PHD

Evolving random graphs in random environment

Lodewijks, Bas

Award date:
2021

Awarding institution:
University of Bath

Link to publication

Evolving random graphs in random environment

submitted by

Bas Lodewijks

for the degree of Doctor of Philosophy

of the

University of Bath

Department of Mathematical Sciences

July 2021

Declaration of any previous Submission of the Work

The material presented here for examination for the award of a higher degree by research has not been incorporated into a submission for another degree.

. . . .

Bas Lodewijks

Declaration of Authorship

I am the author of this thesis, and the work described therein was carried out by myself personally, with the exception of Chapters 2 to 4, which contain research articles that originated from collaboration with my supervisor Marcel Ortgiese and with Laura Eslava.

. . . .

Bas Lodewijks

Abstract

This thesis is concerned with two classes of evolving random graph models in random environment: preferential attachment models with additive fitness, as originally defined in [51], and weighted recursive graphs, as defined in [71]. These models are generalisations of affine preferential attachment models and random recursive trees, respectively, and the random environment represents the inhomogeneity naturally present in real-world networks. In this thesis we study the properties of these models to understand the effect of the random environment on the evolution of the graph, and we indicate how and why the behaviour of the models in random environment differs from the classical models. In particular, we focus on the behaviour of the degree distribution and the maximum degree of these models.

For the preferential attachment model with additive fitness we consider a heavy-tailed fitness distribution and observe a phase transition in the tail exponent of the fitness distribution with respect to the behaviour of the degree distribution and maximum degree. When the fitness distribution has a light tail, we observe behaviour similar to the classical models in the sense that one of the old vertices attains the maximum degree irrespective of fitness, whereas significantly different behaviour is observed for sufficiently heavy-tailed fitness distributions, in which case the maximum degree vertex has to satisfy the right balance between fitness and age.

For the weighted recursive graph model we consider a wide range of vertex-weight distributions for which different behaviour can be observed. For distributions with unbounded support we observe that the maximum degree vertex again has to satisfy the right balance between a high vertex-weight and age. For distributions with bounded support we observe behaviour similar to the random recursive tree, at least to first order. Higher-order corrections of the maximum degree are highly dependent on the underlying vertex-weight distribution and here the behaviour can again differ significantly from what is observed for the random recursive tree.

Acknowledgements

This thesis forms the apotheosis of almost four years of work and research carried out at the Department of Mathematical Sciences of the University of Bath. Though presented as my original work, this thesis would not have come to fruition had it not been for the support of many people, be it professional or personal, whom I hope to all thank here.

First of all, I would like to thank my supervisor Marcel Ortgiese for providing me with the opportunity to come to Bath to pursue a PhD and work on the fascinating topics presented in this thesis. Your guidance and support were crucial to my development over these four years, especially during the many times that I thought there was no way to overcome the mathematical issues and problems we were facing. Your unwavering optimism especially helped me to stay motivated and focussed whilst working from home due to COVID. Thank you so much for everything, I have thoroughly enjoyed working with you these last four years and hope to continue to do so in the future.

I also want to thank Laura Eslava. Just less than a year ago I approached you to ask if you would be interested to collaborate on the topic of weighted recursive trees. Though I was hopeful, I did not expect you to respond so enthusiastically and I am grateful that you decided to work with me. You have introduced me to a lot of new mathematics that I was not familiar with and it has been great to collaborate with you.

To my doctoral examiners, Cécile Mailler and Markus Heydenreich, I would like to express my gratitude for taking the time to carefully read this thesis, for the interesting discussions we had during the Viva and for the feedback and advice they have provided.

I would like to thank all the (PhD-)students that I have had the fortune of meeting as well. Being part of such an international, diverse and vibrant community of people has been one of the highlights of my time in Bath. Most importantly, I would like to thank Martin, Isaac and Andrea, with whom I have shared an office and spent many an hour discussing, procrastinating, and enjoying life. Prior to COVID, when going to the office was all the craze, you made work all that more enjoyable.

To carry out a PhD successfully, staying sane surely is key (or at least it was for me). Luckily, I had ample opportunity in Bath to blow off steam, clear my mind and think about anything but mathematics whenever I needed to. Joining the university's Wine Society was one of the best decisions I have made since I started my PhD and I want to thank all the amazing people I have met and worked with in this fantastic society. A special thanks goes out to the Vintage Club, who have welcomed me into their circle with open arms and whose weekly (online) wine tastings during the lockdown period made it feel as if things were as normal when in fact life had pretty much come to a halt. Additionally, I want to thank Aitor for the many hours we have spent making music together during my first two years at Bath. You are an amazing pianist and have been an incredible source of inspiration. I hope we can one day continue our collaboration.

The COVID-induced reclusion that life underwent during the last one and a half years was unexpected and unfortunate to say the least. Still, with you by my side, Alisha, even the darkest of days shine bright. I am endlessly grateful for the love you have given me. All that I am is yours and I can't wait to see what the future holds for us.

Lastly, I want to thank my mother, father and brother. This PhD, and everything else I have ever achieved, pales in comparison to the infinite support, wisdom and love I have received from you. Nothing I do will ever make up for that and in the end, I can only hope that I have made you proud.

Bath, August 27, 2021

Contents

Chapter 1

Introduction

1.1 Models for real-world networks

In many areas of science, the objects or systems of interest consist of a large number of individual parts or components that are linked in a particular way. Often, beyond understanding what the individual components are, what they behave like and how individual components are linked, the structure and patterns of the connections can provide crucial insight into the behaviour of the system as a whole as well. As an example, the Internet consists of computers and other devices linked by cables, and the structure of the Internet as a whole has implications on the routes that information and data can take and hence on the efficiency of the communication of information and data.

The structure of the connections between components of a system can be represented as a network, or a graph in mathematical terminology. A graph G consists of a set of vertices $V(G)$ and a set of edges $E(G) \subseteq \{\{u, v\} | u, v \in V(G)\}$. The vertices represent the components of the system (e.g. people in a social network) and the edges represent the connections between nodes/agents in the networks (e.g. people who are friends). A graph G is directed when edges have an orientation (a hyperlink in the World Wide Web points from one webpage to another) or undirected when edges have no such orientation (a friendship between two people in a social network).

According to Newman [116], real-world networks can be roughly divided into the following four classes:

(1) Technological networks: Physical infrastructure networks that form the backbone of modern technological societies, such as the Internet, telecommunication networks and power grids.

(2) Information networks: Networks consisting of items of data linked together in a certain way, such as the World Wide Web and citation networks.

(3) Social networks: Networks in which the nodes represent people and the links represent some kind of relation between them, such as (online) friendship networks and collaboration networks.

(4) Biological networks: Networks observed in biological systems, such as the brain and protein interaction networks.

In the last decades there has been an increasing interest in complex networks and their behaviour from a large numbers of scientific areas. The growth in computational power and availability of high-quality data have enabled the improvement of empirical studies of complex networks, which in its turn re-incentivised the study of complex networks from a theoretical perspective. Network research aims to describe the properties real-world networks exhibit as well as understand why these properties are exhibited.

As many real-world complex networks are large and highly irregular, it is not possible to understand them from a global perspective, even with the computational possibilities available today. Rather, researchers aim to describe and understand these networks from a local perspective, using local properties and describing local rules to govern the connectivity among vertices. These rules often are probabilistic due to the unpredictability of how connections are formed in real-world networks and to account for the complexity of these networks. Moreover, it allows for an understanding of the *macroscopic* behaviour of the system that arises from the *microscopic* behaviour. Probability theory can be an effective and useful tool to understand complexity and to derive macroscopic behaviour from probabilistic microscopic rules. As a result, random graphs are a natural way of modelling and understanding real-world complex networks.

In this thesis, we present the findings of our research related to two particular classes of random graph models: preferential attachment models with additive fitness, and weighted recursive graphs. These models are generalisations and extensions of well-known existing random graph models, and our aim is to show why these models show different and more rich behaviour compared to the existing models and how exactly this behaviour differs. We first summarise the existing literature in this introduction to obtain a clear understanding of real-world networks and present an overview of the state-of-the-art models, to then provide the connection to the models which we focus on in this thesis.

1.2 Universal properties of real-world networks

Though probability theory and random graph models can aid in the understanding of complex networks, it is not a priori apparent what such models should look like. The empirical study of real-world networks provides an insight in the properties such models should exhibit in order to reflect the structures and patterns observed in the real world. This kind of research has sparked new interest in models for complex networks, especially since it turned out many of the 'classical' models in random graph theory do not exhibit the properties observed in many real-world networks. Further on, when discussing several random graph models, we focus on whether these properties are indeed exhibited by these models.

Perhaps somewhat surprisingly, it turns out that certain properties can be observed in a wide range of complex networks describing systems in very different contexts. Price [123] argues that the frequent observation of these properties in such a wide range of contexts can point towards universal mechanisms that govern the formation of such complex networks, which offers a further incentive to understand what these properties are and why and how they arise. We discuss a couple of these well-known properties here.

Scale-free property

The degree of a vertex $v \in V(G)$ is defined as the number of edges incident to v, or, more formally, $|\{u \in V(G) : \{v, u\} \in E(G)\}|$. In a directed graph G, the degree is defined as the sum of the in-degree and out-degree, which are the number of edges oriented towards v and the number of edges oriented away from v, respectively.

Many real-world networks are said to exhibit the scale-free property, which is a way of saying that their degree distribution follows a power-law distribution. That is, the proportion p_k of vertices with degree k scales as a regularly-varying function with a negative exponent $-\tau$, where τ is coined the power-law exponent. Hence,

$$p_k = \ell(k)k^{-\tau} \quad \text{or} \quad \log p_k = \log(\ell(k)) - \tau \log k, \tag{1.2.1}$$

where $\ell : \mathbb{N} \to \mathbb{R}_+$ is a slowly-varying function, i.e. $\lim_{x\to\infty} \ell(cx)/\ell(x) = 1$ for any fixed $c > 0$. As a result, since $\lim_{k\to\infty} \ell(k)/\log k = 0$ for any slowly-varying function ℓ, p_k against k yields an asymptotically straight line on a log-log plot with slope $-\tau$. See Figure 1-1 for an example of the degree distribution in a collaboration network of condensed matter physicists.

Due to the slow decay of a power-law, a very large variability of degrees can be observed in such networks. Lacking a typical 'scale' for the degrees, such networks have been coined scale free. More formally, a scale-free network is a network that exhibits a power-law degree sequence as in (1.2.1). Such networks were first observed by Price in 1965 in a network of citations between scientific papers [124]. According to Price, the in-degree distribution showed a power-law exponent τ between 2.5 and 3, later to be determined at $\tau = 3.036$ [123]. The Faloutsos brothers were the first to investigate the degree distribution of the Internet [54], estimating that $\tau \approx 2.15 - 2.20$. Beyond these particular examples, many other scale-free networks have been studied as well, most of which were found to exhibit the scale-free property with $\tau \in (2, 3)$ (see e.g. [2, 6, 34, 134] and [116] and the references therein).

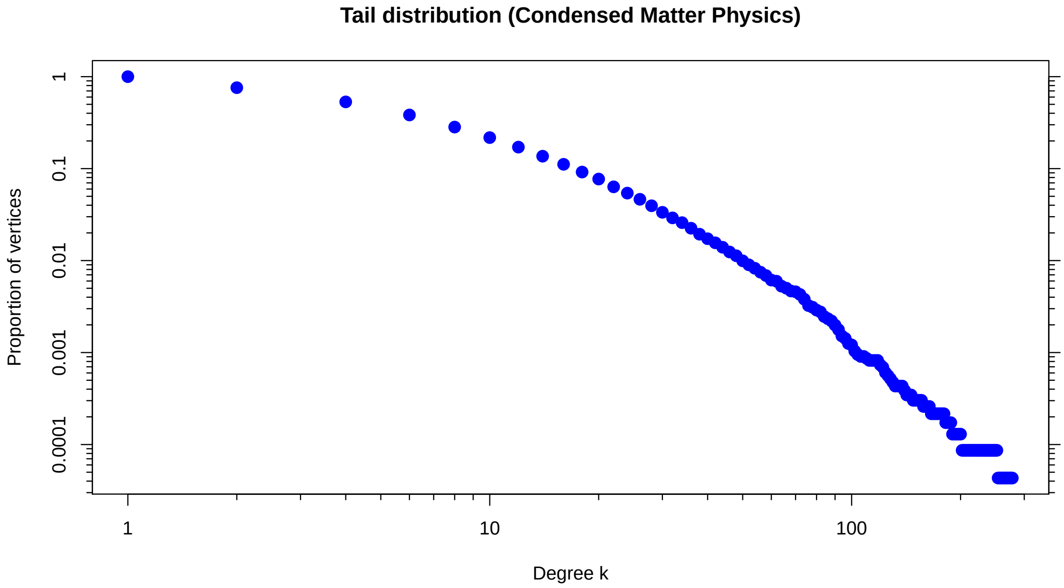

Figure 1-1: The tail distribution of a collaboration network within the field of condensed matter physics. The network consists of 23133 researchers who collaborated on papers in the period January 1993 until April 2003, and 93497 edges. An edge is created when two researchers i and j collaborate on a paper in the aforementioned time period. The plot represents the tail distribution of the number of collaborators of a researcher selected uniformly at random and is presented on log-log axes. Data from Stanford Network Analysis Project [94].

Other than a power-law degree distribution, the presence of a small but significant number of vertices with unusually high degrees can be observed in scale-free networks as well, as a result of the high variability of the degrees. These so-called hubs play an important role in the structure and performance of scale-free networks and are of particular interest in this thesis. For example, many, though not all, scale-free networks are resilient to the removal of randomly selected vertices: The average distance between nodes and the fraction of vertices in the largest connected component in the graph are robust to the random removal of vertices. When thinking of real-world applications, this is a positive result for malicious attacks against power grids, communication networks and other essential infrastructure in today's society, but a negative result for vaccinations against viruses and other diseases.

As an example, [6] discusses the resilience of the Internet to the removal of vertices. Here, simulations show that the network remains unaffected by removal of 2.5% of all vertices when selecting these vertices randomly, confirming the claim made above. However, after a targeted removal of 3% of the vertices with the highest degree, fragmentation of the network occurs. This would imply the loss of connectivity among the vertices in the network and, as a result, the loss of communication via the Internet. The observed behaviour can be explained by the fact that a large proportion of all the vertices has a small degree and hence plays an insignificant role in the connectivity of the network. The hubs with their relatively large degree are the cause of the well-connectedness of the network and hence explain the fragmentation when removed. In other examples of networks where the degree distribution p_k decays not as a power law but as a stretched exponential, this difference in fragmentation due to a different removal strategy is not observed in simulations [6]. In such networks, the degrees are much more homogeneous so that a typical vertex has a degree more comparable to the degree of the best connected vertices in the network. As a result, both the removal of randomly selected vertices and the high-degree vertices leads to a monotonic decrease of the connectedness of the network.

Critique of power-laws

The occurrence and observation of the scale-free property in many real-world networks has incited a tremendous effort to construct mathematical models that reflect and explain this behaviour. On the other hand, there has also been a critical discussion about the validity of the observations and measurements of real-world networks. In particular, the measurements that lead to power laws in the Internet have received a significant amount of criticism.

As the Internet is decentralised and distributed, it is hard to measure the Internet as a whole. Often, an algorithm known as *traceroute* is used to obtain measurements of the structure of the Internet. Traceroute sends messages from a fixed source to a fixed destination, and provides details of all routers visited on the messages' path, as well as the direction of the path. As a result, traceroute uncovers structures within the Internet.

The main critique of the traceroute algorithm, however, is that it is subject to a sampling bias. Lakhina et al. use traceroute on certain sub-graphs of an Erdős-Rényi random graph and power-law sequences can be observed from the data [93]. As we discuss in Section 1.3.1, the Erdős-Rényi random graph model does not exhibit a power-law degree distribution, implying that the observed measurements can only be due to the traceroute algorithm. Similar results were obtained by Achlioptas et al. [2] and Clauset and Moore [33] on Erdős-Rényi random graphs and random regular graphs.

Of course, there are ample examples of networks which can be observed as a whole and for which the complete degree sequence can be measured. In such cases it is harder to refute the measurements themselves. Still, it is not trivial to then conclude the degree sequence follows a power law. Estimators can be biased leading to incorrect estimates, see for example [137] which provides theoretical justification of using particular estimators for the tail exponent.

A recent criticism of scale-free networks by Broido and Clauset considers a large number of real-world network datasets in [27] and compares the degree sequences obtained from these datasets against power law, log-normal, exponential and Weibull distributions. Table 1.1 provides the conclusions from their analysis, and shows that in many cases the alternative density was accepted for the degree sequence in statistical tests, which lead Broido and Clauset to state that "scale-free networks are rare".

This paper has generated a lot of discussion, as well as criticism. Most notably, Barabási provides an in-depth review of the paper in a blog post on his website providing critique on both a conceptual and technical level [12], and in [135] Voitalov et al. consider a much more general class of power-laws rather than just the pure power law $ck^{-\gamma}$ for constants $c > 0, \gamma > 1$ considered by Broido and Clauset (see also [73] for a clear discussion on the matter). Finally, Voitalov et al. conclude: "If we relax the unrealistic requirement that degree distributions in real-world networks must be pure power laws, and allow for real-world impurity via regularly varying distributions, then upon the application of the state of-the-art methods in statistics to detect such distributions in empirical data, we find that one can definitely not call scale-free networks 'rare'."

In this thesis we shall study models which are able to not only produce power-law degree distributions, but a very large range of degree distributions including the ones listed in Table 1.1, which points to its potential use in a wide range of applications.

Alternative	$f(x) \propto$	$\mathrm{M_{PL}}$	Inconclusive	$\mathrm{M_{Alt}}$
Exponential	$\mathrm{e}^{-\lambda x}$	33%	26%	41%
Log-normal	$\frac{1}{x}\exp\{-(\log x-\mu)^2/(2\sigma^2)\}$	12%	40%	48%
Weibull	$\exp\{-(x/b)^a\}$	33%	20%	47%
Power law with cut-off	$x^{-\alpha}\mathrm{e}^{-\lambda x}$	-	44%	56%

Table 1.1: Percentage of network datasets for which the likelihood-ratio test favoured the power-law model $\mathrm{M_{PL}}$, the alternative model $\mathrm{M_{Alt}}$ with density $f(x)$, or neither. Table from [27, Table 1].

Small-world property

Another property of many real-world complex networks is that distances between vertices are, on average, very small in terms of the network size. The typical distance in a network is defined as the graph distance between two vertices that are selected uniformly at random. Let $V_1, V_2 \in V(G)$ be two vertices selected uniformly at random and let $\mathrm{d}_G(\cdot,\cdot)$ denote the graph distance metric on the graph G. The typical distance then equals $\mathrm{d}_G(V_1, V_2)$. Due to the randomness of the vertices V_1, V_2, the typical distance provides information on all distances in the network, even when G itself is a deterministic graph. Networks which have a typical distance of logarithmic order in the size of the network are known as small worlds. Networks with an even smaller typical distance, for example of double logarithmic order, are known as ultra-small worlds. An example is provided in Figure 1-2, where the distribution of the typical distance within the largest connected component in a collaboration network of condensed matter physicists is shown.

Small typical distances were first observed empirically by Milgram in [107] in 1967, who conducted a social experiment in which 60 subjects in Wichita, Kansas, U.S.A. were sent a letter and asked to deliver this letter to a person living at a specified address in Cambridge, Massachusetts. However, participants were only allowed to pass the letter to personal acquaintances, either directly or via a “friend of a friend”. Though only three letters actually reached the desired target, it only required on average six intermediaries to get the letter to the correct person. In later studies, Milgram was able to increase the success rate of the experiment significantly whilst retaining the original conclusion. This small-world phenomenon, later also coined the “six degrees of separation”, is a direct result of the underlying network topology and has been observed in many other networks as well since Milgram’s study.

In 2001, Dodds et al. performed a similar experiment, now with emails instead of physical letters. Data was recorded on 61.168 individuals resulting in 24.163 email chains. Again, the average number of intermediaries was six [45]. Backstrom et al. studied distances in the Facebook network, as well as their evolution over time, in 2012. They found the average distance to stabilise around $3-6$ with an average of 4.74, resulting in 3.74 degrees of separation on average [10].

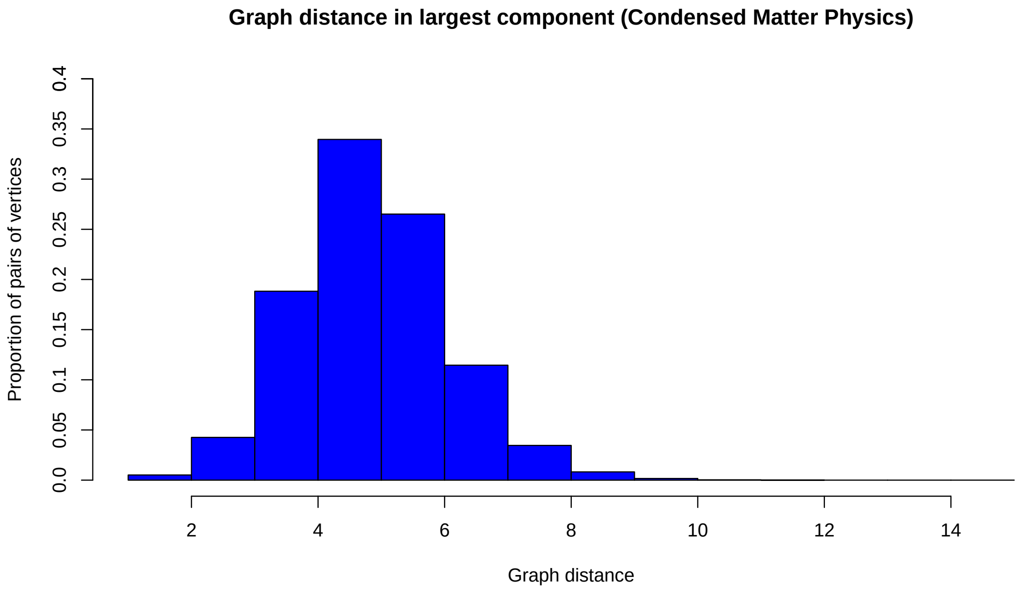

Figure 1-2: The proportion of distances between pairs of vertices in the largest connected component in a collaboration network within the field of condensed matter physics. The network consists of 23133 researchers who collaborated on papers in the period January 1993 until April 2003, and 93497 edges, with 21363 researchers and 91342 edges in the largest connected component. An edge is created when two researchers i and j collaborate on a paper in the aforementioned time period. The average distance is 5.352153 and the diameter is 15 (the SNAP finds a diameter of 14 based on samples over 1000 nodes, we compute the exact diameter here). The plot represents the probability distribution of the distance between two researchers i and j selected uniformly at random from the largest component. Data from Stanford Network Analysis Project [94].

Clustering

Clustering measures the level of transitivity present in a network. That is, when vertex v is connected to u and u is connected to w, the level of transitivity describes how likely it is for v to also be connected to w. Transitivity is often observed in many real-world networks, most notably social networks, though often not perfect but partial (not all direct neighbours of a vertex are neighbours themselves). (Partial) transitivity makes it much more likely that 'a friend of my friend is also my friend', rather than to be connected to a random individual from the population.

To quantify the level of transitivity or clustering, one can compute the clustering coefficient of a graph $G = (\{1, \ldots, n\}, E)$. First, let

$$W_G := \sum_{1 \le i,j,k \le n} \mathbb{1}_{\{(i,j),(i,k) \in E\}}, \qquad \Delta_G := \sum_{1 \le i,j,k \le n} \mathbb{1}_{\{(i,j),(i,k),(j,k) \in E\}},$$

denote the number of wedges and triangles present in the graph, respectively (note that each wedge is counted twice and each triangle counted six times). Then, the clustering coefficient is defined as

$$C_G := \frac{\Delta_G}{W_G}, \tag{1.2.2}$$

the proportion of wedges for which the closing edge to form a triangle is also present. As a result, it can be thought of the probability that two neighbours of a randomly selected vertex are neighbours as well. Alternatively, one can define the local clustering coefficient of a vertex $i \in \{1, \ldots, n\}$ as

$$C_G(i) := \frac{1}{d_i(d_i - 1)} \sum_{1 \le j,k \le n} \mathbb{1}_{\{(i,j),(i,k),(j,k) \in E\}},$$

which denotes the fraction of wedges of which i is the centre vertex for which the closing edge to form a triangle is also present, and where d_i is the degree of vertex i. Another clustering coefficient, also known as the average clustering coefficient proposed by Strogatz and Watts [138], can then be defined as

$$C'_G := \frac{1}{n} \sum_{i=1}^{n} C_G(i).$$

This is an average of all local clustering coefficients, and we note that this definition is not equivalent to the one in (1.2.2). C'_G is often larger than C_G due to the fact that vertices with low degree contribute more to the average value C'_G, so that C_G is often a better characterisation of the transitivity present in networks with many vertices with low degree.

Real-world complex networks exhibit clustering in the sense that C_G (or C'_G) is bounded away from zero for large n. Newman provides a lot of examples of social, technological, biological and information networks in which clustering is observed [118]. To mention a few: In a film actor collaboration network analysed, $C_G = 0.20, C'_G = 0.78$ and in a particular electronic power grid, $C_G = 0.10, C'_G = 0.08$. In the collaboration network for condensed matter physics, of which the degree sequence and typical distance are analysed in Figures 1-1 and 1-2, respectively, $C_G = 0.107, C'_G = 0.6334$ [94].

1.3 Random graph models

The high variability in how vertices can form connections in real-world networks has lead to the use of random graphs to model real-world networks. These models can be roughly split into two categories: static random graph models and evolving random graph models. Where the former model networks of a fixed size and provides a 'snapshot' in time of a network, the latter model networks that grow over time. Here, we discuss some of the well-known random graph models and some of their properties.

1.3.1 Static random graphs

Static random graphs are created on a set of n vertices. One is then interested in, among other things, the structural properties of these graphs as n tends to infinity. Though this provides insight into how these models behave when they are large in size, there is no direct correlation between the graph of size n and the graph of size $n+1$. As a result, the properties of such models do not always arise naturally but are more often imposed, which makes them phenomenological models. We describe some well-known examples here.

Erdős-Rényi random graph

The Erdős -Rényi random graph was proposed independently by Solomonoff and Rapaport in [131], Gilbert in [62] and by Paul Erdős and Alfréd Rényi [48], after whom it was later named. The precise model definition in the three papers varies somewhat, but all are strongly related.

In the Erdős-Rényi model we take n vertices and connect every pair of vertices $i, j \in \{1, \ldots, n\}$ independently with some fixed probability $p \in [0, 1]$. Erdős and Rényi proved several properties of this random graph model when $p = \lambda/n$ and λ is a fixed positive constant. First, it is clear that the degree of each vertex $i \in \{1, \ldots, n\}$ follows a binomial distribution with $n-1$ trials and success probability $p = \lambda/n$. As a result,

$$\lim_{n\to\infty} \mathbb{P}(\text{Degree of } i \text{ equals } k) = \mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}, \qquad k \in \mathbb{N}_0,$$

yielding a limiting Poisson degree distribution. Since $\mathrm{e}^{-\lambda}\lambda^k/k!$ is much smaller than $k^{-\tau}$ for any $\tau > 0$ when k is large, it follows that the Erdős-Rényi random graph model is *not* a scale-free network.

