not converge in the ℓ_{\downarrow}^2 -topology.

Now we show that $\mathbb{E} [S_n^2] = o(n^{1/3})$ is sufficient for convergence in the ℓ_{\downarrow}^2 -topology. Let i_T denote the index of the first component that is explored after time $Tn^{2/3}$. Then, for convergence in the ℓ_{\downarrow}^2 topology it is sufficient to show that for any $\delta > 0$,

$$\lim_{T \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P} \left(\sum_{i \geq i_T} v(\mathcal{C}_i)^2 \geq \delta n^{4/3} \right) = 0. \quad (14.2.52)$$

Let $G^{\geq T}$ denote the graph that is obtained by removing all components that have been explored before time $Tn^{2/3}$. To show that $\mathbb{E} [S_n^2] = o(n^{1/3})$ is sufficient for convergence in the ℓ_{\downarrow}^2 topology, we obtain using the Markov inequality

$$\begin{aligned} \mathbb{P} \left(\sum_{i \geq i_T} v(\mathcal{C}_i)^2 > \delta n^{4/3} \right) &\leq \frac{1}{\delta n^{4/3}} \mathbb{E} \left[\sum_{i \geq i_T} v(\mathcal{C}_i)^2 \right] = \frac{1}{\delta n^{1/3}} \mathbb{E} [v(\mathcal{C}^{\geq T}(V_n))] \\ &= \frac{1}{\delta n^{1/3}} \mathbb{E} \left[S_{H_n}^{\geq T} v(\mathcal{C}^{\geq T}(H_n)) \right], \end{aligned} \quad (14.2.53)$$

where V_n denotes a randomly chosen vertex of $G^{\geq T}$, and H_n denotes a randomly chosen community. Furthermore,

$$\mathbb{E} \left[S_{H_n}^{\geq T} v(\mathcal{C}^{\geq T}(H_n)) \right] = \sum_{k,s} p_{k,s}^{\geq T}(n) s \mathbb{E} \left[v(\mathcal{C}_{(k,s)}^{\geq T}) \right], \quad (14.2.54)$$

where $v(\mathcal{C}_{(k,s)}^{\geq T})$ denotes the size of a component where the first explored community has size s and inter-community degree k . By [71], the criticality parameter of $G^{\geq T}$, \bar{v}_n , satisfies

$$\bar{v}_n \leq \nu_{D_n} - CTn^{-1/3} + o_{\mathbb{P}}(n^{-1/3}), \quad (14.2.55)$$

with ν_{D_n} as in Condition 14.2(iii). Then, combining Lemma 14.6 and (14.2.54) gives

$$\mathbb{E} [S_{H_n} v(\mathcal{C}^{\geq T}(H_n))] \leq \mathbb{E} \left[(S_n^{\geq T})^2 \right] + K \mathbb{E} [D_n^{\geq T} S_n^{\geq T}] \frac{n^{1/3}}{CT - \lambda}. \quad (14.2.56)$$

Furthermore, $\mathbb{E} [(S_n^{\geq T})^2] \leq \mathbb{E} [S_n^2] / (n - Tn^{2/3}) = \mathbb{E} [S_n^2] (1 + O(n^{-1/3}))$. By assumption, $\mathbb{E} [S_n^2] = o(n^{1/3})$. Combining this with (14.2.53) and (14.2.56) yields

$$\mathbb{P} \left(\sum_{i \geq i_T} v(\mathcal{C}_i)^2 > \delta n^{4/3} \right) \leq o_{\mathbb{P}}(1) + \frac{K}{\delta(CT - \lambda)}, \quad (14.2.57)$$

so that

$$\lim_{T \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P} \left(\sum_{i \geq i_T} v(\mathcal{C}_i)^2 > \delta n^{4/3} \right) = 0. \quad (14.2.58)$$

Thus, if $\mathbb{E} [S_n^2] = o(n^{1/3})$, then Theorem 14.1 also holds in the ℓ_{\downarrow}^2 topology. \square

14.3 Percolation on the HCM

In this section we prove Theorem 14.3, which identifies the scaling limit for the cluster sizes of a HCM under critical percolation. As described in Section 14.1.3, it is convenient to percolate first only the edges inside communities. This percolation results in a HCM with percolated communities. These percolated communities may be disconnected. However, if we define the connected components of the percolated communities as new communities, we retrieve an updated HCM. After this, we percolate the inter-community connections. These edges are distributed as in the CM. Therefore, for this second step of percolation, we follow a similar approach as in [115]. Combining these two steps of percolation results in the following algorithm that constructs a percolated HCM:

Algorithm 14.2.

- (S1) For each community H , remove every intra-community edge of H independently with probability $1 - \pi$. Let \bar{n} denote the number of connected components of communities after percolation inside the communities. Then, define the connected components of the percolated communities as new communities $(H_i^{(\pi)})_{i \in [\bar{n}]}$.
- (S2) Let $H_e^{(\pi)}$ be the percolated community attached to inter-community half-edge e . Then, every inter-community half-edge e explodes with probability $1 - \sqrt{\pi}$, it detaches from $H_e^{(\pi)}$, and is associated to a new community $H_e^{I(\pi)}$ of the same shape, but with e as its only inter-community half-edge. Let n_{H+} denote the number of new communities of shape H that are created in this way, and $\tilde{n} = \bar{n} + \sum_H n_{H+}$. Let $(\tilde{H}_i^{(\pi)})_{i \in [\tilde{n}]}$ be the new communities after detaching the half-edges.
- (S3) Construct a hierarchical configuration model with community sequence $(\tilde{H}_i^{(\pi)})_{i \in [\tilde{n}]}$.
- (S4) For all community shapes H , delete the exploded communities with inter-community degree one.

Figure 14.3 illustrates Algorithm 14.2. By [115], a similar algorithm creates a percolated CM. Therefore, by adding the extra step of percolation inside communities, Algorithm 14.2 creates a percolated HCM. In this bond-percolation procedure, there are three sources of randomness: the percolated communities $H^{(\pi)}$, the explosion procedure, and the pairing of the edges to construct a HCM.

Remark 14.8. In percolation on the regular configuration model, instead of deleting the exploded vertices n_+ , it is also possible to choose n_+ vertices uniformly at random from all vertices with degree one, and to delete them [115]. This procedure also results in a multigraph with the same distribution as a percolated configuration model. Similar to this, in our setting it is possible to replace step (S4) of Algorithm 14.2 by

- (S4') For all community shapes H , choose n_{H+} communities uniformly at random from all communities of shape H and inter-community degree one. Delete these communities.

Remark 14.9. In the CM, each exploded half-edge is replaced by a single half-edge attached to a new vertex. In Algorithm 14.2, each exploded inter-community half-edge is attached to a new community of the same shape as the original community,

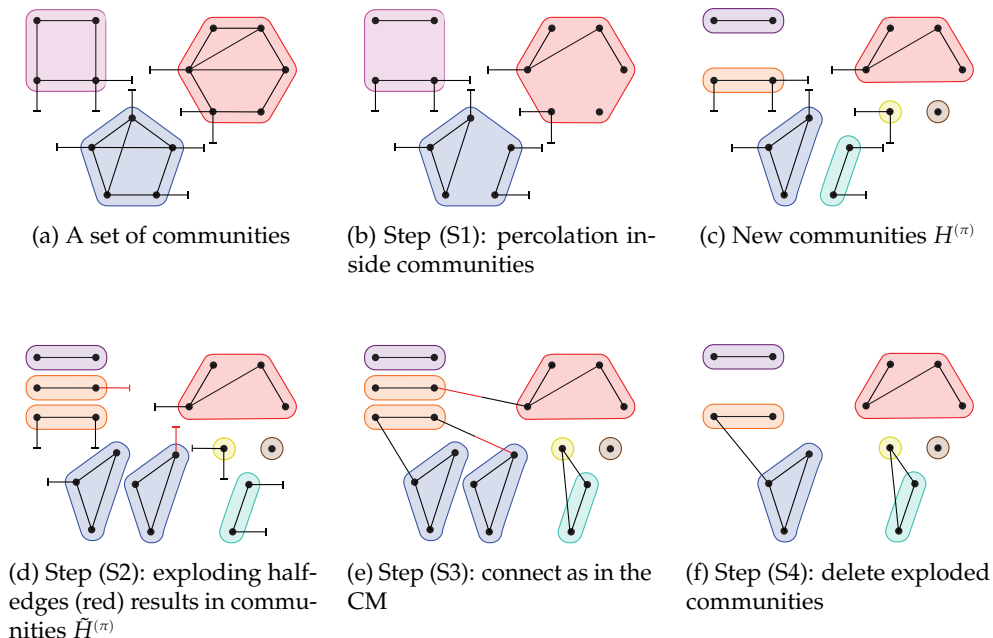


Figure 14.3: Illustration of Algorithm 14.2. In this example $n = 3, \bar{n} = 7, \tilde{n} = 9$.

but with only one half-edge adjacent to it (see Figure 14.3d). This difference is caused by the fact that in the HCM, communities of the same inter-community degree may not be equal. Different communities with inter-community degree k may have different sizes. The effect on the component sizes of percolating a half-edge of a community of inter-community degree k depends on the community size. Percolating the half-edge adjacent to a larger community has more effect on the component sizes than percolating the half-edges adjacent to a smaller community. For this reason, we replace exploded half-edges by half-edges attached to a community of the same size as the original community, instead of replacing it by a vertex of degree one.

14.3.1 The sizes of critical percolation clusters

We now analyze Algorithm 14.2 to prove Theorem 14.3. Let $S_n^{(\pi_n)}$ and $D_n^{(\pi_n)}$ denote the size and degree of communities after percolation only inside the communities with probability π_n , and $S^{(\pi)}$ and $D^{(\pi)}$ their infinite size limits. Furthermore, let $g(H, v, k, \pi)$ denote the probability that after percolating community H with parameter π , the connected component containing vertex v contains k half-edges. By (11.3.1) of Chapter 11,

$$\mathbb{P}(D^{(\pi)} = k) = \frac{\sum_H \sum_{v \in V_H} P(H) d_v^{(b)} g(H, v, k, \pi) / k}{\sum_H \sum_{v \in V_H} \sum_l P(H) d_v^{(b)} g(H, v, l, \pi) / l}. \quad (14.3.1)$$

We denote the number of communities in the original graph by n , the number of communities after percolating only the intra-community edges by \bar{n} , and the number

of communities after the explosion procedure by \tilde{n} . After percolation inside the communities, the number of vertices N is still the same. Furthermore, similarly to [71], let $\mathbb{P}_{\pi_n}^{\tilde{n}}$ denote the probability measure containing the shapes of the exploded communities after Algorithm 14.2, step (S2). Then \mathbb{P}_π denotes the product measure of $(\mathbb{P}_{\pi_n}^{\tilde{n}})_{\tilde{n} \geq 1}$. Since $n_{H+} \sim \text{Bin}(d_H \tilde{n}_H, 1 - \sqrt{\pi_n})$, a.s. with respect to \mathbb{P}_π

$$n_{H+} = d_H \tilde{n}_H (1 - \sqrt{\pi_n}) + o(d_H \tilde{n}_H). \quad (14.3.2)$$

Therefore,

$$\frac{\tilde{n}}{\bar{n}} = 1 + \frac{\sum_H n_{H+}}{\bar{n}} = 1 + \mathbb{E} [D^{(\pi)}] (1 - \sqrt{\pi}) + o(1), \quad (14.3.3)$$

a.s. with respect to \mathbb{P}_π .