Furthermore, Erdős and Rényi show the existence of a phase transition of the size of the largest connected component expressed in the mean degree λ. When $\lambda < 1$, the largest connected component is of order $\log n$, whereas the largest component is of order n (and is also the unique component of this size) when $\lambda > 1$ (see [49, 50, 48]).

Though the scale-free property is not satisfied by the Erdős-Rényi random graph, it is a small world when a largest component of order n exists, i.e. when $\lambda > 1$. Then, as Newman discusses in [116], the typical distance is of order $\log n/\log\lambda$. Finally, as all edges are present independently, it is not reasonable to expect clustering in the network. Indeed, the clustering coefficient as defined in (1.2.2) is of order $1/n$ (see e.g. [72, Exercise 4.9]).

Inhomogeneous random graphs

The Erdős-Rényi model, as discussed above, is not a good model for real-world networks, among other reasons due to the fact that it is not scale free. The Erdős-Rényi graph is egalitarian in the sense that all its vertices are, in distribution, identical. It thus fails to model the heterogeneous nature of many real-world networks.

To incorporate the lacking heterogeneity in the Erdős-Rényi model, several models which we classify here as inhomogeneous random graphs have been introduced. These models aim to generalise the Erdős-Rényi random graph model and introduce more heterogeneity to the individual vertices in the graph by using vertex-weights. Each vertex i is assigned a weight w_i, and an edge (i, j) is present, independently of all other edges, with a probability proportional to the product of w_i and w_j. Examples of such models are the generalised random graph introduced by Britton, Deijfen and Martin-

Löf in [26], the Chung-Lu model [31, 32], and the Norros-Reitu model [119]. Bollobás, Janson and Riordan introduced the most general model in 2007 in [19].

The vertex-weights quantify the ability of a vertex to form edges, and they can be thought of as the expected degree of a vertex. A vertex with a large weight is more likely to form edges with other vertices than a vertex with a small weight, resulting in degrees that are not (necessarily) equal in distribution.

Though inhomogeneous random graphs do not produce graphs that are scale free and small worlds in general, such graphs can be obtained when, for example, sampling the weights from a power-law distribution, see e.g. the analysis of the model introduced by Bollobás, Janson and Riordan [19]. When the weights follow a power-law distribution, typical distances are either of logarithmic order when the power law has finite variance, yielding a small-world model, or of double logarithmic order when the power law has finite mean but infinite variance, yielding an ultra-small-world model (see [19] and [73] and the references therein). The clustering coefficient converges to zero as the graph grows infinitely large, see e.g. [20, 73, 77].

Configuration model

The configuration model, first introduced by Bollobás in [17] to study properties of random regular graphs and later generalised by Molloy and Reed in [108, 109], is a random graph model that is based on a predetermined degree distribution. One fixes a sequence of degrees $d_1, \dots, d_n$ and proceeds to draw a graph uniformly from all graphs with said degree sequence. This can be obtained by assigning each vertex $i \in \{1, \dots, n\}$ d_i half-edges and pairing half-edges uniformly at random to create edges. We note that the order of the pairing does not influence the distribution of the graph obtained, since the process of pairing half-edges is exchangeable. In such graphs it is possible for self-loops (edges from a vertex to itself) and multiple edges between a pair of vertices to arise. Van der Hofstad discusses the probability of obtaining simple graphs for the configuration model, as well as techniques to create a simple graph by deleting or switching edges in [72].

As one can choose a degree distribution for the random graph created by the configuration model, it is possible to obtain a scale-free graph simply by imposing a power-law degree distribution. Furthermore, when the degree distribution is a power law, Van der Hofstad, Hooghiemstra and Znamenski [52, 76] and Van der Hofstad, Hooghiemstra and Van Mieghem [75] study typical distances in the configuration model. They observe that such scale-free graphs are small worlds when the power law has finite variance [75], ultra-small worlds when the power law has finite mean and infinite variance [52], and that typical distances are bounded when the power law has an infinite mean [76]. The cluster coefficient vanishes in the erased configuration model (where self-loops and multiple edges are deleted), though it does not vanish in when self-loops and multiple edge are retained and $\tau \in (2, 7/3)$, as established by Van der Hofstad, Van der Hoorn, Litvak and Stegehuis [77] and Newman [116]. This is mainly due to the fact that triangles are counted ‘multiple times’ due to the high number of multiple edges.

Small-world model

Strogatz and Watts introduced the small-world model in [138] as a way to interpolate between completely deterministic and completely random graphs, as well as to take into account the underlying geometry that forms the basis for many real-world networks. One starts with a regular graph of size n in a cycle where each vertex connects to its k nearest neighbours, and rewires each edge independently with some fixed probability

p, selecting the new recipient of the edge uniformly at random. The case $p = 0$ yields a regular graph, $p = 1$ yields an (almost) completely random graph (as only one of the vertices of a rewired edge is altered, it is not considered 'fully' random), whilst $p \in (0, 1)$ allows for a mix of deterministic edges to nearest neighbours and random edges that (typically) form long-range connections or 'short-cuts'.

In the case that $p = 0$, the typical distance is rather large (of order $n/(2k)$), whereas the case $p = 1$ reminds us of the Erdős-Rényi model in which typical distances are of logarithmic order, see [116]. On the other hand, clustering can be observed when $p = 0$ (as long as $k > 2$) but not when $p = 1$. The essential quality of this model is that both clustering and small distances can be observed for a significant range of values of $p \in (0, 1)$, due to the fact that logarithmic distances can already be observed for moderate values of p close to zero, whilst clustering can be retained even for values of p close to one.

However, the degree distribution of the small-world model does not reflect degree distributions observed in real-world networks well, as it decays exponentially in the degree size [116]. Moreover, the dichotomy of short-range and long-range edges present in the model is generally speaking not an accurate representation of real-world networks, where often many edges with intermediate ranges are present. As an example, think of friendships in a social network, where certain friends live in your direct vicinity, i.e. short-range connections and some might live in a different country, i.e. long-range connections. However, one often also has connections in neighbouring towns or cities, in cities at a greater distance, and in different provinces/states within the same country, which act as the intermediate-range connections.

Hyperbolic random graphs

The structure of many real-world networks is based on an underlying geometry, think of social networks, distribution networks and transportation networks. This geometry plays a role in the connections the vertices of the network form, and a multitude of models aims to incorporate this geometry. Here, we discuss one of these models, known as the hyperbolic random graph.

Introduced by Krioukov, Papadopoulos, Kitsak, Vahdat and Boguñá in 2010, the model consists of n vertices in the disk of radius $R = 2\log(n/\nu)$, where $\nu > 0$ is a model parameter and can be interpreted as the average degree. The vertices $i \in \{1, \dots, n\}$ can be represented by their hyperbolic polar coordinates (r_i, θ_i) where the θ_i are i.i.d. uniform in $(-\pi, \pi)$ and the r_i are i.i.d. from a distribution with density $f_{\alpha,R}(r) := \alpha\sinh(\alpha r)/(\cosh(\alpha R) - 1)$, where α is again a model parameter and controls the scale-free exponent [92].

The hyperbolic random graph model has been shown to exhibit the scale-free property, the small-world property as well as non-vanishing clustering. Gugelmann, Panagiotou and Peter [66] prove the hyperbolic random graph model exhibits the scale-free property with exponent $2\alpha + 1$ and also show the clustering coefficient is bounded from below by a positive constant with high probability as n tends to infinity. The latter result has recently been improved by Fountoulakis, Van der Hoorn, Müller and Schepers in [58], where convergence in probability to the exact constant is proved. Abdullah, Bode and Fountoulakis [1] prove that the largest connected component is ultra-small when $\alpha < 1$, establishing the small-world property, and Müller and Staps [113] show that the diameter of logarithmic order.

An interesting question is how to embed real-world networks in the hyperbolic plane and how to interpret the hyperbolic geometry in real-world settings. Though the hyperbolic random graph has many of the properties one would desire in a network model, it is not yet quite clear how the hyperbolic model can be used to effectively map real-world networks, a difficult and very relevant question. The Internet has been mapped to the hyperbolic plane in [16] and serves as the best known example. Komjáthy and Lodewijks [88] provide an equivalence between hyperbolic random graphs and geometric inhomogeneous random graphs, where the position of vertices in the hyperbolic plane is translated into a Euclidean position and a vertex-weight, which could possibly allow for a more intuitive interpretation of the hyperbolic geometry. It is interesting to see more examples of such mappings as research on this topic continues.

1.3.2 Evolving random graphs

The models discussed in Section 1.3.1 provide excellent tools for understanding and studying the structural properties of networks, e.g. large and small components, degree distributions, typical distances, etcetera. Also, such models are able to serve as models for real-world networks. However, all these models exhibit particular features as they are fixed from the outset: the number of vertices and edges, degree distributions or expected degrees, the existence and size of giant components and other structural features can often be chosen as desired.

Though these phenomenological models, used to recreate desired properties, are very interesting and useful in their own right, they do not provide insight into the question *why* the properties observed empirically in real-world networks arise, for example scale-free degree distributions or small typical distances.

The models described here are generative network models, or evolving random graphs. These models grow or evolve through the sequential inclusion of vertices and edges to the network, and connections established between vertices are again governed by simple local rules. Though it is not possible to prove, it can be suggested that these rules (or similar generative mechanisms) govern the formation of real-world networks when the structures of these network models are similar to the structures observed empirically in real-world networks.

We discuss some well-known examples of evolving random graphs.

Vertex copying model

Kleinberg, Kumar, Raghavan, Rajagopalan and Tomkins introduce the vertex copying model in [87] as a model for the World Wide Web, based on the process of content-creation on this network. They aimed for their model to capture the fact that some page creators on the Web link to other websites without regard of the content and topics already present on the Web, but most creators use Web pages with content of their interest and link to such pages.

The main attribute of the model is as follows: when creating a Web page, a user finds a resource list of links regarding a particular interest and copies some (or most) of these links to be included on their own page. Newman [116] provides an alternative motivation with respect to citation networks. Here, Newman argues, researchers simply copy (parts of) bibliographies of other papers in their field into their own new papers, rather than carefully selecting papers that are actually worth citing.

The simplest form of the model is defined as follows. We assume each vertex has an identical out-degree $m \in \mathbb{N}$ and we start with some network consisting of $n_0 > m$ vertices. Then, with probability $p \in (0,1)$, a newly-added vertex connects to an existing vertex selected uniformly at random, and with probability $1-p$ it selects a random edge from a uniformly selected vertex and connects with the receiving vertex of said edge. The in-degree distribution of the model converges to a power law with exponent $1/(1-p)$ and thus the model not only exhibits the scale-free property but also provides a possible explanation of how a vertex copying mechanism could be at work within (certain) real-world networks.

Random recursive tree

The random recursive tree (RRT) is likely to be the simplest model of an evolving and growing random graph possible. A recursive tree is a labelled tree on which the labels on a path from the root to any vertex in the tree are strictly increasing (with the root labelled as 1). The random recursive tree is a tree sampled uniformly among all recursive trees of size n. We note that there are $(n-1)!$ of such increasing trees and any such a tree is drawn with probability $1/(n-1)!$. Another way to construct the RRT is via a recursive mechanism. One starts with a single vertex which forms the root of the tree and lets T_0 denote this tree. For any $n \in \mathbb{N}$, the tree T_n is then obtained from T_{n-1} by introducing a new vertex with label n and connecting it to one of the vertices in T_{n-1} uniformly at random. It is readily checked that this procedure also yields a uniform increasing tree on n vertices.

The RRT was introduced by Na and Rapoport in [114] in 1970 and has attracted a wealth of interest and many variations of such trees have been studied since. It has been used to study the spread of epidemics [110], the evolution of languages [115] and for investigating pyramid schemes and chain letters [61].

Meir and Moon prove the convergence of the empirical degree distribution to a geometric distribution in [106]. Mahmoud and Smythe [99] extend this result by showing that the number of vertices with degree $0, 1$, and 2 is asymptotically normal and Janson [82] further generalises this to the asymptotic normality of the number vertices with any fixed degree $k \in \mathbb{N}_0$. It hence follows that the RRT does not exhibit the scale-free property, most likely one of the main reasons why this model is not often considered when modelling real-world networks.

High degrees in RRTs have also gained attention, first by Szymański in [133] who prove the mean of the maximum degree scales as $\log_2 n$. Devroye and Lu [44] extend the convergence in mean to almost sure convergence. Goh and Schmutz [63] prove the distributional convergence of $\Delta_n - \log_2 n$ along suitable subsequences (Δ_n is the maximum degree in T_n) and identify possible limiting distributions. Addario-Berry and Eslava [3] establish a more precise characterisation of these limiting distributions in terms of Poisson point processes on the real line, and prove a phase-transition in the distributional limit between the number of vertices attaining the maximum degree and near-maximum degrees. Finally Eslava [53] shows the joint convergence of the rescaled (near-)maximum degrees and their depth in the tree to Poisson and normal limits, respectively. Recently, Banerjee and Bhamidi [11] obtained the convergence of the rescaled label of the maximum degree, showing it grows asymptotically as $n^{(1-2/(2\log 2))(1+o(1))}$.

Pittel studies the height of the RRT in [122]. The height is the graph distance from a vertex selected uniformly at random to the root, and grows asymptotically as $\mathrm{e}\log n$. Later, Addario-Berry and Ford established higher-order correction terms for the height in [4].

In the construction of the RRT, one can also allow the vertex n to connect to $m \in \mathbb{N}$ vertices, each selected independently and uniformly at random. This yields a more general model known as the m-Directed Acyclic Graph (m-DAG), which was introduced by Devroye and Lu in [44]. Devroye and Lu also study the asymptotic growth of the maximum degree of the m-DAG model in [44], is $\log n / \log_{1+1/m} n$ as n tends to infinity. Moreover, Devroye and Janson [43] study the length of paths to the root, which are of logarithmic order.

Preferential attachment models

In 1955, the economist Herbert Simon proposed a stochastic model aiming to explain observations in data describing phenomena in sociological, ecological and economical contexts [130]. The underlying motivation was that, due to the frequency of the observations and the wide range of contexts within which these observations were made, if these observations share any properties it can only be the similarity of the probability mechanisms that give rise to these observations.

Simon's stochastic model is described in terms of the number of words that appear in a written text. The assumptions of the model are as follows. First, given a text of k words, the probability that the $(k+1)^{\text{st}}$ word is a word that has already appeared exactly i times is proportional to i multiplied with the number of different words that have each appeared exactly i times (i.e. the total number of occurrences of words that each have occurred exactly i times). And second, there is a constant probability that the $(k+1)^{\text{st}}$ word is a new word. Based on these two assumptions, Simon was able to show that the frequency distribution of words obeys a power law and that the power-law exponent can be expressed in terms of the probability of adding a new word to the text.

This "rich-get-richer" effect, where wealthy individuals are able to acquire more wealth at a rate proportional to their current wealth (Simon's stochastic model for word frequencies described in terms of money) was also used by Yule in 1925 in an attempt to explain the distribution of biological genera among animal species [139]. As Yule's research was carried out prior to the development of the field of probability theory, Yule's methods and analysis were difficult and involved. Champernowne used similar ideas in 1953 to construct a stochastic model for income distribution in a population [29], though Simon's model requires weaker assumptions and is applied in a more general setting.

Price adapted Simon's approach and methods and was the first to apply them to a network setting in [123]. He named Simon's mechanism cumulative advantage, and used it to describe citation networks in scientific papers. In Price's model, papers are published sequentially and new papers cite existing ones. Price then assumes that a new paper cites an existing paper with a probability that is proportional to one plus the number of citation the existing paper already has. Since every paper starts out with zero citations, the addition of one ensures that a paper has a non-zero probability of receiving citations and hence one avoids a trivial model. This followed empirical observations Price made in [124] that citation frequencies of scientific papers follow a power-law distribution, making Price the first to observe scale-free networks as well, according to Newman [116].

Other mathematical network models based on a similar principle have been studied since, for example ordered recursive trees [125], non-uniform random recursive trees [132], random plane oriented recursive trees [98, 100] and random heap ordered recursive trees [30]. Despite all this research carried out, cumulative advantage or the "rich get richer" effect did not attract a lot of attention in the scientific community. It was only until the famous paper of Barabási and Albert [13], who independently observed and modelled the scale-free behaviour of links between webpages in the World

Wide Web, that the interest in cumulative advantage, or preferential attachment as Barabási and Albert coined it, increased significantly.

In the preferential attachment model developed by Barabási and Albert, nodes are sequentially added to the graph and are more likely to connect to nodes with high degree. The graph starts with m_0 nodes, and at every step a new node is added to the graph. This node is assigned $m \leq m_0$ edges and connects each edge to a different vertex already present. The probability an edge connects to a node i with degree k_i equals

$$\frac{k_i}{\sum_j k_j}, \tag{1.3.1}$$

where the term in the denominator sums over all vertices present in the graph. After n steps, one obtains a graph with $m_0 + n$ vertices and mn edges.

The description of this model is somewhat informal and does not specify how the first edge is created (if all nodes have degree zero, the probability in (1.3.1) is ill-defined), it is not clear whether self-loops are allowed and it is unclear whether there are dependencies between the m edges of a vertex. Nonetheless, Barabási and Albert did include simulations in their paper so that a specific model must have been used, and they observed their model gave rise to a power-law degree distribution with exponent $\tau = 3$.

Bollobás, Riordan, Spencer and Túsnady [22], and independently Móri [111], study this model in more detail and in a more rigorous manner. They not only make more precise choices for the model definition, but also prove that the degree distribution converges to a power-law distribution with exponent $\tau = 3$ (though Móri already studies a more general model that allows for exponents $\tau \in (2, \infty)$).

Empirical observations for preferential attachment mechanisms governing the growth of networks are presented in, among others, [83, 117]. Though linear preferential attachment, as proposed by Barabási and Albert, is observed in certain contexts such as the Internet and citation networks, other cases such as collaboration networks seem to fit better with other attachment rules, such as a sub-linear dependence on the degrees. Beyond linear preferential attachment rules, such sub-linear and super-linear rules have been studied from a theoretical perspective as well. In these cases, the probability a new vertex attaches to a vertex of degree k is proportional to k^γ with $\gamma \in (0, 1)$ and $\gamma > 1$, respectively. Such models with sub-linear attachment rules give rise to degree distributions with stretched exponential tails, see e.g. [40, 89, 91]. With super-linear attachment, on the other hand, every fixed vertex only acquires a finite number of edges and in particular cases a single vertex acquires all but finitely many edges, see [8, 120].

A more general model of linear preferential attachment, allowing for a random out-degree is studied by Deijfen, Van den Esker, Van der Hofstad and Hooghiemstra in [35]. Even more general attachment rules, where the probability to connect to a vertex of degree k is proportional to $f(k)$ for some function $f : \mathbb{N}_0 \to \mathbb{R}_+$ are studied by Athreya [8], Athreya, Ghosh, Sethuraman [9], Holmgren and Janson [78], Rudas, Tóth, Valkó [127], Oliviera and Spencer [120], Dereich and Mörters [41], Bhamidi [14] and Banerjee and Bhamidi [11]. All these papers use an embedding of preferential attachment trees in Crump-Mode-Jagers branching processes to obtain their results.

Other than the degree distribution, the behaviour of the maximum degree in preferential attachment models is a topic of interest as well. Móri [112] first studies the maximum degree in linear preferential attachment models. Later, Athreya (and Ghosh and Sethuraman) show that there exists a fixed (though possibly random) vertex that attains the maximum degree at all but finitely many steps [8, 9], a property that was later coined 'persistence' by Dereich and Mörters. Dereich and Mörters [40] study maximum degrees in sub-linear preferential attachment models and show persistence does not hold for such attachment rules. Rather, new vertices constantly compete for the maximum degree and the index of the maximum degree at time n is shown to diverge as a stretched exponential in $\log n$. Banerjee and Bhamidi [11] study the persistence of hubs for more general attachment rules as well as for more general sequences of out-degrees, and formulate precise criteria under which persistence does or does not hold, and provide asymptotics for the index of the maximum degree in the latter case.

Persistence is an interesting concept and can influence the behaviour of a network as a whole. In real-world networks, persistence can be exhibited in particular contexts, but does not always seem realistic. In protein-protein interaction (PPI) networks, the essential proteins that interact with many other proteins (and as a result, form hubs in the PPI network) often are present earliest on an evolutionary timescale [47]. In social networks like Twitter, it is, however, not the early adapters that necessarily have the largest number of followers. Instead, new and more popular individuals appear as time passes which obtain large numbers of followers rapidly. We aim to provide a different view and probable explanation for the dichotomy of (non-)persistence in this thesis.

Finally, preferential attachment models have been shown to be (ultra-)small worlds, depending on the scale-free exponent of the degree distribution. Most attention related to typical distances is devoted to linear preferential attachment mechanisms. When the scale-free exponent τ is larger than three, so that the degree distribution has a finite variance, distances are of logarithmic order [46]. A phase transition occurs when $\tau \in (2,3)$ (finite mean, infinite variance degree distribution), in which case double logarithmic typical distances as well as diameters can be observed (see [38, 46] for typical distances and [28, 46] for the diameter of preferential attachment models with infinite variance degree distributions).

Somewhat surprisingly, it turns out that both typical distances and the diameter of these preferential attachment models are twice as long as observed in other models such as the configuration model and inhomogeneous random graphs (when these models also exhibit a power-law degree distribution with infinite variance and finite mean). Contradicting the conjectures at the time that all these models belong to the same universality class, in the sense that their behaviour is very similar and independent of the precise model definitions (see [74]), the behaviour of typical distances in preferential attachment models turned out to behave differently. The main reason is that single edge connections between vertices with high degree are not (much) more likely than single edge connections between a vertex with low degree and a vertex with high degree in the preferential attachment model, whilst this is the case in the configuration model and inhomogeneous random graph models. Instead, two vertices with high degree are likely to be connected on a path of length two, via a young vertex that connects to both high degree vertices. This is exactly what gives rise to the factor two in the typical distances and diameter when compared to the other models mentioned before. Note that this is only possible if every vertex (or at least most vertices) has (have) an out-degree of at least two, so that this observation does not hold for preferential attachment trees. Indeed, here typical distances are always of logarithmic order, independent of the value of $\tau \in (2,\infty)$. The critical value $\tau = 3$ yields typical distances of order $\log n/\log\log n$ distances, where the precise constant depends on lower order terms of the degree distribution, see [21, 39].

In the case of infinite variance degree distributions ($\tau \in (2,3)$), typical distances have been extended to weighted distances (each edge is assigned an i.i.d. edge-length) in [86], where conditions on the edge-length distribution are formulated which are very closely related to the conditions for (non-)explosion in continuous-time Crump-Mode-Jager branching processes. Also, an interesting topic has been introduced in [85], in which the evolution of typical distances as a function of time in preferential attachment models is studied.

Beyond preferential attachment models with linear, sub-linear and super-linear attachment rules, many other varieties have been introduced and studied. We provide a short overview of some of the well-known variants here.

Directed preferential attachment

The discussion of preferential attachment models so far has considered non-directed graphs only. After all, in motivating examples such as collaboration networks the edges have no direction, and in citation networks citations always go backward in time, so that the vertex labels indicate the direction of the edge/citation and a directed graph is not required. There are, however, examples in which directed edge make sense. On social media, edges could represent sending and/or responding to messages or who follows whom, and in the World Wide Web hyperlinks have a clear direction towards webpages.

Bollobás, Borgs, Chayes and Riordan investigate a directed preferential attachment model in [18]. Here, one starts with some graph G_0 and fixed parameters $\alpha, \beta, \gamma, \delta_{\text{in}}$, $\delta_{\text{out}} \geq 0$, with $\alpha + \beta + \gamma = 1$. At each step, with probability α, a new vertex v is introduced and a directed edge (v, w) is created. With probability β, a directed edge (v, w) between two existing vertices v, w is created, and with probability γ a new vertex v is introduced and a directed edge (w, v) is created. In all cases, vertices are selected using a linear preferential attachment mechanism with respect to the in- or out-degrees of vertices. For non-trivial parameters settings, it is shown that both the in-degree as well as the out-degree sequence converge almost surely to power-law distributions.

Davis, Resnick, Wan and Wang consider the same model and fit it to network data, estimating parameters with maximum likelihood estimators in [136]. Resnick and Wang also study the concept of reciprocity, which, for example, characterises communication between users on social media platforms. It is defined as the average number of directed edges going back and forth between all pairs of vertices. In [137] they show reciprocity is not exhibited for most parameter choices of this directed preferential attachment model.

Preferential attachment with types

Antunović, Mossel and Rácz [7] introduced a preferential attachment model where each vertex has one of two types, say type zero and one. New vertices connect to m existing vertices via a linear preferential attachment rule and they obtain type zero with probability p_k, where k equals the number of vertices with type zero they connected with, and type one otherwise. The main topic of interest in the paper is the behaviour and convergence of the fractions of types depending on the initial configuration of vertex types and the probability distribution p_k. Jordan extends these results by allowing vertices of different types to be selected with different preferential attachment rules [84], and Haslegrave and Jordan consider three types [69]. In this case there are conditions under which the fractions of types do not converge almost surely.

Preferential attachment with choice

Paquette and Malishkin [103, 104], Krapivsky and Redner [90] and Haslegrave and Jordan [68] study a choice-based preferential attachment model where a new node selects a fixed number of potential neighbours according to a particular preferential attachment rule and connects to one of the potential candidates via a deterministic criterion, e.g. select the vertex with the highest/lowest degree. The different papers show the degree distribution can exhibit exponential decay or a power law, and that under particular conditions condensation-like behaviour can occur.

Haslegrave, Jordan and Yarrow [70], Grauer, Lüchtrath and Yarrow [65] and Freeman and Jordan [59] study preferential attachment with location-based choice, where each node is assigned an independent uniform location on $(0,1)$ and the choice-criterion is based on the locations of the vertices. The introduction of locations allows for more rich behaviour to be observed.

Spatial preferential attachment

One of the main critiques of preferential attachment models is the lack of clustering. As is the case in the Erdős-Rényi random graph, inhomogeneous random graphs and the configuration model, these graphs are locally tree-like, something which is often not the case in real-world networks. Clustering can arise due to the underlying geometry that governs connections, which, among others, is used in the hyperbolic random graph model described in Section 1.3.1. Spatial preferential attachment models use geometry to allow for more clustering via spatial dependence among vertices.

Several spatial preferential attachment models have been introduced over time. Manna and Sen [105] first considered a spatial preferential attachment model where vertices are assigned a random position in the unit square in two dimensions and a new vertex v establishes connections to vertices u with a probability proportional to the degree of u multiplied with $d(u,v)^\alpha$, where $d(\cdot,\cdot)$ denotes the Euclidean distance and $\alpha \in \mathbb{R} \cup \{\infty, -\infty\}$ is a fixed parameter. For $\alpha = \infty, -\infty$, the node is only linked to the furthest and nearest vertices, respectively. Flaxman, Frieze and Vera study a spatial preferential attachment model where connections are established with a linear preferential attachment rule among vertices at distance at most $r = r(n)$, vertices have a fixed out-degree m and a uniform location in the unit sphere in $\mathbb{R}^3$ [55]. Aiello, Bonato, Cooper, Janssen and Prałat consider a spatial model where vertices have both a location on the hypercube in $\mathbb{R}^m$ and a region of influence [5]. These regions scale proportional with the degree of a vertex and scale inversely with time, so that high degree vertices have a larger region of influence and over time an ageing effect occurs in which vertices become less attractive. Finally, Jacob and Mörters study a model in [80] in which new nodes are born in the one-dimensional torus according to a rate one Poisson process and connect to existing nodes with a probability $\phi(t\rho/f(d))$, where t denotes time, ρ denotes the distance between the two nodes, d denotes the in-degree of the existing vertex, f is an increasing attachment function and ϕ is a decreasing profile function. The attachment rule determines the "strength" of a vertex and the profile function determines the spatial dependence. Jacob and Mörters argue this model is a generalisation of the model introduced by Aiello et al. in [5].

1.4 Evolving models in random environment

In the evolving models described above, the only characteristic that enables us to distinguish between vertices is their age (or perhaps type or location in some of the 'non-classical' models). Indeed, young vertices are often observed to behave very differently compared to old vertices in these models, take the location of the maximum degree in the random recursive tree and preferential attachment model as an example. Similarly, vertices of a similar age are expected to show similar behaviour. In the preferential attachment model above, all vertices with the same degree are equally likely to connect to a new vertex, and in the random recursive tree all vertices are always equally likely to attract new connections.

Clearly, in a real-world setting this is hardly ever the case. In the World Wide Web, some webpages may be intrinsically more attractive and interesting and attract more links. Websites providing useful services, such as directories or encyclopedias, are much more likely to receive new links when compared to personal homepages. In citation networks, certain papers perceived to be of higher quality compared to other papers and as a result attracts more citations from newly written papers. On Twitter, particular people are more active or post more controversial content, leading to more followers than the average person.

Moreover, the number of citations of a paper is often a way to measure how influential a paper has been. Search engines base the importance and usefulness of webpages on the number of links webpages receive. And people with many followers on Twitter are considered interesting and fashionable. This would not be the case unless there is some correlation between the perceived quality of a vertex in the network and the degree of the vertex. Though it is not always clear a priori why certain vertices are better able than others to attract connections, every vertex seems to have an inherent ability to do so.

Allowing for the attractiveness of vertices to vary introduces more heterogeneity in the network. No longer are all vertices the same and is their behaviour identical. The local principles which govern the network formation can still be similar, but the individual dynamics of each vertex can now differ. As a result, one could expect the properties of the network created to change compared to the networks in which all vertices (with the same degree) have the same inherent attractiveness. Degree distributions may no longer be or, instead, become scale free, typical distances could increase or decrease and hubs in the network might behave very differently. In the models introduced below, the attractiveness of a vertex is encapsulated by a single value denoted as the *fitness* or *weight* of the vertex. These fitness variables are non-negative i.i.d. random variables from some underlying distribution. Incorporating this fitness into the construction of the network allows for much richer behaviour.

As we assign a random variable to every node in the graph, such evolving random graph models can be interpreted as models in a *random environment*. The main aim of this thesis is to understand the influence of the random environment on the local and global behaviour of the random graphs.

Weighted recursive graphs

Possibly one of the simplest examples of an evolving random graph in a random environment is the weighted recursive graph (WRG). This is a generalisation of the random recursive tree, in which new vertices do not connect to predecessors selected uniformly at random, but where each vertex is assigned a (random) vertex-weight and new vertices select predecessors with a probability proportional to their weight. Depending on the structure and assumptions on the behaviour of the sequence of weights, very rich and, most importantly, very different behaviour can be observed when compared to the random recursive tree and directed acyclic graph (which can be recovered from the WRG by assigning all vertices the same weight).

A weighted recursive tree (WRT) is a WRG in which each new vertex is allowed to connect to exactly one predecessor, yielding a tree. The weighted recursive tree was first introduced by Borovkov and Vatutin in [25, 24] who named it a random recursive tree in random environment, and where the vertex-weights $(W_i)_{i\in\mathbb{N}}$ satisfy a product form $W_i = \prod_{j=1}^{i} X_j$ for some i.i.d. random variables $(X_i)_{i\in\mathbb{N}}$. Hiesmayr and Işlak [71] used a slightly different definition of what they named the weighted recursive tree in which the weight of the first vertex can be random as well, which is the definition we use in this thesis. They studied the height, depth and branch sizes of the model.

Uribe Bravo and Mailler introduced the name weighted random recursive tree [102] and applied its properties to study random walks with preferential relocation and fading memory. The main topic of interest are the height and profile of the tree when the weights are i.i.d. random variables from certain underlying distributions. Uribe Bravo and Mailler establish that the height of a typical node is highly dependent on the underlying distribution of the vertex-weights. Where the height typical nodes is of logarithmic order in the RRT model (see [53]), the height of typical nodes in the WRT model can range from $\log\log n$ to $(\log n)^\alpha$ with $\alpha > 0$ to even n^δ with $\delta \leq 1/2$. Sénizergues allows for deterministic weight sequences with more general assumptions and studies degree sequences and the height and profile of the tree in [128], and provides more detailed asymptotic behaviour of the height in [121] together with Pain. Sénizergues combines assumptions on the weight sequences with the fact that the in-degree of a vertex is identical in law to a sum of independent indicator random variables to establish the pointwise convergence as well as convergence in ℓ^p of the in-degrees in the WRT model. For the profile and height, Sénizergue proves the convergence of the Laplace transform of the profile, defined on the complex plane, from which he can then obtain the desired properties of the profile and height. For the more refined asymptotic behaviour of the height of the WRT model, Sénizergues and Pain adapt methods used to analyse the maximum displacement of a branching random walk in [121] to obtain more precise results on the asymptotics of the height of the WRT model.

Iyer [79] studies an evolving weighted tree with a more general attachment rule, where a new vertex connects to a predecessor v with a probability proportional to $f(\deg_n(v), W_v)$, for some function f and where $\deg_n(v)$ and W_v are the degree and vertex-weight of vertex v. Note that setting $f(x, y) = y$ for all $x \in \mathbb{N}_0, y \geq 0$ yields the WRT model. As a result, this model can be used to study other examples of evolving weighted trees as well, such as weighted Cayley trees. Iyer uses an embedding of the WRT model in a continuous-time Crump-Mode-Jagers process and provides conditions under which almost sure convergence of the degree distribution is obtained.

Fountoulakis, Iyer, Mailler and Sulzbach study a model for random simplicial complexes in $d \geq 0$ dimensions in [57], which generalises WRTs (though not WRGs). In this model, the vertices of an evolving d-dimensional simplicial complex are equipped with bounded vertex weights and the evolution is determined by a fitness function $f : [0,1]^d \to \mathbb{R}_+$ and the vertex-weight distribution μ. At every step, a face of the complex is selected with a probability proportional to the sum of the connection function applied to each of the weights of the vertices adjacent to the face, and a new vertex is introduced and connected to each of the adjacent vertices (hence creating d more faces). Using measure-valued Pólya urns, they prove the almost sure convergence of the empirical degree distribution under certain conditions on the fitness function f and the vertex-weight distribution μ.

Preferential attachment models with fitness
Several preferential attachment models in random environment, often referred to as preferential attachment models with fitness, have been introduced and studied since the paper by Albert and Barabási. These aim to better explain the mechanisms that govern the formation of real-world networks such as the World Wide Web, citation networks, etcetera. Empirical evidence suggests that the acquirement of hyperlinks by webpages, citations by papers and investment by companies, takes place at different rates for different webpages, papers and companies, respectively. This points towards deviations in the growth mechanisms in the sense that it is (slightly) different among entities (webpages, papers and companies). These deviations are not present in the preferential attachment model, as discussed above. Bianconi and Barabási introduced a model that allows for these deviations to occur [15]. Here, every vertex i is assigned a fitness $\mathcal{F}_i$, and new vertex $n+1$ connects to vertex i with probability

$$\frac{\mathcal{F}_i \deg_n(i)}{\sum_{j=1}^n \mathcal{F}_j \deg_n(j)}, \tag{1.4.1}$$

where $\deg_n(i)$ denotes the degree of vertex i in the graph created up to that point, consisting of n vertices. Due to the multiplicative nature of the connection rule, this model is also known as a preferential attachment model with multiplicative fitness. It follows that the vertices that are most likely to attract edges from new vertices are the ones that are already well-connected and have a high fitness, in other words, the ones that are both popular and fit. At the same time, the fittest vertices are most likely to increase their degree. Bianconi and Barabási conjectured the existence of three different phases which depend on the underlying fitness distribution. Borgs, Chayes, Daskalakis and Roch study these phases rigorously in [23] and denote these phases as:

- First-mover-advantage phase, in which the inclusion of fitness leads to no significant different in behaviour compared to the linear preferential attachment model,
- Fit-get-richer phase, in which the inclusion of fitness allows the more fit vertices to acquire edges at a higher rate than less fit vertices.
- Innovation-pays-off phase, in which a non-zero proportion of the edges is attracted by vertices with higher and higher fitness values.

Dereich and Ortgiese [42] further study the fit-get-richer and innovation-pays-off phases (which they refer to by the Bose-Einstein phase) by determining the almost sure limit of the empirical degree weighted fitness measure,

$$\Gamma_n := \frac{1}{n}\sum_{i=1}^n \mathcal{Z}_n(i)\delta_{\mathcal{F}_i}, \tag{1.4.2}$$

where $\mathcal{Z}_n(i)$ and $\mathcal{F}_i$ denote the in-degree and fitness of vertex i, respectively, and where δ is a Dirac measure. Moreover, they provide conditions for condensation to occur under rather general assumptions on the attachment rules of the model, by applying stochastic approximation arguments to Γ_n. If we let μ be the fitness distribution and we assume that the essential supremum of μ equals one, then with $\lambda > 0$ a model parameter which controls the number of edges in the graph (which is of the order λn when the graph consist of n vertices), and

$$\int \frac{f}{1-f}\mu(\mathrm{d}f) < \lambda,$$

then condensation occurs. This implies that a positive fraction of newly incoming edges attaches itself to a set of vertices with fitness moving closer and closer to the essential supremum of μ. In the limit of the empirical degree distribution, this Dereich [36] furthers this by analysing a slight variant of the model where the normalising term in the connection probabilities is deterministic rather than random. Dereich analyses the *condensate*, a small number of vertices with exceptionally high degree, studying which vertices belong to the condensate and determining the qualitative and quantitative properties of the condensate, under the assumption that the fitness distribution belongs to the Weibull maximum domain of attraction. Most notably, contrary to prior conjecture [15], Dereich finds that, typically, it is not a single vertex (or even a finite number of vertices) that constitute the condensate. Rather, the size of the condensate diverges as the graph size tends to infinity.

Dereich, Mailler and Mörters [37] and Mailler, Mörters and Senkevich [101] study a more general model known as reinforced branching processes, a model which covers branching processes with selection and mutation, generalised Pólya urn models and the Bianconi-Barabási model. In reinforced branching processes individuals are assigned fitness values, are organised into families and members of the same family have the same fitness. A family of size k and fitness f gives birth to new individuals at a rate kf and every new individual starts a new family with probability $\beta \in [0, 1]$ with a fitness drawn from an underlying fitness distribution μ, stays with the family it was born from with probability $\gamma \in [0, 1]$ or both of these events happen (i.e. a new family is started and an already existing family increases its size by one) with probability $\beta + \gamma - 1$ (β and γ are such that $\beta + \gamma \geq 1$). The case $\gamma = 1 - \beta$ yields branching processes with mutation and selection, $\beta = \gamma = 1$ yields the Bianconi-Barabási tree. Dereich, Mailler and Mörters establish a 'Winner does not take it all' principle, which they coin non-extensive condensation, showing that the size of the largest family is negligible compared to the number of individuals. This matches with the findings of Dereich [36] in the sense that the number of families that contribute to the condensate diverges. They also describe the asymptotic size and fitness of the largest family, under the assumption that μ belongs to the Weibull maximum domain of attraction. Mailler, Mörters and Senkevich are able to obtain similar results for a large class of fitness distributions in the Gumbel maximum domain of attraction.

The multiplicative nature of the Bianconi-Barabási model has a very clear effect on the model. Even a fitness distribution with bounded support can already result in behaviour that belongs to the innovation-pays-off phase. A model using a combination of multiplicative fitness and ageing is considered by Garavaglia, Van der Hofstad and Woeginger in [60], for which it is shown that the inclusion of ageing can allow for a wider range of fitness distributions in order to obtain scale-free behaviour.

Introducing ageing, which makes vertices less attractive as time passes, is one way to temper the strong effect of the multiplicative fitness. A different method is to use the fitness in an additive rather than a multiplicative way. That is, in the same setting as before (1.4.1), a new vertex $n + 1$ now connects to a vertex i with a probability

$$\frac{\deg_n(i) + \mathcal{F}_i}{\sum_{j=1}^{n} \deg_n(j) + \mathcal{F}_j}.$$

In this case, the fitness has an initial effect which relatively diminishes as the degree of a vertex increases. This model, known as preferential attachment with additive fitness (PAF), was proposed by Ergün and Rodgers in [51]. This model still allows for a better and more natural explanation of the mechanisms governing the formation of real-world networks compared to the preferential attachment model proposed by Albert and Barabási, but the fitness values interact with the degree evolution in a different manner compared to the Bianconi-Barabási model. It is immediately clear that the effect of the fitness is much more subtle than in the case of multiplicative fitness, where the increase of the degree enhances the effect of the fitness rather than relatively diminishing it. As a result, one would expect a fitness distribution with a heavier tail to be required in order to see significantly different behaviour. It begs the question whether the three phases described above for the Bianconi-Barabási model can still be observed in the additive case as well, or whether fewer or other kinds of phases appear. This is one of the topics of interest in this thesis.

Ergün and Rodgers [51] argue that the empirical degree distribution in the PAF model still converges to a power-law distribution, where the power-law exponent depends linearly on the mean of the fitness distribution [51]. Bhamidi studies a wide range of preferential attachment models in [14] using an embedding in Crump-Mode-Jager branching processes, among which PAF trees. Under the assumption that the fitness values are almost surely bounded, Bhamidi obtains that the empirical degree distribution converges in probability to an explicit limiting degree distribution, which agrees with Ergün and Rodger's observations and non-rigorous arguments, and proves tightness of the rescaled maximum degree.

Iyer [79] studies a general model of evolving random trees with fitnesses, also using an embedding in Crump-Mode-Jager branching processes as Bhamidi, but is able to obtain an almost sure limit of the degree-weighted fitness measure Γ_n as in (1.4.2) under more general assumptions on the fitness distributions and more general attachment functions f which depends on the in-degree and fitness of the vertex. Under the additional assumption that $f(i, W) = ig(W) + h(W)$ for some functions g, h (here i denotes the in-degree of a vertex and W its vertex-weight), Iyer present conditions for the occurrence of condensation.

Fountoulakis and Iyer also study the degree-weighted fitness measure Γ_n as in (1.4.2) of a more general model which includes neighbourhood influence in terms of the vertex-weights of neighbours [56], using embedding in Crump-Mode-Jager branching processes. In this model a new vertex $n+1$ connects to a vertex i (the edge is directed towards $n+1$) with a probability proportional to

$$f(\mathcal{F}_i) + \sum_{j\sim i} g(\mathcal{F}_i, \mathcal{F}_j),$$

where $j \sim i$ denotes all out-neighbours of i and $f, g : [0, w^*] \to \mathbb{R}_+$ are two functions and $w^* > 0$ is a fixed constant. The WRT is recovered by setting $f(x) = x, g \equiv 0$, the PAF model by setting $f(x) = x, g \equiv 1$, and the Bianconi-Barabási model by setting $f(x) = g(x, y) = y$. Under assumptions on the functions f and g, the almost sure convergence to a limiting degree distribution is proved.

Equivalence between WRTs and PAFs

Sénizergues discusses an interesting equivalence between the WRT model and the PAF model (in the tree case) in [128]. For a fitness sequence $(\mathcal{F}_i)_{i\in\mathbb{N}}$, construct a WRT with weight sequence $(W_i)_{i\in\mathbb{N}}$ defined as

$$W_1 := 1, \qquad \forall\, n \geq 2, \quad W_n := \prod_{k=1}^{n-1} \beta_k^{-1}, \qquad \beta_k \sim \mathrm{Beta}\Big(\sum_{j=1}^{k} \mathcal{F}_j + k, \mathcal{F}_{k+1}\Big).$$

Then, the WRT with weight sequence $(W_i)_{i\in\mathbb{N}}$ and the PAF tree with fitness sequence $(\mathcal{F}_i)_{i\in\mathbb{N}}$ coincide in law. This equivalence is obtained by using a Pólya urn representation of the PAF and WRT models, which explains the beta random variables in the construction of the weights $(W_n)_{n\in\mathbb{N}}$. It allows Sénirzergues (and Pain) to carry over the results on WRTs in [128] (and [121]) (discussed above) to PAF trees when the fitness sequence satisfies $\max_{i\in[n]} \mathcal{F}_i \leq n^{c+o(1)}$ for some $c \in [0, 1/(\mu+1))$, where $\mu < \infty$ denotes the mean of the fitness distribution. This includes the convergence of the rescaled degrees of fixed vertices, the rescaled maximum degree and the asymptotic behaviour of the height of the tree. This same equivalence between the WRT and PAF tree model is also used by Lo in [95] in order to obtain, under the assumption that the fitness distribution has finite support (though Lo states that the results hold for distributions with exponentially decaying tails and for the multigraph case as well without providing further details), the weak local convergence of the PAF model to what she coins the π-Pólya point tree. As a result, the almost sure convergence of the empirical degree distribution is obtained in total variation distance when the fitness distribution has finite fourth moment.

1.4.1 Model definitions

In this thesis, we focus on the study of preferential attachment models with additive fitness and the weighted recursive graph model with random weights. We provide a precise definition of these models here.

First, we let $(\mathcal{G}_n)_{n\in\mathbb{N}}$ be a sequence of graphs, denote by $\mathcal{Z}_n(i)$ the in-degree of vertex i in $\mathcal{G}_n$ and let $[t] := \{i \in \mathbb{N} : i \leq t\}$ for $t \geq 1$.

Definition 1.4.1 (Preferential attachment with additive fitness)**.** Let $(\mathcal{F}_i)_{i\geq 1}$ be a sequence of i.i.d. copies of a random variable $\mathcal{F}$ taking values in $(0, \infty)$. For any $n \in \mathbb{N}$, define

$$S_n := \sum_{i=1}^{n} \mathcal{F}_i.$$

Let $n_0, m_0, m \in \mathbb{N}$. We construct the *preferential attachment graph with additive fitness* as follows:

(a) Start with some graph $\mathcal{G}_{n_0}$ which consists of n_0 vertices and m_0 edges and assign each vertex $i \in [n_0]$ the fitness $\mathcal{F}_i$.

(b) For each $n \geq n_0$, introduce a new vertex $n+1$ and assign it the vertex-weight $\mathcal{F}_{n+1}$ and m half-edges. Conditionally on $\mathcal{G}_n$, independently connect each half-edge to some $i \in [n]$ with probability

$$\frac{\mathcal{Z}_n(i) + \mathcal{F}_i}{m_0 + m(n-n_0) + S_n}.$$

Let $\mathcal{G}_{n+1}$ denote the resulting graph.

For the weighted recursive graph model, we switch from the notation $\mathcal{F}$ to W here for consistency with the upcoming chapters.

Definition 1.4.2 (Weighted Recursive Graph)**.** Let $(W_i)_{i\geq 1}$ be a sequence of i.i.d. copies of a non-negative random variable W taking values in $(0, \infty)$. For any $n \in \mathbb{N}$, define

$$S_n := \sum_{i=1}^{n} W_i.$$

Let $m \in \mathbb{N}$. We construct the *Weighted Recursive Graph* as follows:

1) Initialise the graph with a single vertex 1, the root, and assign to the root a vertex-weight W_1. We let $\mathcal{G}_1$ denote this graph. .

2) For $n \geq 1$, introduce a new vertex $n+1$ and assign to it the vertex-weight W_{n+1} and m half-edges. Conditionally on $\mathcal{G}_n$, independently connect each half-edge to some vertex $i \in [n]$ with probability W_i/S_n. Let $\mathcal{G}_{n+1}$ denote this graph.

1.5 Main results of the research

The results provided in this thesis describe the degree evolution of vertices in the PAF and WRG models. In particular, the main focus is on the largest degree and (in certain cases) near-maximum degrees. Most importantly, we are interested in the description of the three phases by Borgs et al. in [23] for the Bianoni-Barabási model discussed in Section 1.4 and formulate similar phases in which different behaviour can be observed for the PAF models and the WRG model.

For the PAF model, we study the degree distribution and the maximum degree. As already claimed by Ergün and Rodgers in their introduction of their preferential attachment model with additive fitness [51], when the fitness distribution μ has a finite mean we obtain a limiting distribution $(p_k)_{k\in\mathbb{N}_0}$, with

$$p_k := \int_0^\infty \frac{\theta_m}{x+\theta_m}\prod_{\ell=1}^{k}\frac{(\ell-1)+x}{\ell+x+\theta_m}\mu(\mathrm{d}x), \qquad k \in \mathbb{N}_0,$$

though the result presented here holds more generally. Additionally, our methods provide a rigorous proof and almost sure convergence. Unlike the claim made by Ergün and Rodgers, we also note that this limiting degree sequence exhibits a phase transition in the power law exponent, which depends on the precise underlying assumptions for the fitness distribution μ. Moreover, when the fitness distribution has an infinite mean, we obtain that the empirical degree distribution no longer exhibits a power-law. Rather, a typical vertex receives no edges after its introduction to the graph with high probability.

In [95] Lo is able to prove convergence of the empirical degree distribution to the above limit in total variation distance. However, Lo proves this for distributions μ such that the fourth moment is finite only.

The phase transitions observed in the limiting degree distribution p_k can be extended to the behaviour of the maximum degree as well. In particular, we provide conditions which determine whether persistence holds and obtain the growth rate of the maximum degree relative to the growth rate of the degree of fixed degree vertices. We recall that persistence means that there exists a fixed vertex which attains the maximum degree for all but finitely many steps. For a more concise formulation (though not as general as presented later in Chapter 2), we assume that the fitness distribution μ follows a power-law. That is, for some slowly-varying function $\ell : \mathbb{R}_+ \to \mathbb{R}_+$,

$$\mu(x,\infty) := \ell(x)x^{-(\alpha-1)}, \qquad x > 0. \tag{1.5.1}$$

Moreover, we let $\theta_m := 1 + \mathbb{E}[\mathcal{F}]/m$, where m is the out-degree of each vertex and $\mathcal{F}$ is a random variable with law μ. We then identify the following regimes:

- *Weak disorder regime*: When $\alpha > 1+\theta_m$, persistence occurs and the maximum degree grows at the same rate n^{1/θ_m} as the degree of any fixed vertex.
- *Strong disorder regime*: When $\alpha \in (2, 1+\theta_m)$, persistence does not occur and the index of the vertex attaining the maximum degree is of order n. The maximum degree grows at rate $n^{1/(\alpha-1)}$, whereas the degree of any fixed vertex still grows at n^{1/θ_m}, so that the maximum degree grows faster than the degree of any fixed vertex.

- *Extreme disorder regime*: When $\alpha \in (1,2)$, persistence does not occur and the index of the vertex attaining the maximum degree is of order n. The maximum degree grows linear in n, whereas the degree of any fixed vertex is finite almost surely for all $n \in \mathbb{N}$.

Though not entirely the same, the weak, strong, and extreme disorder regimes exhibited by the PAF are comparable to the first-mover-advantage, fit-get-richer and innovation-pays-off phases for the Bianconi-Barabási model described by Borgs et al. [23], though the conditions for the fitness distribution are very different. The identification of the weak disorder regime matches with results obtained by Sénizergues in [128], though here we can allow for an out-degree $m > 1$ as well as somewhat more general attachment rules, which are not included in Definition 1.4.1. The identification of the strong and extreme regimes is novel and here we are able to provide a precise description of the limiting distributions of the rescaled maximum degree and rescaled index of the maximum degree (the latter only in the strong disorder regime).

For the WRG model, we again study the degree distribution and the maximum degree. One can interpret the WRG model as the PAF model *without* the preferential attachment feedback mechanism that makes vertices more likely to attract edges as their degree increases. Instead, vertices have a fixed 'attractiveness' determined by their vertex-weight. The preferential attachment mechanism has a strong effect on the overall behaviour of the network relative to the fitness, in the sense that only sufficiently heavy-tailed fitness distributions are able to cause a significant change in the behaviour of the degree distribution and the maximum degree, as discussed above. Omitting this mechanism therefore allows the more subtle influence of the fitness/vertex-weights to become apparent.

The empirical degree distribution of the WRG model almost surely converges to a limiting distribution $(p_k)_{k\in\mathbb{N}_0}$ when the vertex-weights have finite mean, with

$$p_k := \int_0^\infty \frac{\theta_m - 1}{\theta_m - 1 + x}\Big(\frac{x}{\theta_m - 1 + x}\Big)^k \mu(\mathrm{d}x), \qquad k \in \mathbb{N}_0,$$

where we recall that $\theta_m := 1 + \mathbb{E}[W]/m$ and W is a random variable with law μ. This is also obtained for the WRT model as a special case in the more general 'evolving random trees with fitness' studied by Iyer in [79]. In the case that the vertex-weights have an infinite mean (and the tail distribution is slowly varying as in (1.5.1)), the same result as for the PAF model is obtained; a typical vertex attracts no edges after its introduction to the graph with high probability.

Where the limiting degree distribution for the PAF model is always a power law (in the case of finite mean random variables), the limit $(p_k)_{k\in\mathbb{N}_0}$ here is strongly influenced by the choice of the distribution μ. The asymptotic behaviour of p_k can range from geometric to stretched exponential to log-compressed exponential to power law, depending on the underlying vertex-weight distribution.

As is the case for the degree distribution, the behaviour of the maximum degree is also highly dependent on the vertex-weight distribution μ. Unlike the PAF model, the behaviour of the maximum degree now can be classified based on whether the vertex-weights are bounded or unbounded, and on the maximum domain of attraction (MDA) the vertex-weight distribution belongs to (assuming it belongs to any MDA). A distribution μ belongs an MDA when there exist sequences $(a_n, b_n)_{n\in\mathbb{N}}$, such that

$$\lim_{n\to\infty} \mu(-\infty, a_n x + b_n)^n = G(x), \tag{1.5.2}$$

for some distribution G and for all continuity points x of G. G can be exactly one of three families of distributions: Weibull, Gumbel or Fréchet. The book on extreme value theory by De Haan and Ferreira (and the references therein) provides a comprehensive introduction to maximum domains of attraction and the overarching extreme value theory [67]. We provide a rough outline of the behaviour of the maximum degree based on bounded or unbounded vertex-weights and further distinguish between the different MDAs μ belongs to.

Unbounded vertex weights
When the distribution μ has an unbounded support, which yields unbounded vertex-weights, the behaviour of the maximum degree is mainly determined by the interplay of the behaviour of the vertex-weight and the age of a node (i.e. the time it is introduced to the graph). For distributions with unbounded support, the Fréchet MDA and the Gumbel MDA are considered, where we note that the Gumbel MDA also consists of distributions with finite support, which we discuss later on.

Fréchet. When the distribution μ belongs to the Fréchet MDA, the maximum degree is of order $u_n := \inf\{x \in \mathbb{R} : \mu(x,\infty) < 1/n\}$ and the index of the maximum degree is of order n. Both the limit of the rescaled maximum degree as well as the rescaled maximum degree can be expressed in terms of a Poisson point process. An interesting observation is the following: in case μ has an infinite first moment, the limits are equivalent to the limits of the maximum degree and index in the PAF model when the fitness follows the same distribution. As the effect of the fitness is very strong for such a distribution, the preferential feedback mechanism is, as it were, overpowered by the fitness and the preferential attachment model with additive fitness (or at least its maximum degree) behaves as a weighted recursive graph.