The following lemma proves that the HCM with community sequence $(\tilde{H}_i^{(\pi_n)})_{i \in [\tilde{n}]}$ satisfies Conditions 14.1 and 14.2, so that we can apply Theorem 14.1 to find its component sizes:

Lemma 14.7. *Let G be a hierarchical configuration model satisfying Condition 14.3 with community sequence $(H_i)_{i \in [n]}$. Then the hierarchical configuration model with community sequence $(\tilde{H}_i^{(\pi_n)})_{i \in [\tilde{n}]}$, constructed as described in Algorithm 14.2, satisfies Conditions 14.1 and 14.2.*

Proof. By (14.3.1),

$$\mathbb{E} \left[(D_n^{(\pi_n)})^3 \right] = \frac{\sum_H \sum_{v \in V_H} \sum_k P_n(H) d_v^{(b)} g(H, v, k, \pi_n) k^2}{\sum_H \sum_{v \in V_H} \sum_l P_n(H) d_v^{(b)} g(H, v, k, \pi_n) / l}. \quad (14.3.4)$$

Let $H_v^{(\pi_n)}$ denote the connected component of the percolated community H containing vertex v . Then,

$$\begin{aligned} \sum_{v \in V_H} \sum_k P_n(H) d_v^{(b)} g(H, v, k, \pi_n) k^2 &= \sum_{v \in V_H} P_n(H) d_v^{(b)} \mathbb{E} \left[(\# \text{ outgoing edges of } H_v^{(\pi_n)})^2 \right] \\ &\leq \sum_{v \in V_H} P_n(H) d_v^{(b)} d_H^2 = P_n(H) d_H^3. \end{aligned} \quad (14.3.5)$$

To show that $\mathbb{E} \left[(D_n^{(\pi_n)})^3 \right]$ converges, we use the General Lebesgue Dominated Convergence Theorem (see for example [191, Thm. 19]), which states that if $|f_n(x)| \leq g_n(x)$ for all $x \in E$, $\sum_{x \in E} g_n(x) \rightarrow \sum_{x \in E} g(x) < \infty$, and f_n converges pointwise to f , then also $\sum_{x \in E} f_n(x) \rightarrow \sum_{x \in E} f(x)$. By Condition 14.2, $\mathbb{E} [D_n^3] \rightarrow \mathbb{E} [D^3]$, so by the General Lebesgue Dominated Convergence Theorem and (14.3.5), $\mathbb{E} \left[(D_n^{(\pi_n)})^3 \right] \rightarrow \mathbb{E} [(D^{(\pi)})^3]$. Similarly,

$$\mathbb{E} \left[D_n^{(\pi_n)} S_n^{(\pi_n)} \right] = \frac{\sum_H \sum_{v \in V_H} \sum_k P_n(H) d_v^{(b)} g(H, v, k, \pi_n) s_{H_v^{(\pi)}}}{\sum_H \sum_{v \in V_H} \sum_l P_n(H) d_v^{(b)} g(H, v, k, \pi_n) / l}. \quad (14.3.6)$$

We can bound the summands in the numerator as

$$\begin{aligned} \sum_{v \in V_H} \sum_k P_n(H) d_v^{(b)} g(H, v, k, \pi_n) s_{H_v^{(\pi)}} &\leq \sum_{v \in V_H} \sum_k P_n(H) d_v^{(b)} g(H, v, k, \pi_n) s_H \\ &= P_n(H) d_{HS_H}, \end{aligned} \tag{14.3.7}$$

so that again by the General Lebesgue Dominated Convergence Theorem and Condition 14.1 $\mathbb{E} [S_n^{(\pi_n)} D_n^{(\pi_n)}] \rightarrow \mathbb{E} [S^{(\pi)} D^{(\pi)}]$. By a similar reasoning $\mathbb{E} [S_n^{(\pi_n)}] \rightarrow \mathbb{E} [S^{(\pi)}]$. Thus, we have proved that $D_n^{(\pi_n)}$ and $S_n^{(\pi_n)}$ satisfy Conditions 14.1 and 14.2(i). Hence, after percolating inside the communities, the HCM still satisfies these conditions.

We want to prove that $\tilde{D}_n^{(\pi_n)}$ and $\tilde{S}_n^{(\pi_n)}$ also satisfy Conditions 14.1 and 14.2(i), so that after the explosion process the conditions are still satisfied. Since $D_n^{(\pi_n)}$ satisfies Condition 14.2, [71, Lemma 24] shows that $\tilde{D}_n^{(\pi_n)}$ also satisfies Condition 14.2.

Now we prove the convergence of the first moment of $\tilde{S}_n^{(\pi_n)}$. After explosion, the first \tilde{n} entries of $(\tilde{S}_i^{(\pi_n)})_{i \in [\tilde{n}]}$ are the same as in $(S_i^{(\pi_n)})_{i \in [\tilde{n}]}$, since the community sizes are not changed when percolating the inter-community edges. Furthermore, there are n_{H+} duplicated communities of shape H . Thus, the limiting distribution $(\tilde{S}^{(\pi)}, \tilde{D}^{(\pi)})$ can be written as

$$\begin{aligned} \mathbb{P} (\tilde{S}^{(\pi)} = s, \tilde{D}^{(\pi)} = k) &= \frac{\mathbb{P} (S^{(\pi)} = s, D^{(\pi)} = k)}{1 + \mathbb{E} [D^{(\pi)}] (1 - \sqrt{\pi})} \\ &\quad + \frac{\mathbb{1}_{\{k=1\}} \sum_j j (1 - \sqrt{\pi}) \mathbb{P} (S^{(\pi)} = s, D^{(\pi)} = j)}{1 + \mathbb{E} [D^{(\pi)}] (1 - \sqrt{\pi})}. \end{aligned} \tag{14.3.8}$$

By (14.3.2) and (14.3.3),

$$\begin{aligned} \frac{1}{\tilde{n}} \sum_{i \in [\tilde{n}]} \tilde{s}_i^{(\pi_n)} &= \frac{1}{\tilde{n}} \left(\sum_{i \in [\tilde{n}]} s_i^{(\pi_n)} + \sum_{i=\tilde{n}+1}^{\tilde{n}} \tilde{s}_i^{(\pi_n)} \right) = \frac{\tilde{n}}{\tilde{n}} \mathbb{E} [S_n^{(\pi_n)}] + \frac{\sum_H s_{H^{\mathbb{N}H+}}}{\tilde{n}} \\ &= \frac{\mathbb{E} [S_n^{(\pi_n)}] + (1 - \sqrt{\pi_n}) \mathbb{E} [D_n^{(\pi_n)} S_n^{(\pi_n)}]}{1 + \mathbb{E} [D_n^{(\pi_n)}] (1 - \sqrt{\pi_n})} + o(1), \end{aligned} \tag{14.3.9}$$

so that $\mathbb{E} [\tilde{S}_n^{(\pi_n)}] \rightarrow \mathbb{E} [\tilde{S}^{(\pi)}]$. Furthermore,

$$\frac{1}{\tilde{n}} \sum_{i \in [\tilde{n}]} \tilde{s}_i^{(\pi_n)} \tilde{d}_{H_i}^{(\pi_n)} = \frac{1}{\tilde{n}} \sum_{i \in [\tilde{n}]} s_i^{(\pi_n)} d_{H_i}^{(\pi_n)}, \tag{14.3.10}$$

and therefore the combined moment also converges, and Condition 14.1 is satisfied.

To prove Condition 14.2(iii), note that

$$\begin{aligned} \nu_{\tilde{D}_n^{(\pi_n)}} &= \frac{\sum_{i \in [\tilde{n}]} \tilde{d}_{H_i}^{(\pi_n)} (\tilde{d}_{H_i}^{(\pi_n)} - 1)}{\sum_{i \in [\tilde{n}]} \tilde{d}_{H_i}^{(\pi_n)}} = \frac{\pi_n \sum_{i \in [\tilde{n}]} d_{H_i}^{(\pi_n)} (d_{H_i}^{(\pi_n)} - 1) + o(n^{2/3})}{\sum_{i \in [\tilde{n}]} d_{H_i}^{(\pi_n)}} \\ &= \pi_n \nu_{D_n^{(\pi_n)}} + o(n^{-1/3}) = 1 + \lambda n^{-1/3} + o(n^{-1/3}), \end{aligned} \tag{14.3.11}$$

where the second equality follows from [71, equation (7.2)]. □

Remark 14.10. In Lemma 14.7, we have assumed that the HCM satisfies Conditions 14.1 and 14.2 (i) and (ii) before percolation. Then, we have shown that $S_n^{(\pi)}$ and $D_n^{(\pi)}$ also satisfy these conditions. However, it is also possible to assume from the start that $S_n^{(\pi)}$ and $D_n^{(\pi)}$ satisfy these conditions. This means for example that the inter-community degrees only need to have finite third moment after percolating inside the communities, they may have an infinite third moment before percolating inside the communities. We gave an example of such a community in Remark 14.7.