Gumbel. When the distribution μ belongs to the Gumbel MDA, the precise expression of the tail distribution is essential in determining the behaviour of the maximum degree. We consider three different classes of tail distributions: log-compressed exponential ($\exp\big\{-(\log x)^\tau\big\}$ with $\tau > 1$), stretched and compressed exponential ($\exp\big\{-x^\tau\big\}$ with $\tau > 0$), and super-exponential (e.g. $\exp\{-\exp\{x\}\}$). In all cases, we show that the first order asymptotic behaviour of the maximum degree is of order $b_{k_n}\log(k_n/n)$ and the location of the maximum degree is roughly k_n, where b_n is as in (1.5.2) and $(k_n)_{n\in\mathbb{N}}$ is a suitable sequence for which $b_{k_n}\log(k_n/n)$ is maximised. In the log-compressed exponential case and the (stretched) exponential case, we also obtain the second order asymptotic behaviour, which depends on a_{k_n} (as in (1.5.2)) and a fine interplay of vertices having a high vertex-weight and being sufficiently young to allow for enough time to obtain a high degree.

Bounded vertex-weights
When the distribution μ has bounded support, which yields almost surely bounded vertex-weights, the effect of the tail distribution of μ is (even) more subtle than in the previous case. Here, large degrees arise via a different mechanism, involving a balance between the age of vertices and large deviation events in which a vertex' degree significantly outgrows its mean degree, like in the random recursive tree (though all vertex-weights are equal in law there). As a result, the first order asymptotic behaviour of the maximum degree is determined exactly by the first moment of μ only, and is of order $\log_{\theta_m} n$, where we recall that $\theta_m := 1 + \mathbb{E}[W]/m$, where W is a random variable with law μ.

Only when we consider higher-order asymptotic behaviour of the (near-)maximum degrees, do we observe different behaviour, though we can only prove this in the tree case (i.e. $m = 1$). These differences arise due to the tail distribution of the vertex-weight distribution μ and are closely linked to the tail distribution of the limiting degree distribution $(p_k)_{k\in\mathbb{N}_0}$. The distributions considered belong to the Weibull MDA or the Gumbel MDA, or have an atom at one (by the construction of the WRG we can assume without loss of generality that bounde vertex-weights are at most one), and for ease of writing we set $\theta := \theta_1$.

Atom at one. When the distribution μ has an atom at one, vertices need to have a vertex-weight equal to one in order to obtain a high degree (compared to other vertices). As a result, the behaviour of the WRG (in the tree case) with such a vertex-weight distribution is similar to the random recursive tree. We show that the difference of the maximum degree and $\lfloor \log_\theta n \rfloor$ converges in distribution along particular subsequences, without rescaling. The sub-sequential convergence is due to a lattice effect caused by the floor function applied to $\log_\theta n$ and as degrees only take integer values. Moreover, precise asymptotics for the distribution of the maximum degree, as well as asymptotic normality of the number of (near-)maximum degrees are obtained.

When μ does not contain an atom at one, the second-order behaviour of the maximum degree diverges to $-\infty$. As it is unlikely to observe a vertex with a weight larger than $1-\varepsilon_n$ for some sequence $(\varepsilon_n)_{n\in\mathbb{N}}$ in this case, the largest degrees do not grow as quickly as in the case when μ has an atom at one. How much smaller than $\log_\theta n$ the (near-)maximum degrees are depends on ε_n and thus on the tail of μ.

Weibull. When the distribution μ belongs to the Weibull MDA with parameter $\alpha > 1$, the second order asymptotic behaviour is shown to be $-(\alpha-1)\log_\theta \log_\theta n$. Higher-order behaviour, as well as the precise asymptotics for the distribution of the maximum degree and asymptotic normality of the number of (near-)maximum degrees, requires more assumption on the tail of μ, for which we provide an example.

Gumbel. When the distribution μ belongs to the Gumbel MDA, similar cases as for the unbounded distributions in the Gumbel MDA discussed above can be observed. The higher-order terms depend on the precise assumptions of the distribution, and we again provide an example for which this can be obtained.

1.6 Structure of the thesis

The thesis is structured as follows: In Chapter 2, we present our results related to the preferential attachment model with additive fitness. The contents of Chapter 2 are published in [97]. In Chapter 3 and 4 we investigate the weighted recursive graph and the weighted recursive tree model (the weighted recursive graph model with $m = 1$). Chapter 3 presents the content of [96], which deals with both unbounded and bounded vertex-weights, whereas Chapter 4 deals with bounded vertex-weights only and provides more precise results compared to what is discussed in Chapter 3 related to bounded vertex-weights. Finally, in Chapter 5 we provide a conclusion and discuss several open problems related to the models investigated in this thesis.

Chapter 2

A phase transition for preferential attachment models with additive fitness

In this chapter we consider the influence of additive fitness on several affine preferential attachment models and we are able to present a phase transition in terms of the behaviour of the degree distribution and the maximum degree. The findings presented in this chapter have been published by the author of this thesis and Marcel Ortgiese in *Electronic Journal of Probability* as an open access publication [97].

2.1 Outline of the article

We consider preferential attachment models with additive fitness, as first introduced by Ergün and Rodgers in [51]. Starting from an arbitrary graph $\mathcal{G}_{n_0}$ with n_0 vertices with labels $1,\dots,n_0$ and fitness values $\mathcal{F}_1,\dots,\mathcal{F}_{n_0}$, i.i.d. copies of some strictly positive random variable $\mathcal{F}$, and $m_0 \geq 1$ edges, the graph $\mathcal{G}_n$ is obtained at step $n > n_0$ from the graph $\mathcal{G}_{n-1}$ by introducing a new vertex n and assigning it a fitness value $\mathcal{F}_n \in (0,\infty)$ (which also is an independent copy of $\mathcal{F}$). This new vertex then connects to $m \in \mathbb{N}$ predecessors, each chosen with a probability proportional to the in-degree plus fitness of the predecessor. We can also allow for a random out-degree, where the vertex n connects to each predecessor with a probability proportional to its in-degree plus fitness and where connections to different vertices have a negative correlation, in spirit of the Bernoulli preferential attachment models studied by Dereich and Mörters in [41, 40]. The model derives its name from the fact that the fitness plays an additive role in these models, in contrast to the multiplicative nature of the fitness in the Bianconi-Barabási model.

In the literature it is well-known that the limiting degree distribution of the affine preferential attachment model, which can be interpreted as a special case of the model under investigation here, is a power law with an exponent τ that can be expressed in terms of the model parameters δ and m, see e.g. [22, 111] and [72] and the references therein. Moreover, the maximum degree scales polynomially with exponent $1/(\tau-1)$ and is attained at a fixed vertex for all but finitely many steps, a property known as *persistence*, see e.g. [8, 112].

We investigate how the properties of the degree distribution and the maximum degree change under the influence of the additive fitness. In particular, we outline a phase transition for the model depending on how quickly the tail of the fitness distribution decays, consisting of three phases in which different behaviour is observed. We coin these phases the *weak disorder regime*, *strong disorder regime* and *extreme disorder regime*. In Section 2 we present a detailed overview of the results related to the degree distribution and maximum degree, as well as the phase transition itself. In Section 4 we provide the details of the proof of the existence of this phase transition for the degree distribution, which uses a stochastic approximation approach in the weak and strong disorder regimes and a more straightforward first moment approach in the extreme disorder regime. Section 5 consists of technical preparations required for proving the convergence of the maximum degree in the extreme disorder regime. The methodology for this proof is a combination of the weak convergence of point processes and the concentration of the maximum degree around the conditional mean maximum degree. Section 6 presents similar technical preparations for the weak and strong disorder regime, which again use point process convergence and concentration, and martingale techniques, respectively. Finally, the main results related to the maximum degree are proved in Section 7.

Appendix 6B: Statement of Authorship

<table>
<tr><td colspan="4">This declaration concerns the article entitled:</td></tr>
<tr><td colspan="4">A phase transition for preferential attachment models with additive fitness</td></tr>
<tr><td colspan="4">Publication status (tick one)
Draft manuscript [ ] Submitted [ ] In review [ ] Accepted [ ] Published [x]</td></tr>
<tr><td>Publication details (reference)</td><td colspan="3">Journal: Electronic Journal of Probability, DOI: 10.1214/20-EJP550

Authors: Bas Lodewijks, Marcel Ortgiese</td></tr>
<tr><td colspan="4">Copyright status (tick the appropriate statement)
I hold the copyright for this material [x] Copyright is retained by the publisher, but I have been given permission to replicate the material here [ ]</td></tr>
<tr><td>Candidate's contribution to the paper (provide details, and also indicate as a percentage)</td><td colspan="3">The candidate predominantly executed the

Formulation of ideas:

60%. The candidate drove the formulation of ideas to go into this paper for a large part after the start of the PhD.

Design of methodology:

80%. The candidate has largely contributed to the development of the theoretical methodology in this paper, working out most proofs in detail to a large extent.

Experimental work:

N/A

Presentation of data in journal format:

80%. The candidate has written the initial draft, after which adjustments and changes were proposed by the other author.</td></tr>
<tr><td>Statement from Candidate</td><td colspan="3">This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.</td></tr>
<tr><td>Signed</td><td>Bas Lodewijks</td><td>Date</td><td>01/07/2021</td></tr>
</table>

Electronic Journal of Probability

Electron. J. Probab. **0** (2012), article no. 0, 1–54.
ISSN: 1083-6489 https://doi.org/vVOL-PID

A phase transition for preferential attachment models with additive fitness

Bas Lodewijks[*] Marcel Ortgiese[*]

Abstract

Preferential attachment models form a popular class of growing networks, where incoming vertices are preferably connected to vertices with high degree. We consider a variant of this process, where vertices are equipped with a random initial fitness representing initial inhomogeneities among vertices and the fitness influences the attractiveness of a vertex in an additive way. We consider a heavy-tailed fitness distribution and show that the model exhibits a phase transition depending on the tail exponent of the fitness distribution. In the weak disorder regime, one of the old vertices has maximal degree irrespective of fitness, while for strong disorder the vertex with maximal degree has to satisfy the right balance between fitness and age. Our methods use martingale methods to show concentration of degree evolutions as well as extreme value theory to control the fitness landscape.

Keywords: Network models; preferential attachment model; additive fitness; scale-free property; maximum degree.
MSC2020 subject classifications: Primary: 05C80 Secondary: 60G42.
Submitted to EJP on March 20, 2020, final version accepted on November 11, 2020.
Supersedes arXiv:2002.12863.

1 Introduction

A distinctive feature of real-world networks is their inhomogeneity, characterized in particular through the presence of hubs. These are nodes with a number of connections that greatly exceeds the average and thus have a great impact on the overall network topology. The existence of hubs in a network is closely linked to the *scale-free property*, that is, the proportion of nodes in the network with degree (number of connections) k scales as a power law $k^{-\tau}$ for some $\tau > 1$.

Preferential attachment models, as popularized by Barabási and Albert [2], form a class of random graphs that shows this behaviour 'naturally', that is, as a result of the dynamics and not because it is imposed otherwise, see also [6] for a first mathematical

[*]University of Bath, United Kingdom.
E-mail: b.lodewijks@bath.ac.uk,m.ortgiese@bath.ac.uk

derivation of this fact. In these evolving random graph models new vertices are introduced to the network over time and they connect to earlier introduced vertices with a probability proportional to their degree. This leads to the so-called *rich-get-richer* effect, which means that vertices with a high degree are more likely to increase their degree. It is exactly this effect that yields the power-law degree distributions and the existence of hubs in the graph.

The study of the emergence of hubs in random graph models such as the preferential model is often focused on the behaviour of the maximum degree in the graph. Móri first showed that for the Barabási-Albert model the maximum degree is of the same order as the degree of the first vertex [20], which was later generalised by Athreya *et al.* to affine preferential attachment models (with random out-degree) and to a larger class of preferential attachment models by Bhamidi in [1] and [3], respectively. A consequence of the way in which preferential attachment graphs evolve, is that the rich-get-richer effect should really be interpreted as an *old-get-richer* effect: it is the old vertices, who are introduced at the beginning of the evolution of the graph, that are able to attract the most connections [15].

However, when compared to real-life networks, it is clearly desirable to have a model where younger vertices can compete with the old ones. One way to achieve this is by assigning to each vertex a random fitness representing its intrinsic attractiveness and then to let the connection probability of a newly incoming vertex be proportional to either the product of the fitness and degree or the sum. These two models were introduced by Barabási and Bianconi in [4] and Ergün and Rodgers in [13], respectively.

Most previous results on preferential attachment models with fitness deal with the multiplicative case for bounded fitness. One of the reasons is that under certain conditions on the fitness distribution, these models exhibit the phenomenon of condensation, where a positive proportion of incoming vertices connects to vertices with fitness closer and closer to the maximal fitness in the system. This phenomenon was first shown in the mathematical literature in [7], later extended in [12] for a wide range of models, by looking at the empirical fitness and degree distribution. A full dynamic description of the condensation is a challenging problem, however see [9] for a very detailed analysis in a slightly modified model. [10] considers a continuous-time embedding of the process into a reinforced branching process, which allows them to control the maximal degree (in the continuous-time setting), which in the non-condensation case can be translated back to the random graph model. Also, under certain assumption on the fitness distribution, they show that condensation is non-extensive in the sense that there is not a single vertex that acquires a positive fraction of the incoming edges. These results are extended by [19] to a larger class of (bounded) fitness distributions (as a special case of a more general set-up).

Here, we consider the model with additive fitness, where a vertex is chosen with probability proportional to the sum of its degree and its intrinsic fitness. Both models of multiplicative and additive fitness can be seen as a way to understand how a random perturbation of the attractiveness of a vertex (due to natural inhomogeneities in the system) changes the behaviour of a standard preferential attachment model. As we have just discussed, in the multiplicative model we observe condensation which is quite a drastic change of the behaviour. This effect is already present for certain cases of bounded fitness, due to the fact that a small perturbation can have a large effect when multiplied by a large degree. For the additive model, we see that the change in behaviour due to random perturbations is of a very different nature. Indeed, the effect of large fitness values is not as immediate and it turns out that we need larger variability in the fitness values (and in particular we need to assume unbounded fitness) to observe a qualitative different behaviour compared to the standard model.

To best of our knowledge the only mathematical result have been [3] and [23], who confirmed the non-rigorous results in [13]. [3] showed that when the fitness is bounded, the degree distribution follows a power law with the same exponent as for the model with an additive constant equal to the expected value of the fitness. Moreover, [3] gives the asymptotics for the maximum degree and shows that it agrees again with the asymptotics for the model with additive constant. [23] considers the case of a deterministic additive sequence and shows that there is an equivalence between the preferential attachment (tree) model and a weighted recursive tree. From this, the author deduces ℓ^p-convergence of the renormalized degree sequence under a growth condition on the additive sequence. Furthermore, he considers geometric properties of the weighted recursive trees. Somewhat related is a model of preferential attachment with random (possibly heavy-tailed) initial degree, for which [8] show convergence of empirical fitness distributions, but the structure of these networks is very different from the additive fitness case due to large out-degrees.

In our work we consider the case of unbounded fitness and show that when the fitness distribution follows a power law, a more complex phase diagram arises. Our first result shows convergence for the empirical degree and fitness distribution. From this we can in particular deduce that if the fitness distribution is sufficiently light-tailed, then we are in a *weak disorder regime*, where the same result as in [3] still holds for both tail exponent of the degree distribution and the asymptotics of the maximum. However, if the tail exponent of the fitness distribution is sufficiently small (but so that the fitness still has a finite first moment), then we are in a *strong disorder regime*, where the tail exponent of the degree distribution is the same as for the fitness distribution. Moreover, the maximal degree grows of the same order as the largest fitness in the system. However, the vertex that maximizes the degree has to satisfy a delicate balance between arriving early and having a large fitness. In the limit this competition is expressed as an optimization of a functional of a Poisson point process.

Finally, we can also consider the *extreme disorder regime* when the fitness does not have a finite first moment. In that case, we show that a uniformly selected vertex has in-degree zero with high probability. Moreover, the maximal degree now scales as order n and the maximising vertex again satisfies the right balance between arriving early and large fitness. We note that our results for the degree distribution improve on those by Ergün and Rodgers [13], where these different regimes are overlooked and only the weak disorder regime is covered.

Our proof for the empirical degree/fitness distributions uses a stochastic approximation argument, which was also used in [12] for the multiplicative case. The analysis of the maximal degree is split into two steps: First we show concentration of the degrees when compared to the expected degree (conditionally on the fitness values) adapting the martingale arguments of Móri [20] (see also [15] for an exposition with more general attachment rules). For the weak disorder case, similar arguments as in [15] are sufficient to control the maximal degree. However, in the strong and extreme disorder case, we have to control the conditional expectation of the degrees, which are a function of the fitness only. We then show that these functionals simplify and converge to a functional of a Poisson point process, so that with the concentration we can deduce convergence of the maximal degree. Finally, our analysis is robust and covers essentially three variants of preferential attachment models: a model with possibly random out-degree as in [11] (and at most one edge between vertices) and two variations where the out-degree of each new vertex is fixed and then the connection probabilities are either updated after each edge is drawn or are kept fixed.

Notation. Throughout we use the following notation. We let $\mathbb{N} = \{1, 2, 3, \ldots\}$ be the natural numbers, we write $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$ if we want to include 0 and let

$[n] := \{1, \dots, n\}$. Moreover, for any sequence a_n and b_n of positive real numbers, we say $a_n = \Theta(b_n)$ if there exists a constant $C > 0$ such that $a_n \le C b_n$ and $b_n \le C a_n$. Moreover, we say $a_n \sim b_n$ if $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$. Also, we use the conditional probability measure $\mathbb{P}_{\mathcal{F}}(\cdot) := \mathbb{P}(\cdot \mid (\mathcal{F}_i)_{i\in\mathbb{N}})$ and expectation $\mathbb{E}_{\mathcal{F}}[\cdot] := \mathbb{E}[\cdot \mid (\mathcal{F}_i)_{i\in\mathbb{N}}]$.

2 Definitions and main results

The preferential attachment model is an evolving random graph model, where vertices are added to the graph consecutively and then connected to older vertices. We denote by $\mathcal{G}_n$ the resulting directed graph at the stage when the vertex set is $[n]$. Moreover, we take edges to be directed from the vertex with high index to the one with lower index. Throughout, we use the following notation,

$$\mathcal{Z}_n(i) := \text{in-degree of vertex } i \text{ in } \mathcal{G}_n.$$

We now introduce three different preferential attachment with fitness models (PAF), the first one which allows for a random out-degree in the spirit of Dereich and Mörters [11], the second one where the out-degree of a new vertex is fixed and we connect edges while keeping the degrees fixed, and the last one with a fixed out-degree, but where we update degrees in between connections (where the latter is the fitness modification of a model closer to [6]).

Definition 2.1 (Preferential attachment with fitness)**.** *Let $(\mathcal{F}_i)_{i\ge1}$ be a sequence of i.i.d. copies of a random variable $\mathcal{F}$ taking values in $(0,\infty)$ with distribution μ. For any $n \in \mathbb{N}$, define*

$$S_n := \sum_{i=1}^{n} \mathcal{F}_i. \tag{2.1}$$

Let $n_0, m_0 \in \mathbb{N}$. We say that a sequence of random graphs $(\mathcal{G}_n)_{n\ge n_0}$ is a preferential attachment model with (additive) fitness *if $\mathcal{G}_n$ is a directed and weighted graph on the vertex set $[n]$ with edges directed from larger to smaller indices. Moreover, we assume that $\mathcal{G}_{n_0}$ has m_0 edges and we assign fitness values $\mathcal{F}_1, \mathcal{F}_2, \dots, \mathcal{F}_{n_0}$ to the vertices $1, 2, \dots, n_0$ respectively.*

To obtain $\mathcal{G}_{n+1}$ from $\mathcal{G}_n$ for some $n \ge n_0$, add vertex $n+1$ to the vertex set and attach fitness $\mathcal{F}_{n+1}$ to $n+1$. Furthermore, we assume that the updating rule satisfies one of the following three assumptions for some fixed $m \in \mathbb{N}$:

(PAFRO) Preferential attachment with fitness and random out-degree. *Conditionally on $\mathcal{G}_n$, vertex $n+1$ is connected to each vertex in $[n]$ by at most one edge and the probability to connect to a given $i \in [n]$ is*

$$\frac{\mathcal{Z}_n(i) + \mathcal{F}_i}{m_0 + (n - n_0) + S_n}. \tag{2.2}$$

Furthermore, conditionally on $\mathcal{G}_n$ the degree increments $(\Delta\mathcal{Z}_n(i) := \mathcal{Z}_{n+1}(i) - \mathcal{Z}_n(i), i \in [n])$ are pairwise non-positively correlated.

(PAFFD) Preferential attachment with fitness and fixed degree. *To vertex $n+1$ we assign m half-edges. Conditionally on $\mathcal{G}_n$, connect each half-edge independently to some $i \in [n]$ with probability*

$$\frac{\mathcal{Z}_n(i) + \mathcal{F}_i}{m_0 + m(n - n_0) + S_n}.$$

(PAFUD) Preferential attachment with fitness and updating degree. *To vertex $n+1$ we assign m half-edges. Let $\mathcal{Z}_{n,j}(i)$ denote the in-degree of vertex i when $n+1$ has attached j of its half-edges, $j = 1, \dots, m$. For $j = 1, \dots, m$, conditionally on the graph of size n including the first $j-1$ half-edges from $n+1$, connect the j^{th} half-edge to $i \in [n]$ with probability*

$$\frac{\mathcal{Z}_{n,j-1}(i) + \mathcal{F}_i}{m_0 + m(n-n_0) + (j-1) + S_n}.$$

Remark 2.2. The quantity in (2.2) is always less than 1, since $\sum_{i=1}^{n_0} \mathcal{Z}_{n_0}(i) = m_0$ and at each step $\mathcal{Z}_n(i)$ increases by at most one. Note also that for the PAFRO model, the exact distribution of $(\Delta\mathcal{Z}_n(i), i \in [n])$ is not specified. For example, for $m = 1$, the PAFFD and the PAFUD model are identical and both satisfy PAFRO. Another possibility is to consider a model with a random out-degree, where $(\Delta\mathcal{Z}_n(i), i \in [n])$ is a vector of independent Bernoulli variables with success probability as given in (2.2).

We have defined our random graph model for an arbitrary fitness distribution. However, for the analysis the most interesting case occurs when we are dealing with heavy-tailed distributions. In this case the fitness can have a significant effect on the behaviour of the system as a whole, whereas the 'fitness effect' is smoothed out when its tail behaviour is too light. In the latter case, one sees no differences in the mean-field behaviour when changing from a deterministic, fixed fitness to random i.i.d. fitness values. Therefore, in the following, we frequently consider the following assumption:

Assumption 2.3. *The fitness distribution is a power law with exponent $\alpha > 1$, i.e.*

$$\mathbb{P}(\mathcal{F} \geq x) = \mu(x, \infty) = \ell(x) x^{-(\alpha-1)}, \quad \textit{for } x > 0,$$

where ℓ is a slowly-varying function at infinity, i.e. for all $c > 0$ $\lim_{x\to\infty} \ell(cx)/\ell(x) = 1$.

We continue by stating our first main result. We define the following measures,

$$\Gamma_n := \frac{1}{n}\sum_{i=1}^{n} \mathcal{Z}_n(i)\delta_{\mathcal{F}_i}, \qquad \Gamma_n^{(k)} := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{\mathcal{Z}_n(i)=k\}}\delta_{\mathcal{F}_i}, \qquad p_n(k) := \Gamma_n^{(k)}([0,\infty)), \tag{2.3}$$

which correspond to the the empirical fitness distribution of a vertex sampled with weight given by its in-degree, then the joint empirical fitness-in-degree distribution and finally the empirical degree distribution.

Theorem 2.4 (Degree distributions in PAF models)**.** *Consider the three PAF models as in Definition 2.1 and suppose the fitness satisfies $\mathbb{E}[\mathcal{F}] < \infty$. Let $\theta_m := 1 + \mathbb{E}[\mathcal{F}]/m$. Then, almost surely, for any $k \in \mathbb{N}_0$, as $n \to \infty$,*

$$\Gamma_n \to \Gamma, \quad \Gamma_n^{(k)} \to \Gamma^{(k)}, \quad \textit{and} \quad p_n(k) \to p(k), \tag{2.4}$$

where the first two statements hold with respect to the topology of weak convergence and the limits are given as

$$\Gamma(\mathrm{d}x) = \frac{x}{\theta_m - 1}\mu(\mathrm{d}x), \qquad \Gamma^{(k)}(\mathrm{d}x) = \frac{\theta_m}{x+\theta_m}\prod_{\ell=1}^{k}\frac{(\ell-1)+x}{\ell+x+\theta_m}\mu(\mathrm{d}x), \tag{2.5}$$

and

$$p(k) = \int_0^\infty \frac{\theta_m}{x+\theta_m}\prod_{\ell=1}^{k}\frac{(\ell-1)+x}{\ell+x+\theta_m}\mu(\mathrm{d}x). \tag{2.6}$$

Remark 2.5. Throughout this article we work with Definition 2.1. However, Theorem 2.4 also holds under the following slightly weaker conditions. Set

$$\bar{\mathcal{F}}_n := \frac{1}{n}\sum_{i=1}^{n}(\mathcal{Z}_n(i) + \mathcal{F}_i),$$

and define the degree increment at step $n+1$ of vertex i by $\Delta\mathcal{Z}_n(i) := \mathcal{Z}_{n+1}(i) - \mathcal{Z}_n(i)$. We assume the graph $\mathcal{G}_{n_0}$ is given deterministically such that $m_0 := \sum_{i\in[n_0]} \mathcal{Z}_{n_0}(i) \geq 1$. Furthermore, we assume for $n \geq n_0$,

(A1) $\mathbb{E}[\Delta\mathcal{Z}_n(i) \mid \mathcal{G}_n] = (\mathcal{Z}_n(i) + \mathcal{F}_i)/(n\bar{\mathcal{F}}_n)\mathbb{1}_{\{i\leq n\}}$.

(A2) $\exists\, C_{\mathrm{var}} > 0 :\ \mathrm{Var}(\Delta\mathcal{Z}_n(i) \mid \mathcal{G}_n) \leq C_{\mathrm{var}}\mathbb{E}[\Delta\mathcal{Z}_n(i) \mid \mathcal{G}_n]$.

(A3) $\sup_{i=1,\ldots,n} n\big|\mathbb{P}(\Delta\mathcal{Z}_n(i) = 1 \mid \mathcal{G}_n) - \mathbb{E}[\Delta\mathcal{Z}_n(i) \mid \mathcal{G}_n]\big| \xrightarrow{a.s.} 0$.

(A4) Conditionally on $\mathcal{G}_n$, $\{\Delta\mathcal{Z}_n(i)\}_{i\in[n]}$ is negatively quadrant dependent in the sense that for any $i \neq j$ and $k, l \in \mathbb{Z}^+$,

$$\mathbb{P}(\Delta\mathcal{Z}_n(i) \leq k, \Delta\mathcal{Z}_n(j) \leq l \mid \mathcal{G}_n) \leq \mathbb{P}(\Delta\mathcal{Z}_n(i) \leq k \mid \mathcal{G}_n)\,\mathbb{P}(\Delta\mathcal{Z}_n(j) \leq l \mid \mathcal{G}_n)\,. \quad (2.7)$$

As can be seen from the proof, Theorem 2.4 holds for any evolving random graph model that satisfies these assumptions. See also Lemma 4.3 below, where we show that the PAFFD and the PAFUD model satisfy the negative quadrant dependency as in (A4).

By comparing with the case where the fitness is constant, we can interpret Theorem 2.4 such that the degree of a typical vertex can be found via a two-step process, where first the fitness is chosen according to μ and then the degree evolves as in the case with an additive constant equal to the fitness.

However, while at first our result looks similar to the constant fitness case, by looking at the tail exponent of the degree distribution we can see that this is only the case when the fitness is not too heavy-tailed. Indeed, suppose that the fitness distribution follows a power law, then we can distinguish three different regimes. As the next theorem shows, if the fitness distribution has finite moments of order $\theta_m = 1 + \mathbb{E}[\mathcal{F}]/m$, then the degree distribution has power law exponent $1 + \theta_m$, which is the same as in the model with constant fitness equal to $\mathbb{E}[\mathcal{F}]$. Using the terminology used in the field of random media, we refer to this situation as the *weak disorder regime*. However, if the fitness distribution is more heavy-tailed, but still with finite first moment, then the degree distribution follows the same power law as the fitness distribution, a situation which we refer to as the *strong disorder regime*. Finally, we can also consider the *extreme disorder* case when the fitness distribution does not have a finite first moment. In this case we show that with high probability, a uniformly chosen vertex has not received any incoming edges (since most connections are made to vertices with very high fitness).

Theorem 2.6. *Suppose* $p(k), k \in \mathbb{N}_0$, *is as in* (2.6) *and* $\theta_m = 1 + \mathbb{E}[\mathcal{F}]/m$.

(i) Weak disorder. *If* $\mathbb{E}[\mathcal{F}^{\theta_m}] < \infty$, *then for* $k \to \infty$,

$$p(k) \sim Ck^{-(1+\theta_m)}, \quad \textit{where } C := \theta_m \int_0^\infty \frac{\Gamma(x+\theta_m)}{\Gamma(x)}\mu(\mathrm{d}x),$$

and where Γ *is the Gamma function.*

(ii) Strong disorder. *Suppose* $\mathcal{F}$ *has a power law distribution as in Assumption 2.3. Then, if* $\alpha = 1 + \theta_m$ *and* $\mathbb{E}[\mathcal{F}^{\theta_m}] = \infty$, *we have as* $k \to \infty$

$$p(k) = \Theta(\ell^\star(k)k^{-(1+\theta_m)}),$$

where $\ell^\star(k) := \int_1^k \ell(x)/x \,\mathrm{d}x$.

If $\alpha \in (2, 1+\theta_m)$*, then as* $k \to \infty$*,*

$$p(k) = \Theta(\ell(k)k^{-\alpha}).$$

(iii) Extreme disorder. *Suppose* $\mathcal{F}$ *has a power law distribution as in Assumption 2.3 with* $\alpha \in (1,2)$ *and consider the three PAF models as in Definition 2.1. Let* U_n *be a uniformly chosen vertex in* $\mathcal{G}_n$*, let* $\varepsilon > 0$ *and let* $E_n := \{\mathcal{Z}_n(U_n) = \mathcal{Z}_{n_0}(U_n)\}$*, be the event that* U_n *has not increased its degree with respect to the initialisation* $\mathcal{G}_{n_0}$*. Then, for* n *sufficiently large,*

$$\mathbb{P}(E_n) \geq 1 - Cn^{-((2-\alpha)\wedge(\alpha-1))/\alpha+\varepsilon},$$

for some constant $C > 0$*.*

Our next main result provides a more detailed analysis of the dynamic behaviour of the system by describing the asymptotics of the maximal degree. As might be expected from the different phases observed for the tail of the degree distribution, there are also three distinct phases for the maximal degree. Again under the assumption that the fitness has a power law, we observe that in the *weak disorder regime*, where the fitness has relatively light tails, the vertex with maximal degree is one of the old vertices, similar to the system with constant fitness. This first result (parts (i) and (iii) in the theorem below) in the special case of the PAFUD/PAFFD model with $m = 1$ is also contained in [23].

However, if the fitness is more heavy-tailed (but still with finite first moment), i.e. in the *strong disorder regime*, the maximal degree grows at the same rate as the maximal fitness in the system (i.e. approximately like $n^{1/(\alpha-1)}$). In this case, the maximal degree satisfies a delicate balance between arriving early enough and having large fitness. Finally, in the *extreme disorder regime*, where the fitness does not have a first moment, the maximal degree grows of order n, again satisfying a non-trivial optimisation between large fitness value and arriving early. The main difference compared to the strong disorder regime is that now the sum of the fitness values in the normalization, e.g. in (2.2), is random to first order and depends on the extreme values of the fitness landscape. As is common in extreme value theory, the limiting variables in the strong and extreme disorder regime are described in terms of a functional of a Poisson point process capturing the extremes of the fitness (in competition with the advantage of arriving early).

Theorem 2.7 ((Maximum) degree behaviour in PAFs)**.** *Consider the three PAF models as in Definition 2.1. First, the following results hold for fixed degrees:*

(i) Suppose $\mathbb{E}[\mathcal{F}^{1+\varepsilon}] < \infty$ *for some* $\varepsilon > 0$*, then for all fixed* $i \in \mathbb{N}$*,*

$$\mathcal{Z}_n(i)n^{-1/\theta_m} \xrightarrow{a.s.} \xi_i, \tag{2.8}$$

where ξ_i *is an almost surely finite random variable with no atom at* 0 *and* $\theta_m := 1 + \mathbb{E}[\mathcal{F}]/m$*.*

(ii) When the fitness distribution satisfies Assumption 2.3 with $\alpha \in (1,2)$*, for all fixed* $i \in \mathbb{N}$*,*

$$\mathcal{Z}_n(i) \xrightarrow{a.s.} \mathcal{Z}_\infty(i), \tag{2.9}$$

for some almost surely finite random variable $\mathcal{Z}_\infty(i)$*.*

In the following let $I_n := \arg\max_{i\in[n]} \mathcal{Z}_n(i)$ *(resolving any ties by taking the smaller index).*

(iii) Weak disorder: *If* $\mathbb{E}[\mathcal{F}^{\theta_m+\varepsilon}] < \infty$ *for some* $\varepsilon > 0$*, then we have*

$$I_n \xrightarrow{a.s.} I, \qquad \max_{i\in[n]} \mathcal{Z}_n(i) n^{-1/\theta_m} \xrightarrow{a.s.} \sup_{i\geq 1} \xi_i, \tag{2.10}$$

for some almost surely finite random variable I*.*

Additionally, assume that the fitness distribution is a power law with parameter α *as in Assumption 2.3 and define* $u_n := \sup\{t \in \mathbb{R} : \mathbb{P}(\mathcal{F} \geq t) \geq 1/n\}$*. Let* Π *be a Poisson point process on* $(0,1)\times(0,\infty)$ *with intensity measure* $\nu(\mathrm{d}t, \mathrm{d}x) := \mathrm{d}t \times (\alpha-1)x^{-\alpha}\mathrm{d}x$*. Then, the following results hold:*

(iv) Strong disorder: *When* $\alpha \in (2, 1+\theta_m)$*,*

$$(I_n/n, \max_{i\in[n]} \mathcal{Z}_n(i)/u_n) \xrightarrow{d} (I, \sup_{(t,f)\in\Pi} f(t^{-1/\theta_m} - 1)), \tag{2.11}$$

where $I \overset{d}{=} B^{\theta_m}$*, with* $B \sim \mathrm{Beta}(\theta_m - (\alpha-1), \alpha)$ *and where* $\max_{(t,f)\in\Pi} f(t^{-1/\theta_m} - 1)$ *has a Fréchet distribution with shape parameter* $\alpha - 1$ *and scale parameter* $(\Gamma(\theta_m - (\alpha-1))\Gamma(\alpha)/\Gamma(\theta_m))^{1/(\alpha-1)}$*, where* Γ *is the Gamma function.*

(v) Extreme disorder: *When* $\alpha \in (1,2)$*, let* Π *be a Poisson point process on* $E := (0,1)\times(0,\infty)$ *with intensity measure* $\nu(\mathrm{d}t, \mathrm{d}x) := \mathrm{d}t \times (\alpha-1)x^{-\alpha}\mathrm{d}x$*. Then,*

$$(I_n/n, \max_{i\in[n]} \mathcal{Z}_n(i)/n) \xrightarrow{d} \left(I', m \sup_{(t,f)\in\Pi} f \int_t^1 \left(\int_E g \mathbb{1}_{\{u\leq s\}} \mathrm{d}\Pi(u,g)\right)^{-1} \mathrm{d}s\right), \tag{2.12}$$

for some random variable I' *with values in* $(0,1)$*.*

3 Overview of the proofs

In this section, we give a short overview of the proofs of the main theorems and the structure of the remaining paper.

In Section 4 we prove Theorems 2.4 and 2.6. To prove Theorem 2.4, we use the theory of stochastic approximation in a similar setup as in [12], where it was used for models with multiplicative fitness.

The main idea is to consider, for $0 \leq f < f' < \infty$, the quantities

$$\Gamma_n((f,f']) = \frac{1}{n}\sum_{i=1}^n \mathcal{Z}_n(i)\mathbb{1}_{\{\mathcal{F}_i\in(f,f']\}}, \qquad \Gamma_n^{(k)}((f,f']) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{\mathcal{Z}_n(i)=k, \mathcal{F}_i\in(f,f']\}}, \quad k \geq 0,$$

where $0 < f < f' < \infty$. Then, by considering the conditional increment and using the preferential attachment dynamics, we show that

$$\Gamma_{n+1}((f,f']) - \Gamma_n((f,f']) \leq \frac{1}{n+1}(A_n - B_n\Gamma_n((f,f'])) + (R_{n+1} - R_n),$$

and also a similar lower bound with slightly different sequences A_n, B_n. This should be interpreted as a time-discretisation of a differential inequality. Then, a basic stochastic approximation argument (see also Lemma 4.1 below) shows that if A_n, B_n and R_n converge almost surely, then we obtain an upper bound on the $\limsup$ of $\Gamma_n((f,f'])$ (and similarly a lower bound). By an approximation argument this yields convergence of Γ_n. We obtain similar bounds for $\Gamma_n^{(k)}((f,f'])$ (involving $\Gamma_n^{(k-1)}((f,f'])$) so that with an induction argument we also can deduce convergence of $\Gamma_n^{(k)}$.

In the last part of Section 4 we prove Theorem 2.6 using standard arguments.

The remainder of the paper deals with the asymptotics of the degree of a fixed vertex, as well as the maximal degree, as stated in Theorem 2.7. In the following we only discuss the proof for the PAFUD model, but the proofs for the PAFRO model and PAFFD model follow with minor modifications.

A central tool in the analysis of the degree evolutions is the following martingale introduced by [20] in the context of classical preferential attachment (see also [15]). For $k \geq -\min\{\mathcal{F}_i, 1\}$, define a sequence

$$M_n^k(i) := c_n^k \binom{\mathcal{Z}_n(i) + \mathcal{F}_i + (k-1)}{k}, \tag{3.1}$$

where c_n^k is a carefully chosen normalisation sequence and

$$\binom{a}{b} = \frac{\Gamma(a+1)}{\Gamma(b+1)\Gamma(a-b+1)}, \quad \text{for } a, b > -1 \text{ such that } a - b > -1,$$

is the generalized binomial coefficient defined in terms of the Gamma function Γ. Next, we write

$$\mathbb{P}_{\mathcal{F}} \quad \text{and} \quad \mathbb{E}_{\mathcal{F}}$$

for the (regular) conditional probability measure (and its expectation respectively) when conditioning on the fitness values $\mathcal{F}_1, \mathcal{F}_2, \ldots$. Then, as for the standard preferential model, one can show that $(M_n^k(i), n \geq i)$ is a martingale under the conditional measure $\mathbb{P}_{\mathcal{F}}$.

Note that (3.1) with $k = 1$ gives,

$$\mathcal{Z}_n(i) = (c_n^1)^{-1} M_n^1(i) - \mathcal{F}_i,$$

and $M_n^1(i)$ converges as it is a non-negative martingale. So for fixed i, the leading order is determined by c_n^1. Indeed, we see that

$$c_n^k \approx \prod_{j=1}^{n-1} \Big(1 - \frac{k}{mj + S_j}\Big)^m \approx \exp\Big\{ -\sum_{j=1}^{n-1} \frac{k}{j + S_j/m} \Big\}, \tag{3.2}$$

where $S_j = \sum_{\ell=1}^j \mathcal{F}_\ell$. In Lemma 6.4, we prove that if $\mathbb{E}[\mathcal{F}] < \infty$, then by the law of large numbers the sequence c_n^k rescaled by n^{k/θ_m} converges almost surely. Moreover, if $\alpha \in (1, 2)$ (for a power law fitness distribution), then c_n^k converges almost surely without rescaling. This proves the first two statements (2.8) and (2.9) of Theorem 2.7.

To prove the statements about the maximal degree, we first consider the conditional expectation of $\mathcal{Z}_n(i)$ which, using the martingale $M_n^1(i)$, can be written as

$$\mathbb{E}_{\mathcal{F}}[\mathcal{Z}_n(i)] = \mathcal{F}_i \Big(\frac{c_i^1}{c_n^1} - 1 \Big), \tag{3.3}$$

at least for $i > n_0$, otherwise a small correction is necessary. From this point, the proofs in the three different regimes deviate from each other.

First, if we assume that $\mathbb{E}[\mathcal{F}] < \infty$, then by (3.3) and the asymptotics of c_n^1 in (3.2) we can deduce that

$$\mathbb{E}_{\mathcal{F}}[\mathcal{Z}_n(i)] \approx \mathcal{F}_i \Big(\Big(\frac{n}{i}\Big)^{1/\theta_m} - 1 \Big). \tag{3.4}$$

Now, suppose that $\mathbb{E}\big[\mathcal{F}^{\theta_m + \varepsilon}\big] < \infty$ for some $\varepsilon > 0$. Then, in Lemma 6.6, we show that

$$\lim_{i \to \infty} \sup_{n \geq n_0 \vee i} M_n^1(i) = 0.$$

Intuitively, this follows from (3.4), since under the assumption that $\mathbb{E}\big[\mathcal{F}^{\theta_m+\varepsilon}\big] < \infty$ for some $\varepsilon > 0$, the maximum of the fitness values satisfies $\max_{i\in[n]} \mathcal{F}_i = o(n^{1/\theta_m})$ (with high probability), so that the term $(\frac{n}{i})^{1/\theta_m}$ dominates for i small. We then use the following result for triangular arrays $a_{i,n}$, $1 \le i \le n$,: if $\lim_{n\to\infty} a_{i,n} = a_i$ for all $i \in \mathbb{N}$, and if $\sup_{n\ge i} a_{i,n} = b_i$ and $\lim_{i\to\infty} b_i = 0$, we obtain $\lim_{n\to\infty} \max_{i\in[n]} a_{i,n} = \sup_{i\ge 1} a_i$. Using this result on $a_{i,n} = c_n^1(\mathcal{Z}_n(i) + \mathcal{F}_i) = M_n^1(i)$ yields, together with (3.4), the weak disorder result in (2.10).

Next, we consider the *strong disorder regime*, where the fitness distribution is a power law with parameter α with $\alpha \in (2, 1+\theta_m)$. Extreme value theory tell us that in this case $\max_{i\in[n]} \mathcal{F}_i \approx n^{1/(\alpha-1)}$ so that (3.4) suggests that in this regime vertices with high fitness have a chance to compete with the old vertices. To capture the asymptotics of the peaks of the fitness landscape more precisely, we consider the point process

$$\Pi_n := \sum_{i=1}^{n} \delta_{(i/n, \mathcal{F}_i/u_n)}, \tag{3.5}$$

where $u_n := \sup\{t \ge 0 \,:\, \mathbb{P}(\mathcal{F} \ge t) \ge 1/n\}$. Then, classical extreme value theory (see e.g. [21, Corollary 4.19]) tells us that

$$\Pi_n \Rightarrow \Pi,$$

in the vague topology, where Π is a Poisson point process on $(0,1)\times(0,\infty)$ with intensity measure $\nu(\mathrm{d}t, \mathrm{d}x) := \mathrm{d}t \times (\alpha-1)x^{-\alpha}\mathrm{d}x$ (see also Section 5 below for more details). From this convergence, we can then deduce using (3.4) that

$$\max_{i\in[n]} \mathbb{E}_{\mathcal{F}}[\mathcal{Z}_n(i)/u_n] \xrightarrow{d} \sup_{(t,f)\in\Pi} f(t^{-1/\theta_m} - 1),$$

see the first part of Proposition 6.1 for details. A non-trivial part of the proof is showing that the approximation in (3.4) works sufficiently well for the relevant range of i. The proof of Theorem 2.7 is then completed by showing concentration of $\mathcal{Z}_n(i)$ around its conditional mean, so that

$$\max_{i\in[n]} \mathcal{Z}_n(i)/u_n - \max_{i\in[n]} \mathbb{E}_{\mathcal{F}}[\mathcal{Z}_n(i)/u_n] \xrightarrow{\mathbb{P}} 0.$$

The concentration argument relies on the martingale $M_n^k(i)$ for carefully chosen k (which correspond approximately to k^{th} moments of $\mathcal{Z}_n(i)$), see the first part of Proposition 6.2.

Finally, we consider the *extreme disorder regime*, where $\alpha \in (1,2)$ so that the fitness does not have finite first moments. In particular, the law of large numbers no longer applies to the sum $S_n = \sum_{i=1}^n \mathcal{F}_i$ appearing in the normalizing constant in the attachment probabilities. In this case, we obtain from (3.2) that, for i of order n,

$$\frac{c_i^1}{c_n^1} - 1 \approx \exp\left\{ m \sum_{j=i}^{n-1} \frac{1}{S_j} \right\} - 1 \approx m \sum_{j=i}^{n-1} \frac{1}{S_j}.$$

Then, it follows from (3.3) with the same Π_n as in (3.5) that

$$\begin{aligned} \frac{\mathbb{E}_{\mathcal{F}}[\mathcal{Z}_n(i)]}{n} &\approx m \frac{\mathcal{F}_i}{u_n} \Big(\frac{1}{n} \sum_{j=i}^{n} \frac{u_n}{S_j} \Big) \\ &= m \frac{\mathcal{F}_i}{u_n} \int_{i/n}^{1} \Big(\int_E f \mathbb{1}_{\{t\le s\}} \mathrm{d}\Pi_n(f,t) \Big)^{-1} \mathrm{d}s \\ &=: m \frac{\mathcal{F}_i}{u_n} T^{i/n}(\Pi_n), \end{aligned} \tag{3.6}$$

where $E := (0,1) \times (0,\infty)$. From this we can eventually deduce that

$$\max_{i\in[n]} \mathbb{E}_{\mathcal{F}}[\mathcal{Z}_n(i)/n] \xrightarrow{d} m \sup_{(t,f)\in\Pi} f \int_t^1 \left(\int_E g \mathbb{1}_{\{u\le s\}} \mathrm{d}\Pi(u,g) \right)^{-1} \mathrm{d}s.$$

Unfortunately, the corresponding functionals $T^{i/n}(\Pi_n)$ are not directly continuous in Π_n, so that the arguments involve careful cut-off arguments (see Section 5). The final step is to show concentration

$$\max_{i\in[n]} \mathcal{Z}_n(i)/n - \max_{i\in[n]} \mathbb{E}_{\mathcal{F}}[\mathcal{Z}_n(i)/n] \xrightarrow{\mathbb{P}} 0,$$

which again uses the martingale $M_n^1(i)$, but in this case is slightly easier than for $\alpha > 2$ due to the scaling factor n.

Overall, the proof of Theorem 2.7 is structured in the following way. In Section 5, we first show convergence of the functional $T^{i/n}(\Pi_n)$ introduced in (3.6). We also take the opportunity to recap some of the basics of convergence of point process convergence and we also carry out the technical cut-off arguments. In Section 6 we introduce the martingales $M_n^k(i)$ more formally and prove some of their properties. We use these properties to show concentration in all three regimes and the point process convergence in the strong disorder case. Finally, in Section 7 we prove Theorem 2.7 by gathering together all the necessary results from the previous two sections.

4 Degree and fitness distributions

This section is devoted to first proving Theorems 2.4 and 2.6, where the proof of the former theorem uses the ideas of stochastic approximation. Before proving Theorem 2.4, we introduce several preliminary lemmas. The first lemma, which is the main ingredient in the proof of Theorem 2.4, comes from [12, Lemma 3.1]:

Lemma 4.1. *Let $(X_n)_{n\ge 0}$ be a non-negative stochastic process. We suppose that the following estimate holds:*

$$X_{n+1} - X_n \le \frac{1}{n+1}(A_n - B_n X_n) + R_{n+1} - R_n, \quad \text{a.s.}$$

where

(i) $(A_n)_{n\ge0}$ and $(B_n)_{n\ge0}$ are almost surely convergent stochastic processes with deterministic limits $A, B > 0$.

(ii) $(R_n)_{n\ge0}$ is an almost surely convergent stochastic process.

Then, almost surely,

$$\limsup_{n\to\infty} X_n \le \frac{A}{B}.$$

Similarly, if instead, under the same conditions (i) and (ii),

$$X_{n+1} - X_n \ge \frac{1}{n+1}(A_n - B_n X_n) + R_{n+1} - R_n,$$

then almost surely,

$$\liminf_{n\to\infty} X_n \ge \frac{A}{B}.$$

In the next lemma, we discuss two specific examples of the stochastic process R_n as introduced in Lemma 4.1, which are used in the proof of Theorem 2.4:

Lemma 4.2. *Recall* Γ_n *and* $\Gamma_n^{(k)}$ *from* (2.3) *and let* $0 < f < f' < \infty, k \in \mathbb{N}_0$ *and assume the fitness distribution has a finite mean. We then have the two following results:*

(i) *Set* $X_n := \Gamma_n^{(k)}((f, f'])$*,* $\Delta R_n := X_{n+1} - \mathbb{E}[X_{n+1} \,|\, \mathcal{G}_n]$ *and* $R_n := \sum_{j=n_0}^{n} \Delta R_j$*. Then* R_n *converges almost surely.*

(ii) *Set* $X_n := \Gamma_n((f, f'])$*,* $\Delta R_n := X_{n+1} - \mathbb{E}[X_{n+1} \,|\, \mathcal{G}_n]$ *and* $R_n := \sum_{j=n_0}^{n} \Delta R_j$*. Then* R_n *converges almost surely.*

Before proving Lemma 4.2, we recall the concept of negative quadrant dependence (NQD) as introduced in (2.7). We note that the PAFRO model has been defined with an additional assumption of non-positively correlated degree increments. Note that, since the degree increments in this model are Bernoulli random variables, NQD is equivalent to non-positive correlation. For the PAFFD and PAFUD models, NQD follows directly from the definition of the model, as we show in the following lemma:

Lemma 4.3. *Recall the degree increments* $\Delta\mathcal{Z}_n(i) := \mathcal{Z}_{n+1}(i) - \mathcal{Z}_n(i)$*. For the PAFUD and PAFFD model, the* $(\Delta\mathcal{Z}_n(i))_{i\in[n]}$ *are negative quadrant dependent, in the sense of* (2.7).