Proof of Theorem 14.3. After explosion, the HCM satisfies the assumptions of Theorem 14.1 by Lemma 14.7. Then Theorem 14.1 gives the component sizes of the exploded HCM. To obtain the sizes of the components of the percolated HCM, we need to know how many vertices are deleted in the last step of Algorithm 14.2. We denote the components of the graph after step (S3) of Algorithm 14.2 by \mathcal{C} , and the percolated components after step (S4) of the algorithm by \mathcal{C}' . Furthermore, we denote the number of vertices that are deleted in step (S4) from component \mathcal{C} by $v^d(\mathcal{C})$. If a community of size s is deleted, s vertices are deleted. Thus, this number can be written as

$$v^d(\mathcal{C}) = \sum_{i=1}^{v^{(H)}(\mathcal{C})} s_{\tilde{H}_i} \mathbb{1}_{\{\tilde{H}_i \text{ is deleted}\}}. \quad (14.3.12)$$

Let $\tilde{n}_{H,1}$ denote the number of communities of shape H and inter-community degree one. Using [71, Proposition 29] with $\alpha = 2/3$ and $f_n(i)$ as the indicator function that community i is of shape H and has inter-community degree one, we can show that the number of communities of shape H with inter-community degree one in component \mathcal{C} satisfies

$$v_{H,1}^{(H)}(\mathcal{C}) = v^{(H)}(\mathcal{C}) \frac{\tilde{n}_{H,1}}{\tilde{n}\mathbb{E}[\tilde{D}^{(\pi n)}]} + o_{\mathbb{P}_\pi} \left(n^{1/3} \frac{\tilde{n}_{H,1}}{\sum_H \tilde{n}_{H,1}} \right). \quad (14.3.13)$$

Therefore, the number of vertices in communities of shape H with inter-community degree one in a component \mathcal{C} satisfies

$$v_{H,1}(\mathcal{C}) = v^{(H)}(\mathcal{C}) \frac{s_H \tilde{n}_{H,1}}{\tilde{n}\mathbb{E}[\tilde{D}^{(\pi n)}]} + o_{\mathbb{P}_\pi} \left(n^{1/3} \frac{s_H \tilde{n}_{H,1}}{\sum_H \tilde{n}_{H,1}} \right). \quad (14.3.14)$$

A fraction of $n_{H+}/\tilde{n}_{H,1}$ communities of shape H with outside degree one is removed uniformly. Therefore, for j fixed,

$$\begin{aligned} v^d(\mathcal{C}_{(j)}^{\tilde{\mathcal{C}}}) &= v^{(H)}(\mathcal{C}_{(j)}^{\tilde{\mathcal{C}}}) \frac{\sum_H s_H n_{H+}}{\tilde{n}\mathbb{E}[\tilde{D}^{(\pi n)}]} + o_{\mathbb{P}_\pi} \left(n^{1/3} \frac{\sum_H s_H \tilde{n}_{H+}}{\sum_H \tilde{n}_{H,1}} \right) + o_{\mathbb{P}_\pi} \left(\sum_H n_{H+} \right) \\ &= v^{(H)}(\mathcal{C}_{(j)}^{\tilde{\mathcal{C}}}) \frac{\sum_H s_H n_{H+}}{\tilde{n}\mathbb{E}[\tilde{D}^{(\pi n)}]} + o_{\mathbb{P}_\pi}(n^{1/3}) + o_{\mathbb{P}_\pi}(n^{2/3}), \end{aligned} \quad (14.3.15)$$

where we have used (14.3.2) and the fact that $\tilde{n}_{H,1} \geq n_{H+}$. Thus, by (14.3.15), (14.3.2), (14.3.10) and (14.1.4),

$$\begin{aligned}
 v^d(\mathcal{C}_{(j)}^{\tilde{\mathcal{C}}}) &= v^{(H)}(\mathcal{C}_{(j)}^{\tilde{\mathcal{C}}}) \frac{\sum_H s_H n_{+H}}{\tilde{n} \mathbb{E} [\tilde{D}^{(\pi_n)}]} + o_{\mathbb{P}_\pi}(n^{2/3}) \\
 &= \frac{(1 - \sqrt{\pi_n}) \sum_H d_H s_H \tilde{n}_H}{\tilde{n} \mathbb{E} [\tilde{D}^{(\pi_n)}]} v^{(H)}(\mathcal{C}_{(j)}^{\tilde{\mathcal{C}}}) + o_{\mathbb{P}_\pi}(n^{2/3}) \\
 &= (1 - \sqrt{\pi_n}) \frac{\sum_{k,s} ks \tilde{n}_{k,s}}{\tilde{n} \mathbb{E} [\tilde{D}^{(\pi_n)}]} v^{(H)}(\mathcal{C}_{(j)}^{\tilde{\mathcal{C}}}) + o_{\mathbb{P}_\pi}(n^{2/3}) \\
 &= (1 - \sqrt{\pi_n}) \frac{\sum_{k,s} ks \tilde{n}_{k,s}}{\tilde{n} \mathbb{E} [\tilde{D}^{(\pi_n)}]} v^{(H)}(\mathcal{C}_{(j)}^{\tilde{\mathcal{C}}}) + o_{\mathbb{P}_\pi}(n^{2/3}) \\
 &= (1 - \sqrt{\pi_n}) \frac{\mathbb{E} [\tilde{D}^{(\pi_n)} \tilde{S}^{(\pi_n)}]}{\mathbb{E} [\tilde{D}^{(\pi_n)}]} v^{(H)}(\mathcal{C}_{(j)}^{\tilde{\mathcal{C}}}) + o_{\mathbb{P}_\pi}(n^{2/3}) \\
 &= (1 - \sqrt{\pi_n}) v(\mathcal{C}_{(j)}^{\tilde{\mathcal{C}}}) + o_{\mathbb{P}_\pi}(n^{2/3}). \tag{14.3.16}
 \end{aligned}$$

We now use that the components \mathcal{C}' after step (S4) of the algorithm are obtained by removing $v^d(\mathcal{C}')$ vertices from component \mathcal{C} so that

$$v(\mathcal{C}'_{(j)}) = \sqrt{\pi_n} v(\mathcal{C}_{(j)}^{\tilde{\mathcal{C}}}) + o_{\mathbb{P}_\pi}(n^{2/3}). \tag{14.3.17}$$

Then, Theorem 14.1 gives

$$\tilde{n}^{-2/3} (v(\mathcal{C}'_{(j)}))_{j \geq 1} \xrightarrow{d} \frac{\mathbb{E} [\tilde{D}^{(\pi)} \tilde{S}^{(\pi)}]}{\mathbb{E} [\tilde{S}^{(\pi)}]} \sqrt{\pi} \tilde{\gamma}^\lambda. \tag{14.3.18}$$

Noting that $N = \sum_{i=1}^{\tilde{n}} \tilde{s}_i^{(\pi_n)}$ leads to

$$\frac{N}{\tilde{n}} \xrightarrow{\mathbb{P}} \mathbb{E} [\tilde{S}^{(\pi)}], \tag{14.3.19}$$

so that (14.1.8) follows. □

14.3.2 The critical window

Equation (14.1.7) gives an implicit equation for the critical window. We want to know whether it is possible to write (14.1.7) in the form

$$\pi_n(\lambda) = \pi_n(0) \left(1 + \frac{\lambda c^*}{n^{1/3}} \right) + o(n^{-1/3}), \tag{14.3.20}$$

for some $c^* \in \mathbb{R}$, so that the width of the critical window in the hierarchical configuration model is similar to the width of the critical window in the configuration model.

Since $g(H, v, k, \pi_n(\lambda))$ is not necessarily increasing in λ , we rewrite (14.1.7) as

$$\pi_n(\lambda) = \frac{\mathbb{E}[D_n]}{\sum_H P_n(H) \sum_{v \in V_H} d_v^{(b)} \sum_{k=1}^{D_H-1} B(H, v, k+1, \pi_n(\lambda))} \left(1 + \frac{\lambda}{n^{1/3}} \right), \tag{14.3.21}$$

where $B(H, v, k, \pi_n(\lambda))$ is the probability that after percolating H with probability $\pi_n(\lambda)$, the connected component of v contains at least k inter-community half-edges, which is increasing in λ .

Lemma 14.8. *For a hierarchical configuration model satisfying Condition 14.3 as well as $\lim_{n \rightarrow \infty} \mathbb{E}[D_n^2 S_n] = \mathbb{E}[D^2 S] < \infty$,*

$$\pi_n(\lambda) = \pi_n(0) \left(1 + \frac{\lambda c^*}{n^{1/3}} \right) + o(\lambda n^{-1/3}), \quad (14.3.22)$$

where

$$c^* = \frac{\mathbb{E}[D]}{\mathbb{E}[D] + \pi^2 \sum_H P(H) \sum_v d_v^{(b)} \sum_k \frac{d}{dp} B(H, v, k+1, p)_{p=\pi}}. \quad (14.3.23)$$

Remark 14.11. Equation (14.3.23) shows that $c^* \leq 1$, so that the critical window of a HCM is smaller than the critical window of a CM where no communities are inserted. Here $\frac{d}{dp} B(H, v, k, p)_{p=\pi}$ captures how vulnerable community H is to percolation inside the community. The larger $\frac{d}{dp} B(H, v, k, p)_{p=\pi}$ will be, the larger the difference between λ and λc^* will be. Intuitively, when $\frac{d}{dp} B(H, v, k, p)_{p=\pi}$ is small, this indicates that changing the percolation probability changes the degrees of the percolated communities very little. Therefore, the critical behavior is almost entirely explained by the macroscopic CM in that case. On the other hand, when $\frac{d}{dp} B(H, v, k, p)_{p=\pi}$ is large, increasing the percolation probability by a small amount increases the degrees of the communities by a lot. Then λc^* may be much smaller than λ .

Proof. We can write (14.3.21) as

$$\pi_n(\lambda) = L_n(\pi_n(\lambda)) \left(1 + \frac{\lambda}{n^{1/3}} \right), \quad (14.3.24)$$

where

$$L_n(\pi_n(\lambda)) = \frac{\mathbb{E}[D_n]}{\sum_H P_n(H) \sum_{v \in V_H} d_v^{(b)} \sum_{k=1}^{D_H-1} B(H, v, k+1, \pi_n(\lambda))}. \quad (14.3.25)$$

Calculating the derivative gives

$$\pi'_n(\lambda) = \frac{L_n(\pi_n(\lambda))}{n^{1/3} (1 - L'_n(\pi_n(\lambda)) (1 + \lambda/n^{1/3}))} \quad (14.3.26)$$

Then, by the mean value theorem, there exists $\lambda^* \in [0, \lambda]$ such that

$$\pi_n(\lambda) = \pi_n(0) + \frac{\lambda}{n^{1/3}} \frac{L_n(\pi_n(\lambda^*))}{1 - L'_n(\pi_n(\lambda^*)) (1 + \lambda^*/n^{1/3})}. \quad (14.3.27)$$

Since $B(H, v, k, \pi)$ is the probability of an increasing event, $L_n(\pi_n(\lambda))$ is continuous. Calculating the derivative of $L_n(\pi_n(\lambda))$ gives

$$L'_n(\pi_n(\lambda)) = - \frac{\mathbb{E}[D_n] \sum_H P_n(H) \sum_v d_v^{(b)} \sum_k B'(H, v, k+1, \pi_n(\lambda))}{\left(\sum_H P_n(H) \sum_v d_v^{(b)} \sum_k B(H, v, k+1, \pi_n(\lambda)) \right)^2}, \quad (14.3.28)$$

where we have denoted

$$B'(H, v, k, \pi) = \frac{d}{dp} B(H, v, k, p)_{p=\pi}. \quad (14.3.29)$$

Since B is an increasing function of the percolation parameter p , $L'_n(\pi_n(\lambda)) \leq 0$. By Theorem 11.2

$$v_D^{(n)}(\pi_n) = \frac{\sum_H P_n(H) \sum_{v \in V_H} d_v^{(b)} \sum_{k=1}^{D_H-1} B(H, v, k+1, \pi_n(\lambda))}{\mathbb{E}[D_n]}, \quad (14.3.30)$$

Therefore, by (14.1.6) and (14.3.30),

$$\begin{aligned} |L'_n(\pi_n(\lambda))| &= \frac{\mathbb{E}[D_n] \sum_H P_n(H) \sum_v d_v^{(b)} \sum_k B'(H, v, k+1, \pi_n(\lambda))}{\left(\sum_H P_n(H) \sum_v d_v^{(b)} \sum_k B(H, v, k+1, \pi_n(\lambda)) \right)^2} \\ &\leq \frac{\sum_H P_n(H) \sum_v d_v^{(b)} \sum_k B'(H, v, k+1, \pi_n(\lambda))}{\mathbb{E}[D_n]}. \end{aligned} \quad (14.3.31)$$