Proof. The NQD of the PAFFD model directly follows from [17], as $(\Delta\mathcal{Z}_n(i))_{i\in[n]}$ forms a multinomial distribution, for which NQD is known. For the PAFUD model, $(\Delta\mathcal{Z}_n(i))_{i\in[n]}$ is a convolution of unlike multinomial distributions (the probabilities of the multinomial distribution change at each step/sampling). In the case that the change in the probabilities in independent of the previous samplings (where previous edges are attached), [17] provides a proof of NQD. However, in this case, the changes in the sampling probabilities are *dependent*, so that a more careful argument is required. Let us write $\Delta\mathcal{Z}_n(i) := X_1 + \ldots + X_m, \Delta\mathcal{Z}_n(j) := Z_1 + \ldots + Z_m$, where the X_k, Z_k are Bernoulli random variables which take value 1 if the k^{th} edge of vertex $n+1$ connects to i, j, respectively, $k \in [m]$. Since X_1, Z_1 are part of a multinomial vector with one trial, (2.7) holds for these random variables. Then, we investigate $X_1 + X_2, Z_1 + Z_2$, where we prove (2.7) for $X_1 + X_2, Z_1 + Z_2$, but with $\geq$ rather than $\leq$ in the event, which is an equivalent definition of NQD. We write, for $k, \ell \geq 0$,

$$\mathbb{P}(X_1 + X_2 \geq k, Z_1 + Z_2 \geq \ell \mid \mathcal{G}_n) = \mathbb{E}[\mathbb{P}(X_2 \geq k - X_1, Z_2 \geq \ell - Z_1 \mid \mathcal{G}_n, X_1, Z_1) \mid \mathcal{G}_n] \,.$$

Since, conditionally on $\mathcal{G}_n$ and (X_1, Z_1), the random variables (X_2, Z_2) are part of a multinomial vector with a single trial, the same argument we use for X_1, Z_1 gives the upper bound

$$\mathbb{E}[\mathbb{P}(X_2 \geq k - X_1 \mid \mathcal{G}_n, X_1, Z_1)\, \mathbb{P}(Z_2 \geq \ell - Z_1 \mid \mathcal{G}_n, X_1, Z_1) \mid \mathcal{G}_n] \,. \tag{4.1}$$

It follows from the definition of the PAFUD model that, conditionally on X_1, X_2 is independent of Z_1 and, conditionally on Z_1, Z_2 is independent of X_1. As the probabilities in (4.1) are increasing functions of X_1, Z_1, respectively, it follows from the definition of negative association in [17], which is equivalent to NQD, that

$$\begin{aligned} &\mathbb{E}[\mathbb{P}(X_2 \geq k - X_1 \mid \mathcal{G}_n, X_1, Z_1)\, \mathbb{P}(Z_2 \geq \ell - Z_1 \mid \mathcal{G}_n, X_1, Z_1) \mid \mathcal{G}_n] \\ &\quad \leq \mathbb{E}[\mathbb{P}(X_2 \geq k - X_1 \mid \mathcal{G}_n, X_1) \mid \mathcal{G}_n]\, \mathbb{E}[\mathbb{P}(Z_2 \geq \ell - Z_1 \mid \mathcal{G}_n, Z_1) \mid \mathcal{G}_n] \\ &\quad = \mathbb{P}(X_1 + X_2 \geq k \mid \mathcal{G}_n)\, \mathbb{P}(Z_1 + Z_2 \geq \ell \mid \mathcal{G}_n) \,. \end{aligned}$$

We can iterate the same argument to obtain the same inequality for the m terms in $\Delta\mathcal{Z}_n(i) = X_1 + \ldots + X_m, \Delta\mathcal{Z}_n(j) = Z_1 + \ldots + Z_m$. We then recall that this result is equivalent to (2.7), as required. $\square$

Proof of Lemma 4.2. First note that, in both cases, R_n is a zero-mean martingale with respect to $\mathcal{G}_n$. The convergence of R_n can be proved by showing its martingale increments $\Delta R_n = R_{n+1}-R_n$ have summable conditional second moments, or have summable second moments. Define, for $0<f<f'<\infty$, $\mathbb{I}_n := \{i\in[n]\,|\,\mathcal{F}_i\in(f,f']\}$. We first deal with case (i). We write ΔR_n as the difference of two martingales. For $k\geq 1$,

$$\Delta R_n = \frac{1}{n+1}\sum_{i\in\mathbb{I}_n}\big(\mathbb{1}_{\{\mathcal{Z}_{n+1}(i)=k\}} - \mathbb{P}(\mathcal{Z}_{n+1}(i)=k \mid \mathcal{G}_n)\big) = \Delta M_n^{(1)} - \Delta M_n^{(2)},$$

where $\Delta M_n^{(i)}$ is a martingale difference, i.e. $\Delta M_n^{(i)} = M_{n+1}^{(i)} - M_n^{(i)},\ i\in\{1,2\}$, and

$$\begin{aligned}\Delta M_n^{(1)} &= \frac{1}{n+1}\bigg(\sum_{i\in\mathbb{I}_n}\mathbb{1}_{\{\mathcal{Z}_n(i)<k,\mathcal{Z}_{n+1}(i)\geq k\}} - \mathbb{E}\bigg[\sum_{i\in\mathbb{I}_n}\mathbb{1}_{\{\mathcal{Z}_n(i)<k,\mathcal{Z}_{n+1}(i)\geq k\}}\,\bigg|\,\mathcal{G}_n\bigg]\bigg),\\ \Delta M_n^{(2)} &= \frac{1}{n+1}\bigg(\sum_{i\in\mathbb{I}_n}\mathbb{1}_{\{\mathcal{Z}_n(i)\leq k,\mathcal{Z}_{n+1}(i)> k\}} - \mathbb{E}\bigg[\sum_{i\in\mathbb{I}_n}\mathbb{1}_{\{\mathcal{Z}_n(i)\leq k,\mathcal{Z}_{n+1}(i)> k\}}\,\bigg|\,\mathcal{G}_n\bigg]\bigg).\end{aligned} \tag{4.2}$$

Indeed, we have used that for all $i\in\mathbb{N}, k\in\mathbb{N}_0$,

$$\begin{aligned}\mathbb{1}_{\{\mathcal{Z}_{n+1}(i)=k\}} &= \mathbb{1}_{\{\mathcal{Z}_{n+1}(i)=k,\mathcal{Z}_n(i)\leq k\}} = \mathbb{1}_{\{\mathcal{Z}_{n+1}(i)\geq k,\mathcal{Z}_n(i)\leq k\}} - \mathbb{1}_{\{\mathcal{Z}_{n+1}(i)> k,\mathcal{Z}_n(i)\leq k\}}\\ &= \mathbb{1}_{\{\mathcal{Z}_n(i)=k\}} + \mathbb{1}_{\{\mathcal{Z}_{n+1}(i)\geq k,\mathcal{Z}_n(i)< k\}} - \mathbb{1}_{\{\mathcal{Z}_{n+1}(i)> k,\mathcal{Z}_n(i)\leq k\}}.\end{aligned}$$

Note that, as the indicators in $M_n^{(1)}, M_n^{(2)}$ only differ by one index k, it is sufficient to prove the summability of the conditional second moment of $\Delta M_n^{(2)}$ for all fixed $k\geq 1$. So, we write

$$\begin{aligned}&\mathbb{E}\Big[(\Delta M_n^{(2)})^2\,\Big|\,\mathcal{G}_n\Big]\\ &= \frac{1}{(n+1)^2}\mathbb{E}\bigg[\Big(\sum_{i\in\mathbb{I}_n}\big(\mathbb{1}_{\{\mathcal{Z}_n(i)\leq k,\mathcal{Z}_{n+1}(i)>k\}} - \mathbb{P}(\mathcal{Z}_n(i)\leq k, \mathcal{Z}_{n+1}(i)>k|\mathcal{G}_n)\big)\Big)^2\bigg|\mathcal{G}_n\bigg].\end{aligned} \tag{4.3}$$

Using the non-positive correlation of the degree increments for the PAFRO model and Lemma 4.3 for the PAFFD and PAFUD models, we can bound this from above by,

$$\begin{aligned}&\frac{1}{(n+1)^2}\sum_{i\in\mathbb{I}_n}\mathbb{E}\Big[\Big(\mathbb{1}_{\{\mathcal{Z}_n(i)\leq k,\mathcal{Z}_{n+1}(i)>k\}} - \mathbb{P}(\mathcal{Z}_n(i)\leq k,\mathcal{Z}_{n+1}(i)>k \mid \mathcal{G}_n)\Big)^2\,\Big|\,\mathcal{G}_n\Big]\\ &\qquad\leq \frac{1}{(n+1)^2}\sum_{i\in\mathbb{I}_n}\mathbb{1}_{\{\mathcal{Z}_n(i)\leq k\}}\mathbb{P}(\Delta\mathcal{Z}_n(i)\geq 1\mid\mathcal{G}_n)\\ &\qquad\leq \frac{1}{(n+1)^2}\sum_{i=1}^n\mathbb{E}[\Delta\mathcal{Z}_n(i)\mid\mathcal{G}_n] = \frac{m}{(n+1)^2},\end{aligned} \tag{4.4}$$

where we use Markov's inequality in the final step and use that the increments of all in-degrees is exactly m by the definition of the PAFFD and PAFUD models. Hence, combining (4.3) and (4.4) yields the almost sure summability of the conditional second moments of $\Delta M_n^{(2)}$, which implies the almost sure convergence of R_n. For the PAFRO model, we use the same steps as in (4.3) and (4.4), but take the expected value on the left- and right-hand sides. Then, using the definition of the PAFRO model, we arrive at

$$\mathbb{E}\big[(\Delta M_n^{(2)})^2\big] \leq \frac{1}{(n+1)^2}\sum_{i=1}^n\mathbb{E}[\Delta\mathcal{Z}_n(i)] \leq \frac{1}{(n+1)^2}\sum_{i=1}^n\frac{\mathbb{E}[\mathcal{Z}_n(i)+\mathcal{F}_i]}{m_0+(n-n_0)}. \tag{4.5}$$

By using the tower rule and conditioning on $\mathcal{G}_{n-1}$, we find

$$\mathbb{E}[\mathcal{Z}_n(i)+\mathcal{F}_i] = \mathbb{E}[\mathbb{E}[\mathcal{Z}_n(i)+\mathcal{F}_i\,|\,\mathcal{G}_{n-1}]] \leq \mathbb{E}[\mathcal{Z}_{n-1}(i)+\mathcal{F}_i]\left(1+\frac{1}{m_0+(n-1-n_0)}\right).$$

Continuing this recursion yields

$$\mathbb{E}[\mathcal{Z}_n(i)+\mathcal{F}_i] \leq \mathbb{E}[\mathcal{Z}_{i\vee n_0}(i)+\mathcal{F}_i]\prod_{j=i\vee n_0}^{n-1}\Big(1+\frac{1}{m_0+(j-n_0)}\Big) \leq \frac{(m_0+\mathbb{E}[\mathcal{F}])(m_0+(n-n_0))}{m_0+(i\vee n_0-n_0)}.$$

Using this upper bound in (4.5), we obtain

$$\mathbb{E}\big[(\Delta M_n^{(2)})^2\big] \leq \frac{1}{(n+1)^2}\Big(C_1+\sum_{i=n_0+1}^{n}\frac{m_0+\mathbb{E}[\mathcal{F}]}{m_0+(i-n_0)}\Big) \leq \frac{C_1+C_2\log n}{(n+1)^2}, \tag{4.6}$$

for some constants $C_1, C_2 > 0$, which is indeed summable.

For $k=0$, we can write ΔR_n as

$$\Delta R_n := \Delta M_n^{(1)} + \Delta M_n^{(2)} + (\mathbb{1}_{\{\mathcal{F}_{n+1}\in(f,f']\}} - \mu((f,f']))/(n+1),$$

where $\Delta M_n^{(1)}=0$ and $\Delta M_n^{(2)}$ is as in (4.2) with $k=0$. We already proved the summability of the second conditional moment of $M_n^{(2)}$ which follows for $k=0$ as well, and the last term has a second conditional moment bounded by $\mu((f,f'])/(n+1)^2$, which is summable too. This proves the almost sure convergence of R_n.

For (ii), we have

$$\Delta R_n = \frac{1}{n+1}\sum_{i\in\mathbb{I}_n}(\mathcal{Z}_{n+1}(i)-\mathbb{E}[\mathcal{Z}_{n+1}(i)\mid\mathcal{G}_n]) = \frac{1}{n+1}\sum_{i\in\mathbb{I}_n}(\Delta\mathcal{Z}_n(i)-\mathbb{E}[\Delta\mathcal{Z}_n(i)\mid\mathcal{G}_n]),$$

as $\mathcal{Z}_{n+1}(i)=\mathcal{Z}_n(i)+\Delta\mathcal{Z}_n(i)$. We now bound the conditional second moments of ΔR_n by

$$\begin{aligned}\mathbb{E}\big[\Delta R_n^2 \,|\, \mathcal{G}_n\big] &= \frac{1}{(n+1)^2}\mathbb{E}\Big[\Big(\sum_{i\in\mathbb{I}_n}(\Delta\mathcal{Z}_n(i)-\mathbb{E}[\Delta\mathcal{Z}_n(i)\mid\mathcal{G}_n])\Big)^2 \,\Big|\, \mathcal{G}_n\Big]\\ &\leq \frac{1}{(n+1)^2}\sum_{i\in\mathbb{I}_n}\mathrm{Var}(\Delta\mathcal{Z}_n(i)\mid\mathcal{G}_n).\end{aligned} \tag{4.7}$$

The second line follows from Lemma 4.3 for the PAFFD and PAFUD models and from the conditional non-positive correlation of the $\mathcal{Z}_n(i)$ for the PAFRO model. Then, for the PAFUD and PAFFD models, we use that $\Delta\mathcal{Z}_n(i)$ is a sum of m indicator random variables and hence that its variance can be bounded by m times its mean. Also noting that the sum of all the increments of the in-degrees equals m, we obtain the upper bound $(m/(n+1))^2$, which is summable almost surely. For the PAFRO model, we again take the expected value on both sides of (4.7) to get rid of the conditional statement. Then, as the variance of $\Delta\mathcal{Z}_n(i)$ is bounded by its mean for the PAFRO model, and the same approach as used in (4.5) through (4.6) works here as well to arrive at a summable upper bound. □

With these lemmas at hand, we can prove Theorem 2.4:

Proof of Theorem 2.4. We provide a proof for the PAFFD and PAFUD models, the proof for the PAFRO model follows by setting $m=1$; the additional required adjustments are all included in the proof of Lemma 4.2.

First, we show that Γ_n converges in the weak* topology to Γ, defined in (2.5). To this end, we let $0<f<f'<\infty$, and set

$$\mathbb{I}_n := \{i\in[n]\mid\mathcal{F}_i\in(f,f']\},\quad X_n := \frac{1}{n}\sum_{i\in\mathbb{I}_n}\mathcal{Z}_n(i) = \Gamma_n((f,f']). \tag{4.8}$$

We develop a recursion for $X_{n+1} - X_n$. By writing $\mathcal{Z}_{n+1}(i) = \mathcal{Z}_n(i) + \Delta\mathcal{Z}_n(i)$ and $\bar{\mathcal{F}}_n := (m_0 + m(n-n_0) + S_n)/n$, we find

$$\mathbb{E}[X_{n+1} \mid \mathcal{G}_n] = \frac{1}{n+1}\Big(\sum_{i\in\mathbb{I}_n}\mathbb{E}[\mathcal{Z}_{n+1}(i) \mid \mathcal{G}_n]\Big) = X_n + \frac{1}{n+1}\Big(\sum_{i\in\mathbb{I}_n}\frac{\mathcal{Z}_n(i)+\mathcal{F}_i}{n\bar{\mathcal{F}}_n/m} - X_n\Big),$$

where we note that this holds for both the PAFFD as well as the PAFUD model. Then,

$$X_{n+1} - X_n = \frac{1}{n+1}\Big(\sum_{i\in\mathbb{I}_n}\frac{\mathcal{Z}_n(i)+\mathcal{F}_i}{n\bar{\mathcal{F}}_n/m} - X_n\Big) + \Delta R_n,$$

with $\Delta R_n := X_{n+1} - \mathbb{E}[X_{n+1} \mid \mathcal{G}_n]$. It is now possible to write the following two bounds:

$$\begin{aligned} X_{n+1} - X_n &\geq \frac{1}{n+1}\Big(-\Big(1 - \frac{m}{\bar{\mathcal{F}}_n}\Big)X_n + \frac{|\mathbb{I}_n|}{n}\frac{mf}{\bar{\mathcal{F}}_n}\Big) + \Delta R_n, \\ X_{n+1} - X_n &\leq \frac{1}{n+1}\Big(-\Big(1 - \frac{m}{\bar{\mathcal{F}}_n}\Big)X_n + \frac{|\mathbb{I}_n|}{n}\frac{mf'}{\bar{\mathcal{F}}_n}\Big) + \Delta R_n. \end{aligned}$$

We note that, by the strong law of large numbers, $|\mathbb{I}_n|/n$ converges almost surely to $\mu((f,f'])$ and $\bar{\mathcal{F}}_n$ converges almost surely to $m\theta_m$, where we recall that $\theta_m = 1 + \mathbb{E}[\mathcal{F}]/m$. From Lemma 4.2 it follows that $R_n := \sum_{k=n_0}^{n}\Delta R_n$ converges almost surely, so it follows from Lemma 4.1 that almost surely

$$\liminf_{n\to\infty} X_n \geq \frac{f}{\theta_m - 1}\mu((f,f'])), \qquad \limsup_{n\to\infty} X_n \leq \frac{f'}{\theta_m - 1}\mu((f,f'])). \tag{4.9}$$

We now take a countable subset $\mathbb{F} \subset [0,\infty)$ that is dense, such that for each $f \in \mathbb{F}$, $\mu(\{f\}) = 0$. As $\mathbb{F}$ is countable, there exists an almost sure event Ω_0 on which both statements in (4.9) hold for any pair $f, f' \in \mathbb{F}$ such that $f < f'$. Take an arbitrary open set U, and approximate U from below by a sequence of sets $(U_m)_{m\in\mathbb{N}}$, where each U_m is a finite union of small disjoint intervals $(f, f']$, with $f, f' \in \mathbb{F}$. Then, for any $m \in \mathbb{N}$, applying a Riemann approximation to (4.9),

$$\liminf_{n\to\infty}\Gamma_n(U) \geq \liminf_{n\to\infty}\Gamma_n(U_m) \geq \Gamma(U_m) \text{ on } \Omega_0. \tag{4.10}$$

Hence, by the monotone convergence theorem, it follows that $\liminf_{n\to\infty}\Gamma_n(U) \geq \Gamma(U)$. Likewise, for any closed set C, a similar argument shows that $\limsup_{n\to\infty}\Gamma_n(C) \leq \Gamma(C)$. It hence follows from the Portmanteau lemma [18, Theorem 13.16] that Γ_n converges to Γ a.s. in the weak* topology.

The approach to prove the other two parts in (2.4) is to apply induction on k to the convergence of the measures $\Gamma_n^{(k)}$ (and thus $p_n(k)$). We prove the statements in (2.4) hold for $k = 0$, the initialisation of the induction, below, and show the induction step first. Let us assume that the last two statements in (2.4) hold for all $0 \leq i < k$, for some $k \geq 1$. We now advance the induction hypothesis.

Let us take $0 < f < f' < \infty$, and define $X_n := \Gamma_n^{(k)}((f,f'])$. Then, we can write the following recurrence relation, using $\mathbb{I}_n$ as in (4.8):

$$\begin{aligned}
\mathbb{E}\big[X_{n+1}\,\big|\,\mathcal{G}_n\big] &= \frac{1}{n+1}\sum_{i=1}^{n+1}\mathbb{P}\big(\mathcal{Z}_{n+1}(i)=k,\mathcal{F}_i\in(f,f']\;\big|\;\mathcal{G}_n\big)\\
&= \frac{1}{n+1}\sum_{i\in\mathbb{I}_n}\sum_{\ell=0}^{k}\mathbb{1}_{\{\mathcal{Z}_n(i)=\ell\}}\mathbb{P}\big(\Delta\mathcal{Z}_n(i)=k-\ell\;\big|\;\mathcal{G}_n\big)\\
&= \frac{1}{n+1}\Big(\sum_{i\in\mathbb{I}_n}\sum_{\ell=0}^{k-1}\mathbb{1}_{\{\mathcal{Z}_n(i)=\ell\}}\mathbb{P}\big(\Delta\mathcal{Z}_n(i)=k-\ell\;\big|\;\mathcal{G}_n\big)\\
&\qquad+\sum_{i\in\mathbb{I}_n}\mathbb{1}_{\{\mathcal{Z}_n(i)=k\}}\big(1-\mathbb{P}\big(\Delta\mathcal{Z}_n(i)\geq 1\;\big|\;\mathcal{G}_n\big)\big)\Big),
\end{aligned}\tag{4.11}$$

where in the second step we note that $\mathcal{Z}_{n+1}(n+1)=0<k$ by definition and where we isolated the $\mathcal{Z}_n(i)=k$ case in the last step. We do this, as this proves to be the only part that does not converge to zero almost surely. We can then write

$$\begin{aligned}
\mathbb{E}\big[X_{n+1}\,\big|\,\mathcal{G}_n\big] &= X_n+\frac{1}{n+1}\Big(\sum_{i\in\mathbb{I}_n}\sum_{\ell=0}^{k-1}\mathbb{1}_{\{\mathcal{Z}_n(i)=\ell\}}\mathbb{P}\big(\Delta\mathcal{Z}_n(i)=k-\ell\;\big|\;\mathcal{G}_n\big)\\
&\qquad-\sum_{i\in\mathbb{I}_n}\mathbb{1}_{\{\mathcal{Z}_n(i)=k\}}\mathbb{P}\big(\Delta\mathcal{Z}_n(i)\geq 1\;\big|\;\mathcal{G}_n\big)-X_n\Big)\\
&= X_n+\frac{1}{n+1}\Big(\sum_{i\in\mathbb{I}_n}\sum_{\ell=0}^{k-1}\mathbb{1}_{\{\mathcal{Z}_n(i)=\ell\}}\mathbb{P}\big(\Delta\mathcal{Z}_n(i)=k-\ell\;\big|\;\mathcal{G}_n\big)\\
&\qquad-\sum_{i\in\mathbb{I}_n}\mathbb{1}_{\{\mathcal{Z}_n(i)=k\}}\Big(\mathbb{P}\big(\Delta\mathcal{Z}_n(i)\geq 1\;\big|\;\mathcal{G}_n\big)-\frac{k+\mathcal{F}_i}{n\bar{\mathcal{F}}_n/m}\Big)\\
&\qquad+\sum_{i\in\mathbb{I}_n}\mathbb{1}_{\{\mathcal{Z}_n(i)=k\}}\Big(\frac{f'-\mathcal{F}_i}{n\bar{\mathcal{F}}_n/m}\Big)-\Big(1+\frac{k+f'}{\bar{\mathcal{F}}_n/m}\Big)X_n\Big).
\end{aligned}\tag{4.12}$$

We can therefore write, using that $f'-\mathcal{F}_i\geq 0$ holds almost surely for all $i\in\mathbb{I}_n$,

$$X_{n+1}-X_n\geq\frac{1}{n+1}\left(A_n-B_nX_n\right)+R_{n+1}-R_n,\tag{4.13}$$

where

$$\begin{aligned}
A_n &:= \sum_{i\in\mathbb{I}_n}\sum_{\ell=0}^{k-1}\mathbb{1}_{\{\mathcal{Z}_n(i)=\ell\}}\mathbb{P}\big(\Delta\mathcal{Z}_n(i)=k-\ell\,\big|\,\mathcal{G}_n\big)\\
&\qquad-\sum_{i\in\mathbb{I}_n}\mathbb{1}_{\{\mathcal{Z}_n(i)=k\}}\Big(\mathbb{P}\big(\Delta\mathcal{Z}_n(i)\geq 1\,\big|\,\mathcal{G}_n\big)-\frac{k+\mathcal{F}_i}{n\bar{\mathcal{F}}_n/m}\Big),\\
B_n &:= 1+\frac{k+f'}{\bar{\mathcal{F}}_n/m},\\
\Delta R_n &:= R_{n+1}-R_n=X_{n+1}-\mathbb{E}[X_{n+1}\,|\,\mathcal{G}_n]\,.
\end{aligned}\tag{4.14}$$

We now prove the convergence of all three terms. First, we prove the convergence of A_n to

$$A:=\frac{1}{\theta_m}\int_{(f,f']}(k-1+x)\;\Gamma^{(k-1)}(\mathrm{d}x).\tag{4.15}$$

We note that, by the induction hypothesis, almost surely,

$$\lim_{n\to\infty}\Big|\int_{(f,f']}(k-1+x)\;\Gamma^{(k-1)}(\mathrm{d}x)-\int_{(f,f']}(k-1+x)\;\Gamma_n^{(k-1)}(\mathrm{d}x)\Big|=0.\tag{4.16}$$

We now deal with the two terms in A_n separately. We start with the second term. By the definition of the PAFFD and PAFUD models in Definition 2.1, it follows that for both models,

$$\mathbb{P}(\Delta\mathcal{Z}_n(i)\geq 1\,|\,\mathcal{G}_n)\leq 1-\Big(1-\frac{\mathcal{Z}_n(i)+\mathcal{F}_i}{n\bar{\mathcal{F}}_n}\Big)^m=\sum_{\ell=1}^m\binom{m}{\ell}(-1)^{\ell+1}\Big(\frac{\mathcal{Z}_n(i)+\mathcal{F}_i}{n\bar{\mathcal{F}}_n}\Big)^\ell.$$

Using this in the second term of A_n in (4.14), we obtain

$$\sum_{i\in\mathbb{I}_n}\mathbb{1}_{\{\mathcal{Z}_n(i)=k\}}\sum_{\ell=2}^m\binom{m}{\ell}(-1)^{\ell+1}\Big(\frac{k+\mathcal{F}_i}{n\bar{\mathcal{F}}_n}\Big)^\ell\leq C_m\sum_{\ell=2}^m n^{1-\ell}\Big(\frac{k+f'}{\bar{\mathcal{F}}_n}\Big)^\ell, \tag{4.17}$$

where $C_m>0$ is a constant. We note that this expression tends to zero almost surely as n tends to infinity, and that a similar lower bound that tends to zero almost surely can be constructed as well. For the first term, we write,

$$\begin{aligned}
&\lim_{n\to\infty}\Big|\sum_{i\in\mathbb{I}_n}\sum_{\ell=0}^{k-1}\mathbb{1}_{\{\mathcal{Z}_n(i)=\ell\}}\mathbb{P}\big(\Delta\mathcal{Z}_n(i)=k-\ell\,\big|\,\mathcal{G}_n\big)-\frac{1}{\theta_m}\int_{(f,f']}(k-1+x)\,\Gamma^{(k-1)}(\mathrm{d}x)\Big|\\
&\quad\leq\lim_{n\to\infty}\Bigg[\,\Big|\frac{1}{\theta_m}-\frac{1}{\bar{\mathcal{F}}_n/m}\Big|\int_{(f,f']}(k-1+x)\,\Gamma^{(k-1)}(\mathrm{d}x)\\
&\qquad+\frac{1}{\bar{\mathcal{F}}_n/m}\Big|\int_{(f,f']}(k-1+x)\,\Gamma^{(k-1)}(\mathrm{d}x)-\int_{(f,f']}(k-1+x)\,\Gamma_n^{(k-1)}(\mathrm{d}x)\Big|\\
&\qquad+\Big|\sum_{i\in\mathbb{I}_n}\mathbb{1}_{\{\mathcal{Z}_n(i)=k-1\}}\mathbb{P}\big(\Delta\mathcal{Z}_n(i)=1\,\big|\,\mathcal{G}_n\big)-\frac{m}{\bar{\mathcal{F}}_n}\int_{(f,f']}(k-1+x)\,\Gamma_n^{(k-1)}(\mathrm{d}x)\Big|\\
&\qquad+\sum_{i\in\mathbb{I}_n}\sum_{\ell=0}^{k-2}\mathbb{1}_{\{\mathcal{Z}_n(i)=\ell\}}\mathbb{P}\big(\Delta\mathcal{Z}_n(i)\geq 2\,\big|\,\mathcal{G}_n\big)\ \Bigg].
\end{aligned} \tag{4.18}$$

The first line converges to zero almost surely by the strong law of large numbers. By the induction hypothesis as used in (4.16), the second line converges to zero almost surely and by a similar argument as in (4.17) the last line converges to zero almost surely. For the third line, we use the definition of $\Gamma_n^{(k-1)}$, as defined in (2.3), to find

$$\begin{aligned}
&\sum_{i\in\mathbb{I}_n}\mathbb{1}_{\{\mathcal{Z}_n(i)=k-1\}}\mathbb{P}\big(\Delta\mathcal{Z}_n(i)=1\,\big|\,\mathcal{G}_n\big)-\frac{m}{\bar{\mathcal{F}}_n}\int_{(f,f']}(k-1+x)\,\Gamma_n^{(k-1)}(\mathrm{d}x)\\
&\quad=\sum_{i\in\mathbb{I}_n}\mathbb{1}_{\{\mathcal{Z}_n(i)=k-1\}}\Big(\mathbb{P}\big(\Delta\mathcal{Z}_n(i)=1\,\big|\,\mathcal{G}_n\big)-\frac{k-1+\mathcal{F}_i}{n\bar{\mathcal{F}}_n/m}\Big),
\end{aligned}$$

and so, again using similar steps as in (4.17), the third line in (4.18) converges to zero almost surely, which finishes the proof of the almost sure convergence of A_n to A, as in (4.15). Now, for B_n we immediately conclude that

$$\lim_{n\to\infty}B_n=1+\frac{k+f'}{\theta_m}=:B,$$

almost surely. Finally, the almost sure convergence of R_n again follows from Lemma 4.2. We thus obtain from Lemma 4.1,

$$\liminf_{n\to\infty}X_n\geq\frac{A}{B}=\frac{1}{k+f'+\theta_m}\int_{(f,f']}(k-1+x)\,\Gamma^{(k-1)}(\mathrm{d}x). \tag{4.19}$$

Likewise, the upper bound

$$\limsup_{n\to\infty}X_n\leq\frac{1}{k+f+\theta_m}\int_{(f,f']}(k-1+x)\,\Gamma^{(k-1)}(\mathrm{d}x) \tag{4.20}$$

can be established from (4.12), too, when we replace the f' by f in (4.12) and note that $f - \mathcal{F}_i \leq 0$ holds almost surely for all $i \in \mathbb{I}_n$.

We now again take a countable subset $\mathbb{F} \subset [0, \infty)$ that is dense, such that for each $f \in \mathbb{F}$, $\mu(\{f\}) = 0$. As $\mathbb{F}$ is countable, there exists an almost sure event Ω_0 on which both (4.19) and (4.20) hold for any pair $f, f' \in \mathbb{F}$ such that $f < f'$. A similar argument as in (4.9) and (4.10) can be made, using Riemann approximations and the Portmanteau lemma, which yields for any open set $U \subseteq [0, \infty)$ and any closed set $C \subseteq [0, \infty)$,

$$\begin{aligned} \liminf_{n\to\infty} \Gamma_n^{(k)}(U) &\geq \int_U \frac{k-1+x}{k+x+\theta_m} \Gamma^{(k-1)}(\mathrm{d}x), \\ \limsup_{n\to\infty} \Gamma_n^{(k)}(C) &\leq \int_C \frac{k-1+x}{k+x+\theta_m} \Gamma^{(k-1)}(\mathrm{d}x), \end{aligned} \tag{4.21}$$

and thus $\Gamma_n^{(k)}$ converges in the weak* topology to $\Gamma^{(k)}$, given by

$$\Gamma^{(k)}(\mathrm{d}x) = \frac{(k-1)+x}{k+x+\theta_m} \Gamma^{(k-1)}(\mathrm{d}x) = \ldots = \prod_{\ell=1}^{k} \frac{(\ell-1)+x}{\ell+x+\theta_m} \Gamma^{(0)}(\mathrm{d}x).$$

What remains is to perform the initialisation of the induction, regarding $\Gamma_n^{(0)}$. Analogous to the steps in (4.11), we now set $X_n := \Gamma_n^{(0)}((f, f'])$, with $0 < f < f' < \infty$, to obtain

$$\begin{aligned} \mathbb{E}[X_{n+1} \mid \mathcal{G}_n] &= \frac{1}{n+1} \bigg(\sum_{i \in \mathbb{I}_n} \mathbb{P}(\mathcal{Z}_{n+1}(i) = 0 \mid \mathcal{G}_n) + \mathbb{P}(\mathcal{F}_{n+1} \in (f, f']) \bigg) \\ &= \frac{1}{n+1} \bigg(\sum_{i \in \mathbb{I}_n} \mathbb{1}_{\{\mathcal{Z}_n(i)=0\}} \mathbb{P}(\Delta \mathcal{Z}_n(i) = 0 \mid \mathcal{G}_n) + \mu((f, f']) \bigg) \\ &= X_n + \frac{1}{n+1} \bigg(- \sum_{i \in \mathbb{I}_n} \mathbb{1}_{\{\mathcal{Z}_n(i)=0\}} \mathbb{P}(\Delta \mathcal{Z}_n(i) \geq 1 \mid \mathcal{G}_n) - X_n + \mu((f, f']) \bigg). \end{aligned}$$

Similar to (4.12), (4.13) and (4.14), we find

$$X_{n+1} - X_n \geq \frac{1}{n+1}(A_n - B_n X_n) + \Delta R_n, \tag{4.22}$$

where $A_n \to \mu((f, f'])$, $B_n \to (f' + \theta_m)/\theta_m$ a.s. as $n \to \infty$, and $\Delta R_n = R_{n+1} - R_n := X_{n+1} - \mathbb{E}[X_{n+1} \,|\, \mathcal{G}_n]$. As before, the almost sure convergence of R_n follows from Lemma 4.2. Analogously to (4.22),

$$X_{n+1} - X_n \leq \frac{1}{n+1}(A_n - B_n' X_n) + \Delta R_n$$

holds, with $B_n' \to (1 + f + \theta_m)/\theta_m$ almost surely. Hence, using Lemma 4.1,

$$\liminf_{n\to\infty} X_n \geq \frac{\theta_m}{f' + \theta_m} \mu((f, f']), \qquad \limsup_{n\to\infty} X_n \leq \frac{\theta_m}{f + \theta_m} \mu((f, f']),$$

and thus, with a similar reasoning as in (4.21), almost surely $\Gamma_n^{(0)}$ converges weakly in the weak* topology to

$$\Gamma^{(0)}(dx) := \frac{\theta_m}{x + \theta_m} \mu(\mathrm{d}x),$$

which yields

$$\Gamma^{(k)}(dx) = \frac{\theta_m}{x + \theta_m} \prod_{\ell=1}^{k} \frac{(\ell-1)+x}{\ell+x+\theta_m} \mu(\mathrm{d}x).$$

Then,

$$p(k) := \lim_{n\to\infty} p_n(k) = \int_0^\infty \frac{\theta_m}{x+\theta_m}\prod_{\ell=1}^k \frac{(\ell-1)+x}{\ell+x+\theta_m}\mu(\mathrm{d}x),$$

which proves (2.4) and concludes the proof. □

We now prove Theorem 2.6:

Proof of Theorem 2.6. We start by proving (i). The integrand of the integral in (2.6) can be written as

$$\frac{\theta_m}{x+\theta_m}\prod_{\ell=1}^k \frac{(\ell-1)+x}{\ell+x+\theta_m} = \theta_m \frac{\Gamma(x+\theta_m)}{\Gamma(k+x+1+\theta_m)}\frac{\Gamma(k+x)}{\Gamma(x)}.$$

From [16, Theorem 1] it follows that $k^{1+\theta_m}\Gamma(k+x)/\Gamma(k+x+1+\theta_m) \leq 1$ for all $x, k \geq 0$. By also using that $\Gamma(t+a)/\Gamma(t) = t^a(1+\mathcal{O}(1/t))$ as $t\to\infty$ and a fixed, we find that the dominated convergence theorem yields

$$\lim_{k\to\infty} p(k)k^{1+\theta_m} = \int_0^\infty \theta_m \frac{\Gamma(x+\theta_m)}{\Gamma(x)}\mu(\mathrm{d}x),$$

which is finite since $\mathbb{E}[\mathcal{F}^{\theta_m}] < \infty$.

We now prove (ii), so the fitness distribution satisfies Assumption 2.3. First, let $\alpha \in (2, 1+\theta_m)$. We write the integral in (2.6) as two separate integrals by splitting the domain into $(0,k)$ and (k,∞). We first concentrate on an upper bound. We note that, by symmetry, it also follows that $x^{1+\theta_m}\Gamma(k+x)/\Gamma(k+x+1+\theta_m) \leq 1$. Hence, we obtain the upper bound

$$k^{-(1+\theta_m)}\int_0^k \theta_m \frac{\Gamma(x+\theta_m)}{\Gamma(x)x^\theta}x^\theta\mu(\mathrm{d}x) + \int_k^\infty \theta_m \frac{\Gamma(x+\theta_m)}{\Gamma(x)x^{\theta_m}}x^{-1}\mu(\mathrm{d}x). \tag{4.23}$$

We note that there exists a constant $c > 1$ such that $\Gamma(x+\theta_m)/(\Gamma(x)x^{\theta_m}) \in [1, c]$ when $x \geq 1$. Hence, using Assumption 2.3, we can bound (4.23) from above by

$$\begin{aligned} \theta_m k^{-(1+\theta_m)}\int_0^1 &\frac{\Gamma(x+\theta_m)}{\Gamma(x)}\mu(\mathrm{d}x) + c\theta_m k^{-(1+\theta_m)}\int_1^k x^{\theta_m}\mu(\mathrm{d}x) + c\theta_m k^{-1}\int_k^\infty \mu(\mathrm{d}x) \\ &= o(k^{-\alpha}) + c\theta_m k^{-(1+\theta_m)}\mathbb{E}\big[\mathcal{F}^{\theta_m}\mathbb{1}_{\{1\leq \mathcal{F}^{\theta_m}\leq k\}}\big] + c\theta_m \ell(k)k^{-\alpha} \\ &= o(k^{-\alpha}) + c\theta_m^2 k^{-(1+\theta_m)}\int_1^k x^{\theta_m-1}\ell(x)x^{-(\alpha-1)}\mathrm{d}x + c\theta_m\ell(k)k^{-\alpha}, \end{aligned} \tag{4.24}$$

where the first term follows from the fact that $\alpha < 1+\theta_m$ and that the integral from 0 to 1 is finite. Hence, by [5, Proposition 1.5.8], as k tends to infinity, this is asymptotically

$$\frac{c\theta_m(2\theta_m-(\alpha-1))}{\theta_m-(\alpha-1)}\ell(k)k^{-\alpha}.$$

For a lower bound, we bound the second integral in (4.23) from below by zero, and bound the first integral, using similar steps as before, from below by

$$o(k^{-\alpha}) + \theta_m^2 k^{-(1+\theta_m)}\int_1^k x^{\theta_m-1}\ell(x)x^{-(\alpha-1)}\mathrm{d}x, \tag{4.25}$$

which is asymptotically, as k tends to infinity, $(\theta_m^2/(\theta_m-(\alpha-1))\ell(k)k^{-\alpha}$. Finally, for $\alpha = 1+\theta_m$, we note that the first term of (4.24) is no longer $o(k^{-\alpha})$, but of the same order as the other terms. Furthermore, since the argument of the integral in the last line of (4.24) (as well as in (4.25)) now equals $\ell(x)/x$, the integral equals $\ell^\star(k)$ and it follows from [5, Proposition 1.5.9a] that either $\ell^\star$ converges, in which case this falls under the first case (i) as the θ_m^{th} moment exists, or that $\ell^\star$ is slowly varying itself. Thus, in the latter case, we obtain an upper and lower bound with asymptotics, respectively,

$$\Big(\theta_m\int_0^1 \frac{\Gamma(x+\theta_m)}{\Gamma(x)}\mu(\mathrm{d}x)+c\theta_m\ell(k)+c\theta_m^2\ell^\star(k)\Big)k^{-(1+\theta_m)}=:\overline{L}(k)k^{-(1+\theta_m)},$$
$$\Big(\theta_m\int_0^1 \frac{\Gamma(x+\theta_m)}{\Gamma(x)}\mu(\mathrm{d}x)+\theta_m^2\ell^\star(k)\Big)k^{-(1+\theta_m)}=:\underline{L}(k)k^{-(1+\theta_m)}.$$

We also have from [5, Proposition 1.5.9a] that, in the case that $\ell^\star$ diverges as k tends to infinity, $\ell^\star(k)/\ell(k)\to\infty$ as $k\to\infty$ as well, so that $\overline{L}(k),\underline{L}(k)=\Theta(\ell^\star(k))$ as $k\to\infty$, which finishes the proof of (ii).

Finally, we tend to (iii). We provide a proof for the PAFFD and PAFUD models with $m\geq 1$ first, and then show how the results follows for the PAFRO model as well.

Recall that U_n is a uniformly chosen vertex from $[n]$. We first condition on the size of the fitness of U_n. Let $0<\beta<((2-\alpha)/(\alpha-1)\wedge 1)$. Note that when $U_n>n_0$, E_n denotes the event that $\mathcal{Z}_n(U_n)=0$. Then,

$$\mathbb{P}(E_n)\geq \mathbb{P}\big(E_n\cap\{\mathcal{F}_{U_n}\leq n^\beta\}\big)=\mathbb{P}\big(\mathcal{F}_{U_n}\leq n^\beta\big)-\mathbb{P}\big(E_n^c\cap\{\mathcal{F}_{U_n}\leq n^\beta\}\big)\,. \tag{4.26}$$

Clearly, for $\varepsilon>0$ fixed and n large,

$$\mathbb{P}\big(\mathcal{F}_{U_n}\leq n^\beta\big)=\mathbb{P}\big(\mathcal{F}\leq n^\beta\big)=1-\ell(n^\beta)n^{-(\alpha-1)\beta}\geq 1-n^{-(\alpha-1)\beta+\varepsilon}, \tag{4.27}$$

where we use Potter's theorem [5, Theorem 1.5.6], which states that for any fixed $\varepsilon>0$ and any function ℓ, slowly-varying at infinity,

$$\lim_{x\to\infty}\ell(x)x^{\varepsilon}=\infty,\qquad \lim_{x\to\infty}\ell(x)x^{-\varepsilon}=0. \tag{4.28}$$

For the second probability on the right-hand side of (4.26), we write

$$\begin{aligned}\mathbb{P}\big(E_n^c\cap\{\mathcal{F}_{U_n}\leq n^\beta\}\big)&=\mathbb{P}\Big(\bigcup_{j=U_n\vee n_0}^{n-1}\{\Delta\mathcal{Z}_j(U_n)\geq 1\}\cap\{\mathcal{F}_{U_n}\leq n^\beta\}\Big)\\&=\sum_{k=1}^n\frac1n\mathbb{P}\Big(\bigcup_{j=k\vee n_0}^{n-1}\{\Delta\mathcal{Z}_j(k)\geq 1\}\cap\{\mathcal{F}_k\leq n^\beta\}\Big)\\&\leq\sum_{k=1}^n\sum_{j=k\vee n_0}^{n-1}\frac1n\mathbb{P}\big(\{\Delta\mathcal{Z}_j(k)\geq 1\}\cap\{\mathcal{F}_k\leq n^\beta\}\big)\,.\end{aligned}$$

Now, using Markov's inequality, applying the tower rule and switching the summations yields the upper bound, writing $\bar{\mathcal{F}}_n=(m_0+m(n-n_0)+S_n)/n$,

$$\begin{aligned}&\frac1n\sum_{j=n_0}^{n-1}\sum_{k=1}^j\mathbb{E}\big[(\mathcal{Z}_j(k)+n^\beta)/(j\bar{\mathcal{F}}_j)\mathbb{1}_{\{\mathcal{F}_k\leq n^\beta\}}\big]\\&\quad=\frac1n\sum_{j=n_0}^{n-1}\sum_{k=1}^j\Big(\mathbb{E}\big[\mathcal{Z}_j(k)/(j\bar{\mathcal{F}}_j)\mathbb{1}_{\{\mathcal{F}_k\leq n^\beta\}}\big]+n^\beta\mathbb{E}\big[(j\bar{\mathcal{F}}_j)^{-1}\mathbb{1}_{\{\mathcal{F}_k\leq n^\beta\}}\big]\Big)\\&\quad\leq\frac1n\sum_{j=n_0}^{n-1}\sum_{k=1}^j\Big(\mathbb{E}[\mathcal{Z}_j(k)/(m_0+M_j)]+n^\beta\mathbb{E}\big[(m_0+M_j)^{-1}\big]\Big),\end{aligned} \tag{4.29}$$

where $M_j := \max_{k\le j}\mathcal{F}_k$, we bound $j\bar{\mathcal{F}}_j$ from below by m_0+M_j and we bound the indicator variables from above by 1. We now bound the first moment from above. Note that, for the PAFFD and PAFUD models,

$$\sum_{k=1}^{j}\mathbb{E}[\mathcal{Z}_j(k)] = m_0 + m(j-n_0), \tag{4.30}$$

since every vertex $i > n_0$ has out-degree m. Hence, combining (4.29) and (4.30), we obtain the upper bound, by using the tower rule and conditioning on the fitness,

$$\frac{1}{n}\sum_{j=n_0}^{n-1}(m+m_0+n^\beta)j\mathbb{E}[1/(m_0+M_j)] \le Cn^{\beta-1}\sum_{j=n_0}^{n-1}j\mathbb{E}[1/(m_0+M_j)], \tag{4.31}$$

when n is sufficiently large, for some constant $C>0$. We now bound $\mathbb{E}[1/(m_0+M_j)]$ from above.

$$\begin{aligned}\mathbb{E}[1/(m_0+M_j)] &= \mathbb{E}\Big[1/(m_0+M_j)\mathbb{1}_{\{M_j\le j^{1/(\alpha-1)-\varepsilon}\}}\Big]\\ &\quad+\mathbb{E}\Big[1/(m_0+M_j)\mathbb{1}_{\{M_j\ge j^{1/(\alpha-1)-\varepsilon}\}}\Big]\\ &\le \mathbb{P}\Big(M_j\le j^{1/(\alpha-1)-\varepsilon}\Big)+j^{-1/(\alpha-1)+\varepsilon}\end{aligned} \tag{4.32}$$

where we bound M_j from below by zero and $j^{1/(\alpha-1)-\varepsilon}$ in the first and second expectation, respectively. Then, using $1-x\le \mathrm{e}^{-x}$, for j large,

$$\mathbb{P}\Big(M_j\le j^{1/(\alpha-1)-\varepsilon}\Big)\le \exp\{-\ell(j^{1/(\alpha-1)-\varepsilon})j^{(\alpha-1)\varepsilon}\}\le\exp\{-j^{(\alpha-1)\varepsilon/2}\}, \tag{4.33}$$

where we use Potter's theorem, as in (4.28), in the last step. By combining (4.32) and (4.33), it follows that for j sufficiently large (say $j>j_0$ for some $j_0\in\mathbb{N}$),

$$\mathbb{E}[1/(m_0+M_j)]\le 2j^{-1/(\alpha-1)+\varepsilon},$$

and $\mathbb{E}[1/(m_0+M_j)]\le 1$ for $j\le j_0$. Using this in (4.31) yields

$$\begin{aligned}\mathbb{P}\big(E_n^c\cap\{\mathcal{F}_{U_n}\le n^\beta\}\big)&\le Cj_0n^{\beta-1}+4Cn^{\beta-1}\sum_{j=j_0+1}^{n-1}j^{1-1/(\alpha-1)+\varepsilon}\\ &\le \widetilde{C}n^{\beta+((1-1/(\alpha-1))\vee-1)+\varepsilon}\\ &=\widetilde{C}n^{\beta-((2-\alpha)/(\alpha-1)\wedge1)+\varepsilon},\end{aligned} \tag{4.34}$$

which, by the definition of β and the fact that ε is arbitrarily small, tends to zero as n tends to infinity. Finally, we combine (4.34) and (4.27) in (4.26) to find

$$\mathbb{P}(E_n)\ge 1-n^{-(\alpha-1)\beta+\varepsilon}-\widetilde{C}n^{\beta-((2-\alpha)/(\alpha-1)\wedge1)+\varepsilon}. \tag{4.35}$$

We now finish the proof of Theorem 2.4 by choosing the optimal value of $\beta\in(0,((2-\alpha)/(1-\alpha)\wedge1))$, namely $\beta=(2-\alpha)/(\alpha(\alpha-1))\wedge(1/\alpha)$, and setting $C=1+\widetilde{C}$.

For the PAFRO model, set m to equal 1. Then, there is one adjustment required. Namely, the equality in (4.30) does not hold. Rather, using (4.6) yields the upper bound

$$\sum_{k=1}^{j}\mathbb{E}_{\mathcal{F}}[\mathcal{Z}_j(k)]\le Cj(\log j-1)\le Cj^{1+\varepsilon},$$

for some large constant $C>0$. This adds at most an extra ε in the exponent of the final expression in (4.35) and since ε is arbitrarily small, the result still holds, which concludes the proof. □

5 Convergence of point process functionals

As mentioned in the proof overview in Section 3, in this section we complete an important step in the proof of Theorem 2.7 and show convergence of a functional of a point process as defined in (3.6) in the *extreme disorder regime* ($\alpha \in (1,2)$). At the same time, we take the chance to discuss some of the required theory of point process convergence, which also is useful in the next section when we consider the strong disorder case. A good reference for this theory is the book [21].

Recall u_n from Theorem 2.7 and let $M_p(E)$ be the space of point measures (point processes) on $E := (0,1) \times (0,\infty)$. Let us define the point process

$$\Pi_n := \sum_{i=1}^{n} \delta_{(i/n, \mathcal{F}_i/u_n)}, \tag{5.1}$$

with δ a Dirac measure. It follows from [21, Corollary 4.19] that, when the fitness distribution satisfies Assumption 2.3 for any $\alpha > 1$, Π_n has a weak limit Π, which is a Poisson point process (PPP) on E with intensity measure $\nu(\mathrm{d}t, \mathrm{d}x) := \mathrm{d}t \times (\alpha-1)x^{-\alpha}\mathrm{d}x$. [21, Proposition 4.20] shows that an almost surely continuous functional T_1 applied to Π_n converges in distribution to T_1 applied to Π by the continuous mapping theorem. In this section, we prove a similar result, though a slightly different approach is required.

Let $\varepsilon, \delta > 0, E_\delta := (0,1) \times (\delta, \infty)$. For a point measure $\Pi \in M_p(E)$, define

$$T^\varepsilon(\Pi) := \int_\varepsilon^1 \Big(\int_E f \mathbb{1}_{\{t \le s\}} \mathrm{d}\Pi(t,f) \Big)^{-1} \mathrm{d}s, \quad T_\delta^\varepsilon(\Pi) := \int_\varepsilon^1 \Big(\int_{E_\delta} f \mathbb{1}_{\{t \le s\}} \mathrm{d}\Pi(t,f) \Big)^{-1} \mathrm{d}s, \tag{5.2}$$

whenever these are well-defined. That is, when $\Pi((0,s) \times (0,\infty)) > 0$ for all $s \in (\varepsilon, 1)$ and when $\Pi((0,s) \times (\delta,\infty)) > 0$ for all $s \in (\varepsilon,1)$, respectively. As mentioned above, the reason for studying the functional T^ε is due to (3.6), where we see that the $\mathbb{E}_{\mathcal{F}}[\mathcal{Z}_n(i)/n]$ is (well) approximated by $m(\mathcal{F}_i/u_n)T^{i/n}(\Pi_n)$ in the *extreme disorder regime*, since the law of large numbers no longer applies to the fitness random variables in this regime. As a result, studying the maximum conditional mean in-degree can be done via studying this functional T^ε. Therefore, the main goal in this section is to prove the following proposition:

Proposition 5.1. *Let $(\mathcal{F}_i)_{i \in \mathbb{N}}$ be i.i.d. copies of a random variable $\mathcal{F}$, which follows a power-law distribution as in Assumption 2.3 with $\alpha \in (1,2)$. Consider the point measure Π_n in* (5.1)*, its weak limit Π and the functional T^ε in* (5.2)*. Then,*

$$\max_{i \in [n]} \frac{\mathcal{F}_i}{u_n} T^{i/n}(\Pi_n) \xrightarrow{d} \sup_{(t,f) \in \Pi} f T^t(\Pi).$$

To prove Proposition 5.1, one would normally prove the continuity of the functional T^ε and combine the weak convergence of Π_n with the continuous mapping theorem to yield the required result, as Resnick does in his proof of Proposition 4.20. This does, however, not work in this case. Due to the specific form of the functional, proving its continuity is not directly possible. Therefore, we investigate T_δ^ε as defined in (5.2) and show that this functional is indeed continuous and is 'sufficiently close' to T^ε. This is worked out in the following two lemmas:

Lemma 5.2. *Consider, for $\varepsilon \in (0,1), \delta > 0$ fixed, the operator T_δ^ε as in* (5.2)*. Then, the mapping $\Pi \mapsto \sum_{(t,f) \in \Pi : t > \varepsilon, f > \delta} \delta_{(f T_\delta^t(\Pi))}$ is continuous in the vague topology for measures $\Pi \in M_p(E)$ satisfying the following conditions:*

$$\begin{aligned} &\Pi(\{s\} \times (0,\infty)) = \Pi((s,t) \times \{0\}) = \Pi((s,t) \times \{\infty\}) = 0, && \forall s < t \in [0,1], \\ &\Pi((0,\varepsilon) \times (\delta,\infty)) > 0, \qquad \Pi((s,t) \times (x,\infty)) < \infty, && \forall s < t \in [0,1], x > 0. \end{aligned} \tag{5.3}$$

Remark 5.3. We note that for a PPP Π with intensity measure ν as introduced above, all the conditions in (5.3) are satisfied almost surely, except for $\Pi((0,\varepsilon)\times(\delta,\infty))>0$, which happens with positive probability only.

Proof of Lemma 5.2. We first prove that, for fixed $\varepsilon\in(0,1),\delta>0$, the mapping $\Pi\mapsto\sum_{(t,f)\in\Pi:t>\varepsilon,f>\delta}\delta_{(T^\varepsilon_\delta(\Pi))}$ is continuous in the vague topology for measures $\Pi\in M_p(E)$. We obtain this by taking $\Pi_n,\Pi\in M_p(E)$ such that $\Pi_n\overset{v}{\to}\Pi$, and show that the image of the mapping of Π_n introduced above also converges vaguely to the mapping of Π. Since the image is a point measure with only finitely many points, due to the last condition in (5.3), we can label the points (t,f) in Π such that $f>\delta$, by $(t_i,f_i),1\leq i\leq p$ for some $p\in\mathbb{N}$, where we order the points such that t_i is increasing in i. We can do the same for the points of Π_n, where we add a superscript n. Vague convergence is then equivalent to the convergence of $(t_i^n,f_i^n)\in\Pi_n$ to $(t_i,f_i)\in\Pi$ for all $1\leq i\leq p$, since there are only finitely many points.