Hence, we need to bound $B'(H, v, k, \pi_n(\lambda))$. The event that v is connected to at least k half-edges is increasing in λ . Let $\mathcal{E}(H, v, k)$ denote the event that vertex v is connected to at least k half-edges of community H . An edge is pivotal for $\mathcal{E}(H, v, k)$ in a certain configuration, if the event occurs when the edge is made present, and the event does not occur when the edge is removed. The event that edge e is pivotal does not depend on the actual edge status of e . Furthermore, pivotal edges cannot form a cycle, because otherwise $\mathcal{E}(H, v, k)$ can still occur after removing one of the pivotal edges on the cycle. Then, by Russo's formula [192],

$$\begin{aligned} B'(H, v, k, \pi_n(\lambda)) &= \sum_{e \in H} \mathbb{P}_{\pi_n(\lambda)}(e \text{ pivotal for } \mathcal{E}(H, v, k+1)) \\ &= \frac{1}{\pi_n(\lambda)} \sum_{e \in H} \mathbb{P}_{\pi_n(\lambda)}(e \text{ present and pivotal for } \mathcal{E}(H, v, k+1)) \\ &= \frac{1}{\pi_n(\lambda)} \mathbb{E}_{\pi_n(\lambda)}[\# \text{ pivotal, present edges for } \mathcal{E}(H, v, k+1)] \\ &\leq \frac{1}{\pi_n(\lambda)} (S_H - 1), \end{aligned} \quad (14.3.32)$$

because at most $S_H - 1$ pivotal edges can be present in a community, since otherwise, they would form a cycle. Therefore,

$$|L'_n(\pi_n(\lambda))| \leq \frac{\sum_H P_n(H) \sum_v d_v^{(b)} \sum_{k=1}^{d_H-1} (S_H - 1)}{\mathbb{E}[D_n] \pi_n(\lambda)} \leq \frac{\sum_H P_n(H) d_H^2 S_H}{\mathbb{E}[D_n] \pi_n(\lambda)}. \quad (14.3.33)$$

Since $\mathbb{E}[D_n^2 S_n] \rightarrow \mathbb{E}[D^2 S]$, we can use the General Lebesgue Dominated Convergence Theorem to conclude that

$$\begin{aligned} \lim_{n \rightarrow \infty} L'_n(\pi_n(\lambda^*)) &= - \frac{\mathbb{E}[D] \sum_H P(H) \sum_v d_v^{(b)} \sum_k B'(H, v, k+1, \pi)}{\left(\sum_H P(H) \sum_v d_v^{(b)} \sum_k B(H, v, k+1, \pi) \right)^2} \\ &= - \frac{\pi^2 \sum_H P(H) \sum_v d_v^{(b)} \sum_k B'(H, v, k+1, \pi)}{\mathbb{E}[D]}, \end{aligned} \quad (14.3.34)$$

where the last equality follows from (14.3.30) and (14.1.7). Furthermore, we can use the General Lebesgue Dominated Convergence Theorem (see the proof of Theorem 11.2) to conclude that

$$\lim_{n \rightarrow \infty} L_n(\pi_n(\lambda)) = L(\pi) = \pi. \quad (14.3.35)$$

Inserting this into (14.3.27) proves (14.3.22). \square

Example 14.1 (Star-shaped communities). We now consider the case where all communities are star-shaped, so that every community has one vertex in the middle, connected to l other vertices that all have inter-community degree one (as in Figure 14.4). Then (14.1.7) becomes

$$\pi_n(\lambda) = \frac{1}{(l-1)\pi_n(\lambda)^2} \left(1 + \frac{\lambda}{n^{1/3}}\right), \quad (14.3.36)$$

or

$$\pi_n(\lambda) = \frac{1}{(l-1)^{1/3}} \left(1 + \frac{\lambda}{n^{1/3}}\right)^{1/3}. \quad (14.3.37)$$

A first order Taylor approximation then gives

$$\pi_n(\lambda) = \frac{1}{(l-1)^{1/3}} \left(1 + \frac{\lambda}{3n^{1/3}}\right) + O(n^{-2/3}), \quad (14.3.38)$$

so that $c^* = 1/3$, which is the same result that is obtained when computing (14.3.22). Table 14.1 compares the approximation of (14.3.22) with the exact values of $\pi_n(\lambda)$ from (14.3.37).

λ	$n = 10^5$		$n = 10^6$	
	$\pi_n(\lambda)$	$\pi_n(\lambda)$ appr.	$\pi_n(\lambda)$	$\pi_n(\lambda)$ appr.
-10	0,581	0,585	0,608	0,609
-1	0,625	0,625	0,628	0,628
0	0,630	0,630	0,630	0,630
1	0,634	0,634	0,632	0,632
10	0,672	0,675	0,650	0,651

Table 14.1: Values of $\pi_n(\lambda)$ for star-shaped communities with 5 end points, and the approximation by (14.3.22).

Example 14.2 (Line communities). We now consider the case where all communities are either line communities of length 5 (as in Figure 14.5), or single vertices of degree 3, both with probability 1/2. Here a line community is a community that consists of a line of 5 vertices. The two vertices at the ends of the line have inter-community degree one, the other vertices have inter-community degree zero. It is possible to calculate (14.3.22) analytically in this setting. Table 14.2 compares this approximation with the exact values of $\pi_n(\lambda)$. We can see that the approximation is very close to the actual value of $\pi_n(\lambda)$, especially for n large and λ small.

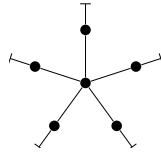


Figure 14.4: A star-shaped community

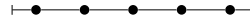


Figure 14.5: A line community of length 5

λ	$n = 10^5$		$n = 10^6$	
	$\pi_n(\lambda)$	$\pi_n(\lambda)$ appr.	$\pi_n(\lambda)$	$\pi_n(\lambda)$ appr.
-10	0,623	0,636	0,696	0,698
-1	0,741	0,741	0,747	0,747
0	0,753	0,753	0,753	0,753
1	0,764	0,764	0,758	0,758
10	0,858	0,870	0,804	0,807

Table 14.2: Values of $\pi_n(\lambda)$ for line communities and single vertex communities, and the approximation by (14.3.22).

14.4 Conclusion

In this chapter we have investigated the influence of mesoscopic community structures on critical component sizes in the hierarchical configuration model (HCM). We have considered the critical component sizes of the HCM when the inter-community connections have a finite third moment. These critical component sizes converge as $n \rightarrow \infty$ to a similar scaling limit as the critical component sizes in the CM, as long as the mesoscopic scales remain smaller than $n^{2/3}$. The critical component sizes of the HCM only depend on the sizes of the communities, and are independent of the precise community shapes. We have also obtained an implicit critical percolation window for the HCM, that depends on both the connections between communities, as well as the connections inside communities. We have found that under stricter conditions on the community sizes and the inter-community edges, the critical window can be written in an explicit form. The question whether this stricter condition is necessary to write the critical window in an explicit form remains open for further research.

The HCM can be used to model real-world networks with a community structure. Since Chapter 12 shows that many real-world networks have diverging third moments of their inter-community connections, it would be worthwhile to investigate the scaling limits of the HCM in this setting.

Bibliography

- [1] R. van der Hofstad, A. Janssen and J. S. H. van Leeuwaarden. “Critical epidemics, random graphs and Brownian motion with a parabolic drift”. In: *Adv. Appl. Probab.* 42.4 (2010), pp. 1187–1206.
- [2] A. Aharony and A. B. Harris. “Absence of self-averaging and universal fluctuations in random systems near critical points”. In: *Phys. Rev. Lett.* 77.18 (1996), pp. 3700–3703.
- [3] W. Aiello, A. Bonato, C. Cooper, J. Janssen and P. Prałat. “A Spatial Web Graph Model with Local Influence Regions”. In: *Internet Mathematics* 5.1-2 (2008), pp. 175–196.
- [4] R. Albert and A. L. Barabási. “Emergence of scaling in random networks”. In: *Science* 286.5439 (1999), pp. 509–512.
- [5] R. Albert, H. Jeong and A.-L. Barabási. “Error and attack tolerance of complex networks”. In: *Nature* 406.6794 (2000), pp. 378–382.
- [6] R. Albert, H. Jeong and A.-L. Barabási. “Internet: Diameter of the world-wide web”. In: *Nature* 401.6749 (1999), pp. 130–131.
- [7] D. Aldous. “Brownian excursions, critical random graphs and the multiplicative coalescent”. In: *Ann. Probab.* 25.2 (1997), pp. 812–854.
- [8] A. Allard, M. Á. Serrano, G. García-Pérez and M. Boguñá. “The geometric nature of weights in real complex networks”. In: *Nat. Commun.* 8 (2017), p. 14103.
- [9] U. Alon. “Network motifs: theory and experimental approaches”. In: *Nat. Rev. Genet.* 8.6 (2007), pp. 450–461.
- [10] R. Arratia and T. M. Liggett. “How likely is an i.i.d. degree sequence to be graphical?” In: *The Annals of Applied Probability* 15.1B (2005), pp. 652–670.
- [11] Y. Artzy-Randrup and L. Stone. “Generating uniformly distributed random networks”. In: *Physical Review E* 72.5 (2005).
- [12] P. Bailey, N. Craswell and D. Hawking. “Engineering a multi-purpose test collection for Web retrieval experiments”. In: *Information Processing & Management* 39.6 (2003), pp. 853–871.
- [13] F. Ball, D. Sirl and P. Trapman. “Analysis of a stochastic SIR epidemic on a random network incorporating household structure”. In: *Math. Biosci.* 224.2 (2010), pp. 53–73.

- [14] F. Ball, D. Sirl and P. Trapman. “Threshold behaviour and final outcome of an epidemic on a random network with household structure”. In: *Adv. in Appl. Probab.* 41.03 (2009), pp. 765–796.
- [15] T. Bannink, R. van der Hofstad and C. Stegehuis. “Switch chain mixing times and triangle counts in simple graphs with given degrees”. In: *Journal of Complex Networks* (2018), cny013.
- [16] A.-L. Barabási. *Network Science*. Cambridge University Press, 2016.
- [17] A. Barrat, M. Barthélemy, R. Pastor-Satorras and A. Vespignani. “The architecture of complex weighted networks”. In: *Proc. Natl. Acad. Sci. U.S.A.* 101.11 (2004), pp. 3747–3752.
- [18] A. Barrat, M. Barthélemy and A. Vespignani. *Dynamical processes on complex networks*. Cambridge University Press, 2008.
- [19] A. R. Benson, D. F. Gleich and J. Leskovec. “Higher-order organization of complex networks”. In: *Science* 353.6295 (2016), pp. 163–166.
- [20] N. Berger, C. Borgs, J. T. Chayes and A. Saberi. “Asymptotic behavior and distributional limits of preferential attachment graphs”. In: *Ann. Probab.* 42.1 (2014), pp. 1–40.
- [21] S. Bhamidi, R. van der Hofstad and G. Hooghiemstra. “Universality for first passage percolation on sparse random graphs”. In: *Ann. Probab.* 45.4 (2017), pp. 2568–2630.
- [22] S. Bhamidi, R. van der Hofstad and J. S. H. van Leeuwaarden. “Novel scaling limits for critical inhomogeneous random graphs”. In: *Ann. Probab.* 40.6 (2012), pp. 2299–2361.
- [23] S. Bhamidi, R. van der Hofstad and J. S. H. van Leeuwaarden. “Scaling limits for critical inhomogeneous random graphs with finite third moments”. In: *Electron. J. Probab.* 15 (2010), no. 54, 1682–1702.
- [24] S. Bhamidi, S. Sen and X. Wang. “Continuum limit of critical inhomogeneous random graphs”. In: *Probab. Theory Related Fields* (2016), pp. 1–77.
- [25] S. Bhamidi, N. Broutin, S. Sen and X. Wang. “Scaling limits of random graph models at criticality: Universality and the basin of attraction of the Erdős Rényi random graph”. In: arXiv:1411.3417 (2014).
- [26] S. Bhamidi, S. Dhara, R. van der Hofstad and S. Sen. “Universality for critical heavy-tailed network models: metric structure of maximal components”. In: arXiv:1703.07145 (2017).
- [27] T. Bläsius, T. Friedrich and A. Krohmer. “Cliques in hyperbolic random graphs”. In: *Algorithmica* 80.8 (2018), pp. 2324–2344.
- [28] T. Bläsius, C. Freiberger, T. Friedrich, M. Katzmann, F. Montenegro-Retana and M. Thieffry. “Efficient Shortest Paths in Scale-Free Networks with Underlying Hyperbolic Geometry”. In: arXiv:1805.03253 (2018).
- [29] V. D. Blondel, J.-L. Guillaume, R. Lambiotte and E. Lefebvre. “Fast unfolding of communities in large networks”. In: *J. Stat. Mech: Theory Exp.* 2008.10 (2008), P10008.

- [30] M. Bloznelis. "Degree and clustering coefficient in sparse random intersection graphs". In: *Ann. Appl. Probab.* 23.3 (2013), pp. 1254–1289.
- [31] M. Bloznelis, J. Jaworski and V. Kurauskas. "Assortativity and clustering of sparse random intersection graphs". In: *Electron. J. Probab.* 18.0 (2013).
- [32] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez and D.-U. Hwang. "Complex networks: Structure and dynamics". In: *Phys. Rep.* 424.4 (2006), pp. 175–308.
- [33] M. Bode, N. Fountoulakis and T. Müller. "On the largest component of a hyperbolic model of complex networks". In: *Electron. J. Combin.* 22.3 (2015), P3–24.
- [34] M. Bode, N. Fountoulakis and T. Müller. "The probability of connectivity in a hyperbolic model of complex networks". In: *Random Structures & Algorithms* 49.1 (2016), pp. 65–94.
- [35] M. Boguñá and R. Pastor-Satorras. "Class of correlated random networks with hidden variables". In: *Phys. Rev. E* 68 (3 2003), p. 036112.
- [36] M. Boguñá, R. Pastor-Satorras and A. Vespignani. "Absence of epidemic threshold in scale-free networks with degree correlations". In: *Phys. Rev. Lett.* 90 (2 2003), p. 028701.
- [37] M. Boguñá, C. Castellano and R. Pastor-Satorras. "Langevin approach for the dynamics of the contact process on annealed scale-free networks". In: *Phys. Rev. E* 79.3 (2009), p. 036110.
- [38] M. Boguñá, F. Papadopoulos and D. Krioukov. "Sustaining the Internet with hyperbolic mapping". In: *Nat. Commun.* 1.6 (2010), pp. 1–8.
- [39] M. Boguñá, R. Pastor-Satorras and A. Vespignani. "Absence of Epidemic Threshold in Scale-Free Networks with Degree Correlations". In: *Phys. Rev. Lett.* 90.2 (2003).
- [40] M. Boguñá, R. Pastor-Satorras, A. Díaz-Guilera and A. Arenas. "Models of social networks based on social distance attachment". In: *Phys. Rev. E* 70.5 (2004).
- [41] B. Bollobás. "A probabilistic proof of an asymptotic formula for the number of labelled regular graphs". In: *European J. Combin.* 1.4 (1980), pp. 311–316.
- [42] B. Bollobás, S. Janson and O. Riordan. "The phase transition in inhomogeneous random graphs". In: *Random Structures & Algorithms* 31.1 (2007), pp. 3–122.
- [43] B. Bollobás and O. Riordan. "An old approach to the giant component problem". In: *Journal of Combinatorial Theory, Series B* 113 (2015), pp. 236–260.
- [44] B. Bollobás and O. Riordan. "Mathematical Results on Scale-Free Random Graphs". In: *Handbook of Graphs and Networks*. Wiley–WCH, 2002, pp. 1–37.
- [45] B. Bollobás, O. Riordan, J. Spencer and G. Tusnády. "The degree sequence of a scale-free random graph process". In: *Random Structures Algorithms* 18.3 (2001), pp. 279–290.
- [46] S. Bonaccorsi, S. Ottaviano, F. De Pellegrini, A. Socievole and P. Van Mieghem. "Epidemic outbreaks in two-scale community networks". In: *Phys. Rev. E* 90.1 (2014), p. 012810.

- [47] M. Borassi, A. Chessa and G. Caldarelli. “Hyperbolicity measures democracy in real-world networks”. In: *Phys. Rev. E* 92.3 (2015).
- [48] P. Brach, M. Cygan, J. Łacki and P. Sankowski. “Algorithmic complexity of power law networks”. In: *Proceedings of the Twenty-seventh Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '16. Arlington, Virginia: Society for Industrial and Applied Mathematics, 2016, pp. 1306–1325.
- [49] K. Bringmann, R. Keusch and J. Lengler. “Sampling geometric inhomogeneous random graphs in linear time”. In: *arXiv:1511.00576v3* (2015).
- [50] T. Britton, M. Deijfen and A. Martin-Löf. “Generating simple random graphs with prescribed degree distribution”. In: *J. Stat. Phys.* 124.6 (2006), pp. 1377–1397.
- [51] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins and J. Wiener. “Graph structure in the Web”. In: *Computer Networks* 33.1-6 (2000), pp. 309–320.
- [52] A. D. Broido and A. Clauset. “Scale-free networks are rare”. In: *arXiv:1801.03400v1* (2018).
- [53] D. Bu, Y. Zhao, L. Cai, H. Xue, X. Zhu, H. Lu, J. Zhang, S. Sun, L. Ling, N. Zhang, et al. “Topological structure analysis of the protein–protein interaction network in budding yeast”. In: *Nucleic Acids Res.* 31.9 (2003), pp. 2443–2450.
- [54] P. G. Buckley and D. Osthus. “Popularity based random graph models leading to a scale-free degree sequence”. In: *Discrete Mathematics* 282.1-3 (2004), pp. 53–68.
- [55] D. S. Callaway, M. E. J. Newman, S. H. Strogatz and D. J. Watts. “Network robustness and fragility: Percolation on random graphs”. In: *Phys. Rev. Lett.* 85.25 (2000), p. 5468.
- [56] E. Candellero and N. Fountoulakis. “Clustering and the hyperbolic geometry of complex networks”. In: *Algorithms and Models for the Web Graph: 11th International Workshop, WAW 2014, Beijing, China, December 17-18, 2014, Proceedings*. Ed. by A. Bonato, F. C. Graham and P. Prałat. Springer International Publishing, 2014, pp. 1–12.
- [57] M. Catanzaro, M. Boguñá and R. Pastor-Satorras. “Generation of uncorrelated random scale-free networks”. In: *Phys. Rev. E* 71 (2 2005), p. 027103.
- [58] M. Catanzaro, G. Caldarelli and L. Pietronero. “Assortative model for social networks”. In: *Phys. Rev. E* 70 (3 2004), p. 037101.
- [59] D. Chakrabarti, Y. Zhan and C. Faloutsos. “R-MAT: A Recursive Model for Graph Mining”. In: *Proceedings of the 2004 SIAM International Conference on Data Mining*. Society for Industrial and Applied Mathematics, 2004, pp. 442–446.
- [60] E. Cho, S. A. Myers and J. Leskovec. “Friendship and mobility: user movement in location-based social networks”. In: *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2011, pp. 1082–1090.

- [61] F. Chung and L. Lu. "The average distances in random graphs with given expected degrees". In: *Proc. Natl. Acad. Sci. USA* 99.25 (2002), pp. 15879–15882.
- [62] F. Chung, L. Lu and V. Vu. "Spectra of random graphs with given expected degrees". In: *Proceedings of the National Academy of Sciences* 100.11 (2003), pp. 6313–6318.
- [63] F. Chung and L. Lu. "Concentration inequalities and martingale inequalities: a survey". In: *Internet Mathematics* 3.1 (2006), pp. 79–127.
- [64] F. Chung and M. Radcliffe. "On the spectra of general random graphs". In: *the electronic journal of combinatorics* 18.1 (2011), p. 215.
- [65] A. Clauset, M. E. J. Newman and C. Moore. "Finding community structure in very large networks". In: *Phys. Rev. E* 70.6 (2004).
- [66] A. Clauset, C. R. Shalizi and M. E. J. Newman. "Power-law distributions in empirical data". In: *SIAM Rev.* 51.4 (2009), pp. 661–703.
- [67] R. Cohen, K. Erez, D. ben Avraham and S. Havlin. "Breakdown of the Internet under Intentional Attack". In: *Phys. Rev. Lett.* 86.16 (2001), pp. 3682–3685.
- [68] E. Coupechoux and M. Lelarge. "How clustering affects epidemics in random networks". In: *Adv. in Appl. Probab.* 46.04 (2014), pp. 985–1008.
- [69] M. Das and J. R. Green. "Self-averaging fluctuations in the chaoticity of simple fluids". In: *Phys. Rev. Lett.* 119.11 (2017).
- [70] M. Deijfen and W. Kets. "Random intersection graphs with tunable degree distribution and clustering". In: *Probab. Engrg. Inform. Sci.* 23.04 (2009), p. 661.
- [71] S. Dhara, R. van der Hofstad, J. S. H. van Leeuwen and S. Sen. "Critical window for the configuration model: finite third moment degrees". In: *Electron. J. Probab.* 22.0 (2017).
- [72] S. Dhara, R. van der Hofstad, J. S. H. van Leeuwen and S. Sen. "Heavy-tailed configuration models at criticality". In: arXiv:1612.00650v2 (2016).
- [73] S. Dommers, R. van der Hofstad and G. Hooghiemstra. "Diameters in preferential attachment models". In: *J. Stat. Phys.* 139.1 (2010), pp. 72–107.
- [74] S. N. Dorogovtsev. "Clustering of correlated networks". In: *Phys. Rev. E* 69 (2004), p. 027104.
- [75] S. N. Dorogovtsev, A. V. Goltsev and J. F. F. Mendes. "Pseudofractal scale-free web". In: *Phys. Rev. E* 65 (6 2002), p. 066122.
- [76] S. N. Dorogovtsev, A. V. Goltsev and J. F. Mendes. "Critical phenomena in complex networks". In: *Rev. Mod. Phys.* 80.4 (2008), p. 1275.
- [77] N. Eggemann and S. D. Noble. "The clustering coefficient of a scale-free random graph". In: *Discrete Applied Mathematics* 159.10 (2011), pp. 953–965.
- [78] H. van den Esker, R. van der Hofstad and G. Hooghiemstra. "Universality for the distance in finite variance random graphs". In: *J. Stat. Phys.* 133.1 (2008), pp. 169–202.
- [79] M. Faloutsos, P. Faloutsos and C. Faloutsos. "On power-law relationships of the internet topology". In: *ACM SIGCOMM Computer Communication Review*. Vol. 29. 4. ACM. 1999, pp. 251–262.

- [80] S. Fortunato. "Community detection in graphs". In: *Phys. Rep.* 486.3 (2010), pp. 75–174.
- [81] N. Fountoulakis. "Percolation on sparse random graphs with given degree sequence". In: *Internet Mathematics* 4.4 (2007), pp. 329–356.
- [82] N. Fountoulakis, P. van der Hoorn, D. Krioukov, T. Müller and M. Schepers. "Local clustering in the hyperbolic random graph". In: *Work in progress* (2018).
- [83] N. Fountoulakis, T. Friedrich and D. Hermelin. "On the average-case complexity of parameterized clique". In: *Theoret. Comput. Sci.* 576 (2015), pp. 18–29.
- [84] O. Frank. "Moment properties of subgraph counts in stochastic graphs". In: *Ann. New York Acad. Sci.* 319.1 (1979), pp. 207–218.
- [85] T. Friedrich and A. Krohmer. "Cliques in hyperbolic random graphs". In: *INFOCOM proceedings 2015*. IEEE, 2015, pp. 1544–1552.
- [86] T. Friedrich and A. Krohmer. "Parameterized clique on inhomogeneous random graphs". In: *Discrete Appl. Math.* 184 (2015), pp. 130–138.
- [87] C. Gao and J. Lafferty. "Testing network structure using relations between small subgraph probabilities". In: *arXiv:1704.06742* (2017).
- [88] P. Gao and N. Wormald. "Uniform generation of random graphs with power-law degree sequences". In: *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*. Society for Industrial and Applied Mathematics, 2018, pp. 1741–1758.
- [89] G. García-Pérez, M. Boguñá, A. Allard and M. Á. Serrano. "The hidden hyperbolic geometry of international trade: World Trade Atlas 1870–2013". In: *Sci. Rep.* 6.1 (2016).
- [90] M. R. Garey, D. S. Johnson and M. R. Garey. *Computers and intractability: A guide to the theory of NP-completeness*. W H FREEMAN & CO, Apr. 11, 2011. 340 pp.
- [91] C. I. D. Genio, H. Kim, Z. Toroczkai and K. E. Bassler. "Efficient and exact sampling of simple graphs with given arbitrary degree sequence". In: *PLoS ONE* 5.4 (2010). Ed. by F. Rapallo, e10012.
- [92] M. Girvan and M. E. Newman. "Community structure in social and biological networks". In: *Proceedings of the National Academy of Sciences* 99.12 (2002), pp. 7821–7826.
- [93] J. P. Gleeson. "Bond percolation on a class of clustered random networks". In: *Phys. Rev. E* 80.3 (2009), p. 036107.
- [94] J. P. Gleeson. "Cascades on correlated and modular random networks". In: *Phys. Rev. E* 77.4 (2008), p. 046117.
- [95] J. P. Gleeson, S. Melnik and A. Hackett. "How clustering affects the bond percolation threshold in complex networks". In: *Phys. Rev. E* 81.6 (2010), p. 066114.
- [96] I. Gradshteyn and I. Ryzhik. *Table of Integrals, Series, and Products*. Ed. by D. Zwillinger and V. Moll. Elsevier, 2015.

- [97] C. Greenhill and M. Sfragara. "The switch Markov chain for sampling irregular graphs and digraphs". In: *Theoretical Computer Science* 719 (2018), pp. 1–20.
- [98] J. A. Grochow and M. Kellis. "Network motif discovery using subgraph enumeration and symmetry-breaking". In: *In RECOMB*. 2007, pp. 92–106.
- [99] P. Groeneboom. "Brownian motion with a parabolic drift and airy functions". In: *Probab. Theory Related Fields* 81.1 (1989), pp. 79–109.
- [100] L. Gugelmann, K. Panagiotou and U. Peter. "Random hyperbolic graphs: Degree sequence and clustering". In: *Automata, Languages, and Programming*. Ed. by A. Czumaj, K. Mehlhorn, A. Pitts and R. Wattenhofer. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 573–585.
- [101] L. Gugelmann, K. Panagiotou and U. Peter. "Random hyperbolic graphs: degree sequence and clustering". In: *ICALP proceedings 2012, Part II*. Springer, Berlin, Heidelberg. 2012, pp. 573–585.
- [102] R. Guimerà, L. Danon, A. Díaz-Guilera, F. Giralt and A. Arenas. "Self-similar community structure in a network of human interactions". In: *Phys. Rev. E* 68.6 (2003).
- [103] R. Guimera and L. A. N. Amaral. "Functional cartography of complex metabolic networks". In: *Nature* 433.7028 (2005), pp. 895–900.
- [104] S. L. Hakimi. "On realizability of a set of integers as degrees of the vertices of a linear graph. I". In: *Journal of the Society for Industrial and Applied Mathematics* 10.3 (1962), pp. 496–506.
- [105] H. Heydari and S. M. Taheri. "Distributed maximal independent set on inhomogeneous random graphs". In: *2017 2nd Conference on Swarm Intelligence and Evolutionary Computation (CSIEC)*. IEEE, 2017.
- [106] R. van der Hofstad. *Random Graphs and Complex Networks Vol. 1*. Cambridge University Press, 2017.
- [107] R. van der Hofstad. *Random Graphs and Complex Networks Vol. 2*. Available on <http://www.win.tue.nl/~rhofstad/NotesRGCNII.pdf>, 2014.
- [108] R. van der Hofstad, G. Hooghiemstra and P. Van Mieghem. "Distances in random graphs with finite variance degrees". In: *Random Structures & Algorithms* 27.1 (2005), pp. 76–123.
- [109] R. van der Hofstad, G. Hooghiemstra and D. Znamenski. "Distances in random graphs with finite mean and infinite variance degrees". In: *Electron. J. Probab.* 12.0 (2007), pp. 703–766.
- [110] R. van der Hofstad and N. Litvak. "Degree-degree dependencies in random graphs with heavy-tailed degrees". In: *Internet Mathematics* 10.3-4 (2014), pp. 287–334.
- [111] R. van der Hofstad, P. van der Hoorn, N. Litvak and C. Stegehuis. "Limit theorems for assortativity and clustering in the configuration model with scale-free degrees". In: *arXiv:1712.08097* (2017).

- [112] P. van der Hoorn and N. Litvak. "Upper bounds for number of removed edges in the erased configuration model". In: *Algorithms and Models for the Web Graph: 12th International Workshop, WAW 2015, Proceedings*. Ed. by D. F. Gleich, J. Komjáthy and N. Litvak. Cham: Springer International Publishing, 2015, pp. 54–65.
- [113] W. Huang and C. Li. "Epidemic spreading in scale-free networks with community structure". In: *J. Stat. Mech: Theory Exp.* 2007.01 (2007), P01014.
- [114] E. Jacob and P. Mörters. "Spatial preferential attachment networks: Power laws and clustering coefficients". In: *Ann. Appl. Probab.* 25.2 (2015), pp. 632–662.
- [115] S. Janson. "On percolation in random graphs with given vertex degrees". In: *Electron. J. Probab.* 14 (2009), pp. 86–118.
- [116] S. Janson. "The probability that a random multigraph is simple". In: *Combin. Probab. Comput.* 18.1-2 (2009), p. 205.
- [117] S. Janson and M. J. Luczak. "A new approach to the giant component problem". In: *Random Structures and Algorithms* 34.2 (2009), pp. 197–216.
- [118] S. Janson and M. J. Luczak. "A simple solution to the k-core problem". In: *Random Structures & Algorithms* 30.1-2 (2007), pp. 50–62.
- [119] S. Janson, T. Łuczak and I. Norros. "Large cliques in a power-law random graph". In: *J. Appl. Probab.* 47.04 (2010), pp. 1124–1135.
- [120] S. Janson and O. Riordan. "Susceptibility in inhomogeneous random graphs". In: *Electron. J. Combin.* 19.1 (2012), # P31.
- [121] A. J.E. M. Janssen and J. S. H. van Leeuwen. "Giant component sizes in scale-free networks with power-law degrees and cutoffs". In: *EPL (Europhysics Letters)* 112.6 (2015), p. 68001.
- [122] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai and A.-L. Barabási. "The large-scale organization of metabolic networks". In: *Nature* 407.6804 (2000), pp. 651–654.
- [123] A. Joseph. "The component sizes of a critical random graph with given degree sequence". In: *Ann. Appl. Probab.* 24.6 (2014), pp. 2560–2594.
- [124] M. Karonski, E. R. Scheinerman and K. B. Singer-Cohen. "On random intersection graphs: The subgraph problem". In: *Combin. Probab. Comput.* 8.1-2 (1999), 131–159.
- [125] R. M. Karp. "Reducibility among combinatorial problems". In: *Complexity of computer computations*. Springer, 1972, pp. 85–103.
- [126] B. Karrer and M. E. J. Newman. "Random graphs containing arbitrary distributions of subgraphs". In: *Phys. Rev. E* 82.6 (2010).
- [127] N. Kashtan, S. Itzkovitz, R. Milo and U. Alon. "Efficient sampling algorithm for estimating subgraph concentrations and detecting network motifs". In: *Bioinformatics* 20.11 (2004), pp. 1746–1758.
- [128] K. Klemm and V. M. Eguíluz. "Growing scale-free networks with small-world behavior". In: *Phys. Rev. E* 65 (5 2002), p. 057102.

- [129] B. Klimt and Y. Yang. "Introducing the Enron Corpus." In: *CEAS*. 2004.
- [130] D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat and M. Boguná. "Hyperbolic geometry of complex networks". In: *Phys. Rev. E* 82.3 (2010), p. 036106.
- [131] D. Krioukov, M. Kitsak, R. S. Sinkovits, D. Rideout, D. Meyer and M. Boguná. "Network cosmology". In: *Sci. Rep.* 2 (2012), p. 793.
- [132] A. Krot and L. Ostroumova Prokhorenkova. "Local clustering coefficient in generalized preferential attachment models". In: *Algorithms and Models for the Web Graph: 12th International Workshop, WAW 2015, Eindhoven, The Netherlands, December 10-11, 2015, Proceedings*. Ed. by D. F. Gleich, J. Komjáthy and N. Litvak. Cham: Springer International Publishing, 2015, pp. 15–28.
- [133] J. Kunegis. "KONECT: The Koblenz Network Collection". In: *Proceedings of the 22Nd International Conference on World Wide Web. WWW '13 Companion*. Rio de Janeiro, Brazil: ACM, 2013, pp. 1343–1350.
- [134] F. S. Labini, N. L. Vasilyev, L. Pietronero and Y. V. Baryshev. "Absence of self-averaging and of homogeneity in the large-scale galaxy distribution". In: *EPL (Europhysics Letters)* 86.4 (2009), p. 49001.
- [135] A. Lancichinetti and S. Fortunato. "Community detection algorithms: a comparative analysis". In: *Phys. Rev. E* 80.5 (2009), p. 056117.
- [136] A. Lancichinetti, S. Fortunato and F. Radicchi. "Benchmark graphs for testing community detection algorithms". In: *Phys. Rev. E* 78.4 (2008), p. 046110.
- [137] J. Leskovec and A. Krevl. *SNAP Datasets: Stanford Large Network Dataset Collection*. <http://snap.stanford.edu/data>. Date of access: 14/03/2017. 2014.
- [138] J. Leskovec, K. J. Lang and M. Mahoney. "Empirical comparison of algorithms for network community detection". In: *Proceedings of the 19th international conference on World wide web*. ACM. 2010, pp. 631–640.
- [139] J. Leskovec, K. J. Lang, A. Dasgupta and M. W. Mahoney. "Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters". In: *Internet Mathematics* 6.1 (2009), pp. 29–123.
- [140] J. Leskovec. *Dynamics of large networks*. ProQuest, 2008.
- [141] N. Litvak and R. van der Hofstad. "Uncovering disassortativity in large scale-free networks". In: *Phys. Rev. E* 87.2 (2013).
- [142] Z. Liu and B. Hu. "Epidemic spreading in community networks". In: *Europhysics Letters (EPL)* 72.2 (2005), pp. 315–321.
- [143] S. Lozano, A. Arenas and A. Sánchez. "Mesoscopic structure conditions the emergence of cooperation on social networks". In: *PLoS One* 3.4 (2008), e1892.
- [144] D. Marcus and Y. Shavitt. "Rage – a rapid graphlet enumerator for large networks". In: *Comput. Networks* 56.2 (2012), pp. 810–819.
- [145] S. Maslov and K. Sneppen. "Specificity and stability in topology of protein networks". In: *Science* 296.5569 (2002), pp. 910–913.

- [146] S. Maslov, K. Sneppen and A. Zaliznyak. "Detection of topological patterns in complex networks: correlation profile of the internet". In: *Phys. A* 333 (2004), pp. 529–540.
- [147] C. Matias, S. Schbath, E. Birmelé, J.-J. Daudin and S. Robin. "Network motifs: mean and variance for the count". In: *Revstat* 4.1 (2006), pp. 31–51.
- [148] M. Mayo, A. Abdelzaher and P. Ghosh. "Long-range degree correlations in complex networks". In: *Computational Social Networks* 2.1 (2015), p. 4.
- [149] P. V. Mieghem, H. Wang, X. Ge, S. Tang and F. A. Kuipers. "Influence of assortativity and degree-preserving rewiring on the spectra of networks". In: *The European Physical Journal B* 76.4 (2010), pp. 643–652.
- [150] G. Miller and C. Fellbaum. *Wordnet: An electronic lexical database*. 1998.
- [151] R. Milo, N. Kashtan, S. Itzkovitz, M. E. J. Newman and U. Alon. "On the uniform generation of random graphs with prescribed degree sequences". In: *arXiv preprint cond-mat/0312028* (2003).
- [152] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii and U. Alon. "Network motifs: Simple building blocks of complex networks". In: *Science* 298.5594 (2002), pp. 824–827.
- [153] R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer and U. Alon. "Superfamilies of evolved and designed networks". In: *Science* 303.5663 (2004), pp. 1538–1542.
- [154] M. Molloy and B. Reed. "A critical point for random graphs with a given degree sequence". In: *Random Structures & Algorithms* 6.2-3 (1995), pp. 161–180.
- [155] M. Molloy and B. Reed. "The size of the giant component of a random graph with a given degree sequence". In: *Comb. Probab. Comput.* 7.03 (1998), pp. 295–305.
- [156] A. Nachmias and Y. Peres. "Critical percolation on random regular graphs". In: *Random Structures & Algorithms* 36.2 (2010), pp. 111–148.
- [157] M. E. J. Newman. "Random graphs with clustering". In: *Phys. Rev. Lett.* 103.5 (2009), p. 058701.
- [158] M. E. J. Newman, S. Forrest and J. Balthrop. "Email networks and the spread of computer viruses". In: *Phys. Rev. E* 66.3 (2002).
- [159] M. E. J. Newman and J. Park. "Why social networks are different from other types of networks". In: *Phys. Rev. E* 68.3 (2003).
- [160] M. E. J. Newman. "Assortative mixing in networks". In: *Phys. Rev. Lett.* 89.20 (2002), p. 208701.
- [161] M. E. J. Newman. "Mixing patterns in networks". In: *Phys. Rev. E* 67.2 (2003), p. 026126.
- [162] M. E. J. Newman. *Networks: An introduction*. Oxford University Press, 2010.
- [163] M. E. J. Newman. "The structure and function of complex networks". In: *SIAM Rev.* 45.2 (2003), pp. 167–256.

- [164] M. E. J. Newman, S. H. Strogatz and D. J. Watts. "Random graphs with arbitrary degree distributions and their applications". In: *Phys. Rev. E* 64.2 (2001), p. 026118.
- [165] X. Niu, X. Sun, H. Wang, S. Rong, G. Qi and Y. Yu. "Zhishi.me - Weaving Chinese linking open data". In: *The Semantic Web – ISWC 2011*. Springer Nature, 2011, pp. 205–220.
- [166] I. Norros and H. Reittu. "On a conditionally Poissonian graph process". In: *Adv. Appl. Probab.* 38.01 (2006), pp. 59–75.
- [167] S. Omid, F. Schreiber and A. Masoudi-Nejad. "MODA: An efficient algorithm for network motif discovery in biological networks". In: *Genes & Genetic Systems* 84.5 (2009), pp. 385–395.
- [168] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész and A.-L. Barabási. "Structure and tie strengths in mobile communication networks". In: *Proceedings of the National Academy of Sciences* 104.18 (2007), pp. 7332–7336.
- [169] J.-P. Onnela, J. Saramäki, J. Kertész and K. Kaski. "Intensity and coherence of motifs in weighted complex networks". In: *Phys. Rev. E* 71 (6 2005), p. 065103.
- [170] C. Orsini, M. M. Dankulov, P. Colomer-de Simón, A. Jamakovic, P. Mahadevan, A. Vahdat, K. E. Bassler, Z. Toroczkai, M. Boguñá, G. Caldarelli, et al. "Quantifying randomness in real networks". In: *Nat. Commun.* 6 (2015).
- [171] M. Ostilli. "Fluctuation analysis in complex networks modeled by hidden-variable models: Necessity of a large cutoff in hidden-variable models". In: *Phys. Rev. E* 89 (2 2014), p. 022807.
- [172] L. Ostroumova Prokhorenkova and E. Samosvat. "Global clustering coefficient in scale-free networks". In: *Algorithms and Models for the Web Graph: 11th International Workshop, WAW 2014, Beijing, China, December 17-18, 2014, Proceedings*. Ed. by A. Bonato, F. C. Graham and P. Prałat. Cham: Springer International Publishing, 2014, pp. 47–58.
- [173] G. Palla, I. Derényi, I. Farkas and T. Vicsek. "Uncovering the overlapping community structure of complex networks in nature and society". In: *Nature* 435.7043 (2005), pp. 814–818.
- [174] J. Park and M. E. J. Newman. "Statistical mechanics of networks". In: *Phys. Rev. E* 70 (6 2004), p. 066117.
- [175] J. Park and M. E. J. Newman. "Origin of degree correlations in the Internet and other networks". In: *Phys. Rev. E* 68.2 (2003), p. 026112.
- [176] R. Pastor-Satorras, A. Vázquez and A. Vespignani. "Dynamical and correlation properties of the internet". In: *Phys. Rev. Lett.* 87 (25 2001), p. 258701.
- [177] R. Pastor-Satorras and A. Vespignani. "Epidemic spreading in scale-free networks". In: *Phys. Rev. Lett.* 86.14 (2001), p. 3200.
- [178] R. Pastor-Satorras, C. Castellano, P. Van Mieghem and A. Vespignani. "Epidemic processes in complex networks". In: *Rev. Mod. Phys.* 87 (3 2015), pp. 925–979.

- [179] L. A. Pastur and M. V. Shcherbina. "Absence of self-averaging of the order parameter in the Sherrington-Kirkpatrick model". In: *J. Statist. Phys.* 62.1-2 (1991), pp. 1–19.
- [180] F. Picard, J.-J. Daudin, M. Koskas, S. Schbath and S. Robin. "Assessing the exceptionality of network motifs". In: *J. Comput. Biol.* 15.1 (2008), pp. 1–20.
- [181] M. A. Porter, J.-P. Onnela and P. J. Mucha. "Communities in networks". In: *Notices of the AMS* 56.9 (2009), pp. 1082–1097.
- [182] V. M. Preciado and M. A. Rahimian. "Moment-Based Spectral Analysis of Random Graphs with Given Expected Degrees". In: *IEEE Transactions on Network Science and Engineering* 4.4 (2017), pp. 215–228.
- [183] L. O. Prokhorenkova. "General results on preferential attachment and clustering coefficient". In: *Optimization Letters* 11.2 (2016), pp. 279–298.
- [184] L. O. Prokhorenkova and A. Krot. "Local clustering coefficients in preferential attachment models". In: *Doklady Mathematics* 94.3 (2016), pp. 623–626.
- [185] L. O. Prokhorenkova, A. Ryabchenko and E. Samosvat. "Generalized Preferential Attachment: Tunable Power-Law Degree Distribution and Clustering Coefficient". In: *Algorithms and Models for the Web Graph*. Ed. by A. Bonato, M. Mitzenmacher and P. Prałat. Springer International Publishing, 2013, pp. 185–202.
- [186] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto and D. Parisi. "Defining and identifying communities in networks". In: *Proceedings of the National Academy of Sciences* 101.9 (2004), pp. 2658–2663.
- [187] E. Ravasz and A.-L. Barabási. "Hierarchical organization in complex networks". In: *Phys. Rev. E* 67 (2 2003), p. 026112.
- [188] S. Redner. "How popular is your paper? An empirical study of the citation distribution". In: *The European Physical Journal B* 4.2 (1998), pp. 131–134.
- [189] O. Riordan. "The phase transition in the configuration model". In: *Comb. Probab. Comput.* 21.1-2 (2012), pp. 265–299.
- [190] M. Rosvall and C. T. Bergstrom. "Maps of random walks on complex networks reveal community structure". In: *Proceedings of the National Academy of Sciences* 105.4th (2008), pp. 1118–1123.
- [191] H. L. Royden and P. Fitzpatrick. *Real Analysis*. 4th ed. Prentice Hall, 2010.
- [192] L. Russo. "On the critical percolation probabilities". In: *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 56.2 (1981), pp. 229–237.
- [193] P. Sah, L. O. Singh, A. Clauset and S. Bansal. "Exploring community structure in biological networks with random graphs". In: *BMC Bioinf.* 15.1 (2014), p. 220.
- [194] M. Salathé and J. H. Jones. "Dynamics and control of diseases in networks with community structure". In: *PLoS Comput. Biol.* 6.4 (2010), e1000736.
- [195] F. Schreiber and H. Schwobbermeyer. "MAVisto: a tool for the exploration of network motifs". In: *Bioinformatics* 21.17 (2005), pp. 3572–3574.

- [196] M. Á. Serrano and M. Boguñá. "Clustering in complex networks. I. General formalism". In: *Phys. Rev. E* 74 (5 2006), p. 056114.
- [197] M. Á. Serrano and M. Boguñá. "Clustering in complex networks. II. Percolation properties". In: *Phys. Rev. E* 74.5 (2006), p. 056115.
- [198] M. Á. Serrano and M. Boguñá. "Percolation and epidemic thresholds in clustered networks". In: *Phys. Rev. Lett.* 97.8 (2006), p. 088701.
- [199] M. Á. Serrano, M. Boguñá, R. Pastor-Satorras and A. Vespignani. "Correlations in complex networks". In: *Large scale structure and dynamics of complex networks: From information technology to finance and natural sciences* (2007), pp. 35–66.
- [200] S. S. Shen-Orr, R. Milo, S. Mangan and U. Alon. "Network motifs in the transcriptional regulation network of *Escherichia coli*". In: *Nat. Genet.* 31.1 (2002), pp. 64–68.
- [201] P. Colomer-de Simon and M. Boguñá. "Clustering of random scale-free networks". In: *Phys. Rev. E* 86 (2 2012), p. 026120.
- [202] P. Colomer-de Simón, M. Á. Serrano, M. G. Beiró, J. I. Alvarez-Hamelin and M. Boguñá. "Deciphering the global organization of clustering in real complex networks". In: *Sci. Rep.* 3 (2013), p. 2517.
- [203] T. Squartini and D. Garlaschelli. "Analytical maximum-likelihood method to detect patterns in real networks". In: *New J. Phys.* 13.8 (2011), p. 083001.
- [204] G. Szabó, M. Alava and J. Kertész. "Structural transitions in scale-free networks". In: *Phys. Rev. E* 67 (5 2003), p. 056102.
- [205] J. Szymanski. "Concentration of vertex degrees in a scale-free random graph process". In: *Random Structures Algorithms* 26 (2005), pp. 224–236.
- [206] P. Trapman. "On analytical approaches to epidemics on networks". In: *Theor. Popul. Biol.* 71.2 (2007), pp. 160–173.
- [207] C. E. Tsourakakis. "Fast Counting of Triangles in Large Real Networks without Counting: Algorithms and Laws". In: *2008 Eighth IEEE International Conference on Data Mining*. IEEE, 2008.
- [208] C. E. Tsourakakis, J. Pachocki and M. Mitzenmacher. "Scalable motif-aware graph clustering". In: *Proceedings of the 26th International Conference on World Wide Web*. WWW '17. Perth, Australia, 2017, pp. 1451–1460.
- [209] T. S. Turova. "Diffusion approximation for the components in critical inhomogeneous random graphs of rank 1." In: *Random Structures & Algorithms* 43.4 (2013), pp. 486–539.
- [210] A. Vázquez. "Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations". In: *Phys. Rev. E* 67 (5 2003), p. 056104.
- [211] A. Vázquez and Y. Moreno. "Resilience to damage of graphs with degree correlations". In: *Phys. Rev. E* 67.1 (2003).
- [212] A. Vázquez, R. Pastor-Satorras and A. Vespignani. "Large-scale topological and dynamical properties of the Internet". In: *Phys. Rev. E* 65 (6 2002), p. 066130.

- [213] F. Viger and M. Latapy. "Efficient and simple generation of random simple connected graphs with prescribed degree sequence". In: *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2005, pp. 440–449.
- [214] B. Viswanath, A. Mislove, M. Cha and K. P. Gummadi. "On the evolution of user interaction in facebook". In: *Proceedings of the 2nd ACM workshop on Online social networks*. ACM, 2009, pp. 37–42.
- [215] L. Weng, F. Menczer and Y.-Y. Ahn. "Virality prediction and community structure in social networks". In: *Sci. Rep.* 3 (2013).
- [216] S. Wernicke and F. Rasche. "FANMOD: a tool for fast network motif detection". In: *Bioinformatics* 22.9 (2006), pp. 1152–1153.
- [217] W. Whitt. *Stochastic-Process Limits*. Springer, New York, Apr. 11, 2006.
- [218] V. V. Williams, J. R. Wang, R. Williams and H. Yu. "Finding four-node subgraphs in triangle time". In: *Proceedings of the Twenty-sixth Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '15. San Diego, California: Society for Industrial and Applied Mathematics, 2015, pp. 1671–1680.
- [219] S. Wiseman and E. Domany. "Finite-size scaling and lack of self-averaging in critical disordered systems". In: *Phys. Rev. Lett.* 81.1 (1998), pp. 22–25.
- [220] S. Wiseman and E. Domany. "Lack of self-averaging in critical disordered systems". In: *Phys. Rev. E* 52.4 (1995), pp. 3469–3484.
- [221] X. Wu and Z. Liu. "How community structure influences epidemic spread in social networks". In: *Phys. A* 387.2 (2008), pp. 623–630.
- [222] X.-Z. Wu, P. G. Fennell, A. G. Percus and K. Lerman. "Degree Correlations Amplify the Growth of Cascades in Networks". In: *arxiv.1807.05472* (2018).
- [223] S. Wuchty, Z. N. Oltvai and A.-L. Barabási. "Evolutionary conservation of motif constituents in the yeast protein interaction network". In: *Nat. Genet.* 35.2 (2003), pp. 176–179.
- [224] G. Yan, Z.-Q. Fu, J. Ren and W.-X. Wang. "Collective synchronization induced by epidemic dynamics on complex networks with communities". In: *Phys. Rev. E* 75.1 (2007), p. 016108.
- [225] J. Yang and J. Leskovec. "Defining and evaluating network communities based on ground-truth". In: *Knowledge and Information Systems* 42.1 (2015), pp. 181–213.
- [226] D. Yao, P. van der Hoorn and N. Litvak. "Average nearest neighbor degrees in scale-free networks". In: *Internet Mathematics* (2018).
- [227] H. Yin, A. R. Benson and J. Leskovec. "Higher-order clustering in networks". In: *Phys. Rev. E* 97.5 (2018).

Summary

A network is a collection of vertices, connected in pairs by edges. Many objects of interest can be thought of as networks. Examples include social networks, the Internet, biological networks, the brain or communication networks. Even though these networks are very different in application, their connectivity patterns often share several universal properties. For example, most networks contain communities: groups of densely connected vertices. In social networks these communities may correspond to groups of friends or groups of people with similar interests, but many other types of networks also contain community structures. Another frequently observed network property is that two neighbors of a vertex are likely to be connected as well. In a social network for example, this means that two of your friends are likely to know each other as well.

Networks are typically modeled with random graphs: mathematical models generating large networks that may serve as null models for real-world networks. Whereas a real-world network often consists of just one network observation, random graphs are able to generate many network samples, allowing for statistical analyses. Furthermore, properties of these random graph models can often be analyzed mathematically. For this reason, random graph models are used to study network properties. In this thesis, we study several observed properties of real-world networks using random graph models, aiming to understand the similarities and the differences between random graph models and real-world networks.

We first investigate degree correlations, investigating whether high-degree vertices typically connect to other high-degree vertices, or to lower-degree vertices. Specifically, we study the average degree of a neighbor of a vertex of degree k , $a(k)$. Chapter 2 shows that $a(k)$ also decays in k in several random graph models, as has also been observed in many real-world networks, indicating that high degree vertices tend to connect to lower-degree vertices.

The second network property we analyze describes the relationships between neighbors of a vertex. In many networks, two neighbors of a vertex are likely to be connected as well. This can be quantified in terms of the clustering spectrum, measuring the probability that two neighbors of a vertex of degree k are connected to one another. In many real-world networks the clustering spectrum decays in k . This indicates for example that two random friends of a popular person are less likely to know each other than two random friends of a less popular person. In Chapters 3-6 we develop methods to analyze the clustering spectrum for different random graph models, and show that all these models display a decaying clustering spectrum.

The clustering spectrum of a network describes the presence of triangles in networks. However, other small subgraphs, such as squares or larger complete graphs can also provide relevant structural information. In Chapters 7-9 we identify the degrees of the vertices where subgraphs are most likely to be present for several random graph models. These optimal subgraph structures resolve the trade off that on the one hand, high degree vertices take part in many subgraphs, but on the other hand, such vertices are rare. Investigating these optimal subgraph structures enables us to count the number of subgraphs and to analyze their fluctuations. In Chapter 10 we use these optimal structures to design an algorithm that efficiently searches for a subgraph in a large random graph.

The third part of this thesis studies community structures, groups of densely connected vertices. In Chapter 11 we define a mathematically tractable random graph model that includes communities. We show that applying this model to real-world network data gives new insights into the community structures of real-world networks in Chapter 12. We then investigate how epidemic processes spread across the random graph model in Chapters 13 and 14. By comparing the spread of an epidemic process on the model with community structures to the behavior of the same process on traditional random graph models without community structures, we find how community structure influences the spread of epidemics. Interestingly, community structures speed up the spread of an epidemic in some networks, but slow it down in other networks, depending on the exact community shapes. This illustrates the importance of using community structures in network models.

About the author

Clara Stegehuis was born on May 24 1991 in Amersfoort, The Netherlands. She completed her secondary education at Corderius College in Amersfoort in 2009, and then started her studies in applied mathematics at Twente University. After obtaining her bachelor's degree in 2012, she continued her studies by pursuing a master's degree in applied mathematics, with the specialization Stochastic Operations Research. She obtained her master's degree in 2014.

Clara started her PhD project at Eindhoven University of Technology in February 2015 under the supervision of Remco van der Hofstad and Johan van Leeuwen. Her research focuses on structures in large networks. The results of this research are presented in this dissertation and provided the basis for several scientific publications. During her PhD she spent four months at the INRIA - Microsoft Research Centre in Paris. Beside her research, Clara is one of the 'Faces of Science', PhD students elected by the Royal Netherlands Academy of Arts and Sciences (KNAW) to promote science among high school students through blogs and public appearances.