By [21, Proposition 3.13], we can fix $\eta>0$ and take n large enough such that the balls $B_i:=B((t_i,f_i),\eta)$, centred around (t_i,f_i) with radii η, contain the points (t_i^n,f_i^n) and $B_i\cap B_j=\varnothing$ for $i\neq j$. Thus, let us set $q:=\Pi((0,\varepsilon)\times(\delta,\infty))>0$ and take n large enough such that $\Pi_n((0,\varepsilon)\times(\delta,\infty))=q$ as well. That is, points $(t_i,f_i),(t_i^n,f_i^n),i=1,\dots,q$, satisfy $t_i^n<\varepsilon$ and points $(t_i,f_i),(t_i^n,f_i^n),i=q+1,\dots,p$, satisfy $t_i^n>\varepsilon$ (due to the first condition in (5.3) there are no points (t,f) such that $t=\varepsilon$ a.s.). We can now express $T^\varepsilon_\delta(\Pi)$ in terms of a sum. Namely,

$$T^\varepsilon_\delta(\Pi)=\int_\varepsilon^1\Big(\int_{E_\delta}f\mathbb{1}_{\{t\leq s\}}\mathrm{d}\Pi(t,f)\Big)^{-1}\mathrm{d}s=\sum_{i=q+1}^{p+1}\Big[(t_i-t_{i-1}\vee\varepsilon)\Big(\sum_{j=1}^{i-1}f_i\Big)^{-1}\Big],\tag{5.4}$$

where we set $t_{p+1}:=1$. A similar expression follows for Π_n, with $t^n_{p+1}:=1$. Since the sum contains a finite number of terms, the convergence of $T^\varepsilon_\delta(\Pi_n)\to T^\varepsilon_\delta(\Pi)$ immediately follows from the convergence of the individual points. As $\Pi_n\overset{v}{\longrightarrow}\Pi$, $f_i^n\to f_i$ as n tends to infinity for all $i=1,\dots,p$ as well. What remains to prove, is that $(T^{t_i^n}_\delta(\Pi_n),1\leq i\leq p)\to(T^{t_i}_\delta(\Pi),1\leq i\leq p)$ as $n\to\infty$. Using the triangle inequality, we obtain

$$|T^{t_i^n}_\delta(\Pi_n)-T^{t_i}_\delta(\Pi)|\leq|T^{t_i^n}_\delta(\Pi_n)-T^{t_i}_\delta(\Pi_n)|+|T^{t_i}_\delta(\Pi_n)-T^{t_i}_\delta(\Pi)|.$$

Let us first consider $2\leq i\leq p$. The second term on the right-hand side tends to zero by the above, as for $i\geq 2$, $\Pi_n((0,t_i)\times(\delta,\infty))>0$ and thus the conditions in (5.3) are satisfied with $\varepsilon=t_i$. The first term can be rewritten using the definition of T^ε_δ in (5.2) as

$$\begin{aligned}|T^{t_i^n}_\delta(\Pi_n)-T^{t_i}_\delta(\Pi_n)|&=\int_{t_i^n\wedge t_i}^{t_i^n\vee t_i}\Big(\int_{E_\delta}f\mathbb{1}_{\{t\leq s\}}\mathrm{d}\Pi_n(t,f)\Big)^{-1}\mathrm{d}s\\&\leq|t_i^n-t_i|\Big(\int_{E_\delta}f\mathbb{1}_{\{t\leq t_i^n\wedge t_i\}}\mathrm{d}\Pi_n(t,f)\Big)^{-1},\end{aligned}$$

where we bound the integrand of the outer integral from above by replacing the integration variable s by $t_i^n\wedge t_i$ in the integral's argument. In the integral that remains, we can bound f from below by δ and therefore, for n sufficiently large, we can bound the integral from below by δ, as there is always at least one particle (t,f) such that $t\leq t_i^n\vee t_i$ since $i\geq 2$ and the balls B_i introduced above are disjoint. We thus obtain the upper bound $|t_i^n-t_i|/\delta$, which tends to zero with n. For $i=1$, we adapt our approach to find

$$\begin{aligned}|T^{t_1^n}_\delta(\Pi_n)-T^{t_1}_\delta(\Pi)|\leq\min\{&|T^{t_1^n}_\delta(\Pi_n)-T^{t_1}_\delta(\Pi_n)|+|T^{t_1}_\delta(\Pi_n)-T^{t_1}_\delta(\Pi)|,\\&|T^{t_1^n}_\delta(\Pi_n)-T^{t_1^n}_\delta(\Pi)|+|T^{t_1^n}_\delta(\Pi)-T^{t_1}_\delta(\Pi)|\}.\end{aligned}$$

When $t_1 < t_1^n$, the first term is infinite and we use the second term, while the second term is infinite when $t_1 > t_1^n$ and we then use the first term. When the first term is finite ($t_1 > t_1^n$), its first term is bounded from above by $(t_1 - t_1^n)\delta^{-1} < \eta/\delta$ and its second term can be bounded by a constant times η, as follows when using (5.4). Similarly, when the second term of the minimum is finite ($t_1 \leq t_1^n$), its second term is bounded from above by $(t_1^n - t_1)\delta^{-1} < \eta/\delta$ and its first term can be bounded by a constant times η. As η is arbitrary, the required result holds. □

We are also interested in how 'close' $T^\varepsilon(\Pi)$ and $T_\delta^\varepsilon(\Pi)$ (resp. $T^\varepsilon(\Pi_n)$ and $T_\delta^\varepsilon(\Pi_n)$) are when δ is small (resp. δ is small and n is large). We formalise this in the following lemma:

Lemma 5.4. *Consider the operator T_δ^ε as in (5.2) and the point process Π_n as in (5.1), let Π be its weak limit and let Assumption 2.3 hold with $\alpha \in (1,2)$. For $\varepsilon \in (0,1), \eta > 0$ fixed,*

$$\begin{aligned} T_\delta^\varepsilon(\Pi) &\xrightarrow{\mathbb{P}} T^\varepsilon(\Pi) \text{ as } \delta \downarrow 0, \\ \lim_{\delta\downarrow 0}\lim_{n\to\infty} \mathbb{P}(|T_\delta^\varepsilon(\Pi_n) &- T^\varepsilon(\Pi_n)| \geq \eta) = 0. \end{aligned} \tag{5.5}$$

Proof. We start by proving the first statement. We fix $\eta > 0$ and define $E_\delta^\xi := (0,\varepsilon) \times (\delta^{(2-\alpha)/2}(1+\delta^{-\xi}), \infty)$, where $\xi \in (0,(2-\alpha)/2)$. Then,

$$\mathbb{P}(|T_\delta^\varepsilon(\Pi) - T^\varepsilon(\Pi)| \geq \eta) \leq \mathbb{P}(|T_\delta^\varepsilon(\Pi) - T^\varepsilon(\Pi)| \geq \eta \,|\, \Pi(E_\delta^\xi) \neq 0) + \mathbb{P}(\Pi(E_\delta^\xi) = 0). \tag{5.6}$$

We condition on $\{\Pi(E_\delta^\xi) \neq 0\}$ to ensure that $T_\delta^\varepsilon(\Pi)$ is finite and show that on $\{\Pi(E_\delta^\xi) \neq 0\}$ the difference in $T_\delta^\varepsilon(\Pi)$ and $T^\varepsilon(\Pi)$ tends to zero in probability as $\delta \downarrow 0$. We first compute the second probability on the right-hand side.

$$\begin{aligned} \mathbb{P}\Big(\Pi(E_\delta^\xi) = 0\Big) &= \exp\Big\{-\int_{E_\delta^\xi} (\alpha-1)y^{-\alpha}\mathrm{d}y\mathrm{d}t\Big\} \\ &= \exp\Big\{-\varepsilon\delta^{-(\alpha-1)(2-\alpha)/2}(1+\delta^{-\xi})^{-(\alpha-1)}\Big\}. \end{aligned} \tag{5.7}$$

Note that, by the choice of ξ, this probability tends to zero with δ. Now, we bound the conditional probability in (5.6). Defining the event $F_{\delta,\xi} := \{\Pi(E_\delta^\xi) \neq 0\}$, we obtain,

$$\begin{aligned} &\mathbb{P}(|T_\delta^\varepsilon(\Pi) - T^\varepsilon(\Pi)| \geq \eta \,|\, F_{\delta,\xi}) \\ &= \mathbb{P}\bigg(\bigg|\int_\varepsilon^1 \bigg(\int_{E_\delta} f\mathbb{1}_{\{t\leq s\}}\mathrm{d}\Pi(t,f)\bigg)^{-1} - \bigg(\int_E f\mathbb{1}_{\{t\leq s\}}\mathrm{d}\Pi(t,f)\bigg)^{-1}\mathrm{d}s\bigg| \geq \eta \,\bigg|\, F_{\delta,\xi}\bigg) \\ &\leq \mathbb{P}\bigg(\int_\varepsilon^1 \bigg(\int_{E\setminus E_\delta} f\mathbb{1}_{\{t\leq s\}}\mathrm{d}\Pi(t,f)\bigg)\Big/\bigg(\int_{E_\delta} f\mathbb{1}_{\{t\leq s\}}\mathrm{d}\Pi(t,f)\bigg)^2\mathrm{d}s \geq \eta \,\bigg|\, F_{\delta,\xi}\bigg) \\ &\leq \mathbb{P}\bigg(\int_{E\setminus E_\delta} f\mathrm{d}\Pi(t,f) \geq \frac{\eta}{1-\varepsilon}\bigg(\int_{E_\delta} f\mathbb{1}_{\{t\leq\varepsilon\}}\mathrm{d}\Pi(t,f)\bigg)^2 \,\bigg|\, F_{\delta,\xi}\bigg), \end{aligned} \tag{5.8}$$

where, in the last line, we replaced the integration variable s with 1 in the integral in the numerator and with ε in the integral in the denominator. We now bound the integral over E_δ on the right-hand side from below using $\Pi(E_\delta^\xi) \geq 1$ and use Markov's inequality to find the upper bound

$$\begin{aligned}\mathbb{P}&\bigg(\int_{E\setminus E_\delta} f\mathrm{d}\Pi(t,f)\geq \frac{\eta}{1-\varepsilon}\delta^{2-\alpha}(1+\delta^{-\xi})^2\,\bigg|\,F_{\delta,\xi}\bigg)\\&=\mathbb{P}\bigg(\int_{E\setminus E_\delta} f\mathrm{d}\Pi(t,f)\geq \frac{\eta}{1-\varepsilon}\delta^{2-\alpha}(1+\delta^{-\xi})^2\bigg)\\&\leq \mathbb{E}\bigg[\int_{E\setminus E_\delta} f\mathrm{d}\Pi(t,f)\bigg]\frac{1-\varepsilon}{\eta}\delta^{-(2-\alpha)}(1+\delta^{-\xi})^{-2}\\&=\int_{E\setminus E_\delta}(\alpha-1)x^{1-\alpha}\mathrm{d}t\,\mathrm{d}x\frac{1-\varepsilon}{\eta}\delta^{-(2-\alpha)}(1+\delta^{-\xi})^{-2}=\frac{(1-\varepsilon)(\alpha-1)}{\eta(2-\alpha)}(1+\delta^{-\xi})^{-2},\end{aligned}\tag{5.9}$$

which tends to zero as $\delta\downarrow 0$. Note that we can omit the conditional statement in the second line, as the integral is independent of $\Pi(E_\delta^\xi)$. Combining (5.7) and the upper bound of (5.9) in (5.6), implies that $T_\delta^\varepsilon(\Pi)\xrightarrow{\mathbb{P}} T^\varepsilon(\Pi)$ as $\delta\downarrow 0$. We now prove the second statement in (5.5), which uses a similar approach. Namely, using analogous steps as in (5.6), (5.8) and (5.9), we obtain

$$\begin{aligned}&\mathbb{P}(|T^\varepsilon(\Pi_n)-T_\delta^\varepsilon(\Pi_n)|\geq\eta)\\&\quad\leq \mathbb{P}\bigg(\int_{E\setminus E_\delta} f\mathrm{d}\Pi_n(t,f)\geq\frac{\eta}{1-\varepsilon}\delta^{2-\alpha}(1+\delta^{-\xi})^{-2}\bigg)+\mathbb{P}\Big(\Pi_n(E_\delta^\xi)=0\Big).\end{aligned}\tag{5.10}$$

The second probability on the right-hand side converges to $\mathbb{P}(\Pi(E_\delta^\xi)=0)$ as n tends to infinity, and then to zero as δ tends to zero by (5.7). Using Markov's inequality, we obtain an upper bound for the first probability on the right-hand side of the form

$$\begin{aligned}&\sum_{i=1}^n \mathbb{E}\big[\mathcal{F}_i/u_n\mathbb{1}_{\{\mathcal{F}_i/u_n\leq\delta\}}\big]\frac{1-\varepsilon}{\eta}\delta^{-(2-\alpha)}(1+\delta^{-\xi})^2\\&\qquad=\frac{1-\varepsilon}{\eta}\delta^{-(2-\alpha)}(1+\delta^{-\xi})^2\frac{n}{u_n}\int_{x_\ell}^{\delta u_n}\ell(x)x^{-(\alpha-1)}\mathrm{d}x,\end{aligned}$$

where $x_\ell:=\inf\{x\in\mathbb{R}:F_{\mathcal{F}}(x)>0\}$. Using [5, Proposition 1.5.8], yields

$$\int_{x_\ell}^{\delta u_n}\ell(x)x^{-(\alpha-1)}\mathrm{d}x\sim\frac{1}{2-\alpha}(\delta u_n)^{2-\alpha}\ell(\delta u_n),\ \text{as } n\to\infty.$$

Thus, as $n\to\infty$, since ℓ is slowly-varying,

$$\frac{1-\varepsilon}{\eta}\delta^{-(2-\alpha)}(1+\delta^{-\xi})^2\frac{n}{u_n}\int_{x_\ell}^{\delta u_n}\ell(x)x^{-(\alpha-1)}\mathrm{d}x\sim\frac{(1-\varepsilon)}{\eta(2-\alpha)}(1+\delta^{-\xi})^{-2}n\ell(u_n)u_n^{-(\alpha-1)}.$$

Using [21, Corollary 4.19 and Proposition 3.21], we conclude that $n\ell(u_n)u_n^{-(\alpha-1)}$ converges to 1 and so the right-hand side tends to zero with δ. Thus,

$$\lim_{\delta\downarrow0}\lim_{n\to\infty}\mathbb{P}(|T^\varepsilon(\Pi_n)-T_\delta^\varepsilon(\Pi_n)|\geq\eta)=0,\tag{5.11}$$

which finishes the proof. □

We now prove Proposition 5.1.

Proof of Proposition 5.1. For a closed set $C\subseteq\mathbb{R}_+$ and $\eta>0$, let $C_\eta:=\{x\in\mathbb{R}:\inf_{y\in C}|x-y|\leq\eta\}$ be the η-enlargement of C and let us define the events

$$\begin{aligned}E_{n,\varepsilon,\delta}(\eta)&:=\Big\{\Big|\max_{i\in[n]}\frac{\mathcal{F}_i}{u_n}T^{i/n}(\Pi_n)-\max_{\varepsilon n\leq i\leq n:\mathcal{F}_i\geq\delta u_n}\frac{\mathcal{F}_i}{u_n}T_\delta^{i/n}(\Pi_n)\Big|<\eta\Big\},\\F_{n,\varepsilon,\delta}&:=\{\Pi_n((0,\varepsilon)\times(\delta,\infty))\geq1\}.\end{aligned}\tag{5.12}$$

We can then write

$$\mathbb{P}\Big(\max_{i\in[n]}\frac{\mathcal{F}_i}{u_n}T^{i/n}(\Pi_n)\in C\Big)\leq \mathbb{P}\bigg(\Big\{\max_{i\in[n]}\frac{\mathcal{F}_i}{u_n}T^{i/n}(\Pi_n)\in C\Big\}\cap E_{n,\varepsilon,\delta}(\eta)\cap F_{n,\varepsilon,\delta}\bigg)\\ +\mathbb{P}(E_{n,\varepsilon,\delta}(\eta)^c)+\mathbb{P}\big(F^c_{n,\varepsilon,\delta}\big). \tag{5.13}$$

Then, on $E_{n,\varepsilon,\delta}(\eta)$ and using C_η, we can bound the first probability on the right-hand side from above by

$$\mathbb{P}\bigg(\Big\{\max_{\varepsilon n\leq i\leq n:\mathcal{F}_i\geq \delta u_n}\frac{\mathcal{F}_i}{u_n}T^{i/n}_\delta(\Pi_n)\in C_\eta\Big\}\cap F_{n,\varepsilon,\delta}\bigg).$$

We note that every term in the maximum is bounded from above by 1. Then, since for n large $\Pi_n((\varepsilon,1)\times(\delta,\infty))=\Pi((\varepsilon,1)\times(\delta,\infty))<\infty$ and on $F_{n,\varepsilon,\delta}$, it follows from the continuous mapping theorem, Lemma 5.2 and Remark 5.3 that

$$\lim_{n\to\infty}\mathbb{P}\bigg(\Big\{\max_{\varepsilon n\leq i\leq n:\mathcal{F}_i\geq \delta u_n}\frac{\mathcal{F}_i}{u_n}T^{i/n}_\delta(\Pi_n)\in C_\eta\Big\}\cap F_{n,\varepsilon,\delta}\bigg)\\ =\mathbb{P}\bigg(\Big\{\sup_{(t,f)\in\Pi:t\geq\varepsilon,f\geq\delta}fT^t_\delta(\Pi)\in C_\eta\Big\}\cap F_{\varepsilon,\delta}\bigg), \tag{5.14}$$

where $F_{\varepsilon,\delta}:=\{\Pi((0,\varepsilon)\times(\delta,\infty))\geq 1\}$. We now claim that it is possible to remove the δ in $T^\varepsilon_\delta(\Pi)$ and the δ and ε constraints in the supremum in (5.14), as well as that the two terms in the last line of (5.13) tend to zero when letting n tend to infinity, and then δ and ε to zero. These two tasks require a very similar approach, as they are essentially the same, one with Π_n and the other with its weak limit Π. We start with the latter claim. We want to show that

$$\Big|\sup_{(t,f)\in\Pi:t\geq\varepsilon,f\geq\delta}fT^t_\delta(\Pi)-\sup_{(t,f)\in\Pi}fT^t(\Pi)\Big|\xrightarrow{\mathbb{P}}0 \text{ as first } \delta\downarrow 0 \text{ and then } \varepsilon\downarrow 0. \tag{5.15}$$

To this end, we write

$$\begin{aligned}\Big|\sup_{(t,f)\in\Pi:t\geq\varepsilon,f\geq\delta}fT^t_\delta(\Pi)-\sup_{(t,f)\in\Pi}fT^t(\Pi)\Big| &\leq \Big|\sup_{(t,f)\in\Pi:t\geq\varepsilon,f\geq\delta}fT^t_\delta(\Pi)-\sup_{(t,f)\in\Pi:t\geq\varepsilon}fT^t(\Pi)\Big|\\ &\quad+\Big|\sup_{(t,f)\in\Pi:t\geq\varepsilon}fT^t(\Pi)-\sup_{(t,f)\in\Pi}fT^t(\Pi)\Big|\\ &=: D_1+D_2.\end{aligned}$$

We first prove D_1 tends to zero in probability as $\delta\downarrow 0$. Namely, using the triangle inequality and the definitions of T^ε_δ and T^ε in (5.2),

$$\begin{aligned}D_1 &\leq \Big|\sup_{(t,f)\in\Pi:t\geq\varepsilon,f\geq\delta}fT^t_\delta(\Pi)-\sup_{(t,f)\in\Pi:t\geq\varepsilon,f\geq\delta}fT^t(\Pi)\Big|\\ &\quad+\Big|\sup_{(t,f)\in\Pi:t\geq\varepsilon,f\geq\delta}fT^t(\Pi)-\sup_{(t,f)\in\Pi:t\geq\varepsilon}fT^t(\Pi)\Big|\\ &\leq \sup_{(t,f)\in\Pi:t\geq\varepsilon,f\geq\delta}f(T^t_\delta(\Pi)-T^t(\Pi))+\sup_{(t,f)\in\Pi:t\geq\varepsilon,f<\delta}fT^t(\Pi)\\ &\leq\Big(\sup_{(t,f)\in\Pi}f\Big)\sup_{(t,f)\in\Pi:t\geq\varepsilon}(T^t_\delta(\Pi)-T^t(\Pi))+\delta T^\varepsilon(\Pi)\\ &\leq\Big(\sup_{(t,f)\in\Pi}f\Big)(T^\varepsilon_\delta(\Pi)-T^\varepsilon(\Pi))+\delta T^\varepsilon(\Pi),\end{aligned} \tag{5.16}$$

where the final inequality follows from the definitions of T^ε and T^ε_δ. Since $\alpha>1$, $\sup_{(t,f)\in\Pi}f<\infty$ almost surely. Furthermore, for any $\varepsilon>0$ fixed, $T^\varepsilon(\Pi)<\infty$ almost surely as well. Finally, by Lemma 5.4, $(T^\varepsilon_\delta(\Pi) - T^\varepsilon(\Pi)) \xrightarrow{\mathbb{P}} 0$ as $\delta \downarrow 0$. Thus, we obtain that $D_1 \xrightarrow{\mathbb{P}} 0$ as $\delta \downarrow 0$. We now show that $D_2 \xrightarrow{a.s.} 0$ as $\varepsilon \downarrow 0$. We discretise the interval $(0,1)$ into smaller sub-intervals $[2^{-(k+1)}, 2^{-k}), k \geq 0$. Then,

$$\lim_{\varepsilon\downarrow 0} D_2 \leq \lim_{\varepsilon\downarrow 0} \sup_{(t,f)\in\Pi: t<\varepsilon} fT^t(\Pi) = \lim_{K\to\infty} \sup_{k\geq K} \sup_{(t,f)\in\Pi: t\in[2^{-(k+1)},2^{-k})} fT^t(\Pi). \tag{5.17}$$

We now bound the inner supremum, by controlling the size of the maximum fitness value in these sub-intervals. That is, we define, for $\xi > 0, k \in \mathbb{Z}^+$,

$$\begin{aligned} \ell_k &:= 2^{-(k+1)/(\alpha-1)} \log((k+2)^{1+\xi})^{-1/(\alpha-1)}, \\ h_k &:= 2^{-(k+1)/(\alpha-1)} \log((1-(k+2)^{-(1+\xi)})^{-1})^{-1/(\alpha-1)}. \end{aligned} \tag{5.18}$$

Now,

$$\begin{aligned} \mathbb{P}\Big(\Pi([2^{-(k+1)}, 2^{-k}) \times (h_k, \infty)) \neq 0\Big) &= 1 - \exp\Big\{ - \int_{2^{-(k+1)}}^{2^{-k}} \int_{h_k}^\infty (\alpha-1)x^{-\alpha} \mathrm{d}x\mathrm{d}t \Big\} \\ &= 1 - \exp\{\log((1-(k+2)^{-(1+\xi)})\} \\ &\leq k^{-(1+\xi)}, \\ \mathbb{P}\Big(\Pi([2^{-(k+1)}, 2^{-k}) \times (\ell_k, \infty)) = 0\Big) &= \exp\Big\{ - \int_{2^{-(k+1)}}^{2^{-k}} \int_{\ell_k}^\infty (\alpha-1)x^{-\alpha} \mathrm{d}x\mathrm{d}t \Big\} \\ &\leq k^{-(1+\xi)}, \end{aligned} \tag{5.19}$$

which are both summable. Therefore, by the Borel-Cantelli lemma, it follows that almost surely there exist a random index L, such that for all $k \geq L$,

$$\sup_{(t,f)\in\Pi: t\in[2^{-(k+1)},2^{-k})} f \in (\ell_k, h_k). \tag{5.20}$$

Now, on the event $\{t \leq 2^{-L}\}$,

$$\begin{aligned} T^t(\Pi) &= \int_t^1 \Big(\int_E f\mathbb{1}_{\{u\leq s\}} \mathrm{d}\Pi(u,f)\Big)^{-1} \mathrm{d}s \\ &= \int_t^{2^{-L}} \Big(\int_E f\mathbb{1}_{\{u\leq s\}} \mathrm{d}\Pi(u,f)\Big)^{-1} \mathrm{d}s + \int_{2^{-L}}^1 \Big(\int_E f\mathbb{1}_{\{u\leq s\}} \mathrm{d}\Pi(u,f)\Big)^{-1} \mathrm{d}s \\ &\leq \int_t^{2^{-L}} (\sup_{(u,f)\in\Pi: u\leq s} f)^{-1} \mathrm{d}s + \Big(\int_E f\mathbb{1}_{\{u\leq 2^{-L}\}} \mathrm{d}\Pi(u,f)\Big)^{-1} \end{aligned} \tag{5.21}$$

By applying (5.20) to the both integrals, we find an upper bound

$$\sum_{j=L}^{\lceil \log_2(1/t) \rceil} 2^{-(j+2)} \ell_{j+1}^{-1} + \ell_L^{-1}.$$

Using the definition of ℓ_j in (5.18), for j large and some $\zeta \in (0, \alpha-1)$, we obtain

$$\begin{aligned} T^t(\Pi) &\leq C \sum_{j=L}^{\lceil \log_2(1/t) \rceil} 2^{(j+1)((1+\zeta)/(\alpha-1)-1)} + \ell_L^{-1} \\ &\leq \widetilde{C} t^{1-(1+\zeta)/(\alpha-1)} + \ell_L^{-1}, \end{aligned} \tag{5.22}$$

for some constant $\widetilde{C} > 0$. Again using (5.20) and on $\{k > L\}$ (similar to $t \leq 2^{-L}$), we find

$$\begin{aligned} \sup_{(t,f)\in\Pi: t\in[2^{-(k+1)},2^{-k})} fT^t(\Pi) &\leq h_k(\widetilde{C} 2^{(k+1)((1+\zeta)/(\alpha-1)-1)} + \ell_L^{-1}) \\ &\leq \widetilde{C} 2^{(k+1)(\zeta/(\alpha-1)-1)} k^\gamma + h_k \ell_L^{-1}, \end{aligned}$$

for some $\gamma > (1+\xi)/(\alpha-1)$. We finish the argument by noting that $L<\infty$ almost surely and hence

$$\begin{aligned}\lim_{K\to\infty}\sup_{k\geq K}\sup_{(t,f)\in\Pi:t\in[2^{-(k+1)},2^{-k})} fT^t(\Pi) &\leq \lim_{K\to\infty}\sup_{k\geq K}\widetilde{C}2^{(k+1)(\zeta/(\alpha-1)-1)}k^\gamma + h_k\ell_L^{-1}\\ &= \lim_{K\to\infty}\widetilde{C}2^{(K+1)(\zeta/(\alpha-1)-1)}K^\gamma + h_K\ell_L^{-1},\end{aligned} \tag{5.23}$$

which equals zero by the choice of ζ. Thus, $D_2 \xrightarrow{a.s.} 0$ as $\varepsilon\downarrow 0$. Together with the convergence of D_1 to zero in probability, we obtain (5.15). Recall $F_{n,\varepsilon,\delta}$ from (5.12) and $F_{\varepsilon,\delta} = \lim_{n\to\infty}F_{n,\varepsilon,\delta}$ under (5.14). Evidently, by a similar argument as in (5.7), $\lim_{\delta\downarrow 0}\mathbb{P}(F_{\varepsilon,\delta}) = 1$ for all $\varepsilon\in(0,1)$, which also shows the third probability in (5.13) tends to zero as $n\to\infty$ and then $\delta\downarrow 0$. Combining this with (5.15) and (5.14) yields

$$\lim_{\varepsilon\downarrow 0}\lim_{\delta\downarrow 0}\lim_{n\to\infty}\mathbb{P}\bigg(\Big\{\max_{\substack{\varepsilon n\leq i\leq n\\ \mathcal{F}_i\geq\delta u_n}}\frac{\mathcal{F}_i}{u_n}T_\delta^{i/n}(\Pi_n)\in C_\eta\Big\}\cap F_{n,\varepsilon,\delta}\bigg) = \mathbb{P}\bigg(\sup_{(t,f)\in\Pi} fT^t(\Pi)\in C_\eta\bigg). \tag{5.24}$$

Recall $E_{n,\varepsilon,\delta}(\eta)$ from (5.12). What remains to prove, is that for all $\eta>0$ fixed,

$$\lim_{\varepsilon\downarrow 0}\lim_{\delta\downarrow 0}\lim_{n\to\infty}\mathbb{P}(E_{n,\varepsilon,\delta}(\eta)^c) = 0,$$

which is very similar to (5.15), though we now deal with Π_n rather than Π. Again, we use the triangle inequality to find

$$\begin{aligned}\mathbb{P}(E_{n,\varepsilon,\delta}(\eta)^c) &\leq \mathbb{P}\bigg(\Big|\max_{\varepsilon n\leq i\leq n:\mathcal{F}_i\geq\delta u_n}\frac{\mathcal{F}_i}{u_n}T_\delta^{i/n}(\Pi_n) - \max_{\varepsilon n\leq i\leq n}\frac{\mathcal{F}_i}{u_n}T^{i/n}(\Pi_n)\Big|\geq\eta/2\bigg)\\ &\quad+\mathbb{P}\bigg(\Big|\max_{\varepsilon n\leq i\leq n}\frac{\mathcal{F}_i}{u_n}T^{i/n}(\Pi_n) - \max_{i\in[n]}\frac{\mathcal{F}_i}{u_n}T^{i/n}(\Pi_n)\Big|\geq\eta/2\bigg)\\ &=: P_1+P_2.\end{aligned} \tag{5.25}$$

We first deal with P_1. As in (5.16), we split this into two terms, namely

$$\begin{aligned}P_1 &\leq \mathbb{P}\bigg(\Big|\max_{\varepsilon n\leq i\leq n:\mathcal{F}_i\geq\delta u_n}\frac{\mathcal{F}_i}{u_n}T_\delta^{i/n}(\Pi_n) - \max_{\varepsilon n\leq i\leq n}\frac{\mathcal{F}_i}{u_n}T_\delta^{i/n}(\Pi_n)\Big|\geq\eta/4\bigg)\\ &\quad+\mathbb{P}\bigg(\Big|\max_{\varepsilon n\leq i\leq n}\frac{\mathcal{F}_i}{u_n}T_\delta^{i/n}(\Pi_n) - \max_{\varepsilon n\leq i\leq n}\frac{\mathcal{F}_i}{u_n}T^{i/n}(\Pi_n)\Big|\geq\eta/4\bigg).\end{aligned} \tag{5.26}$$

To show the first probability tends to zero, we write

$$\Big|\max_{\varepsilon n\leq i\leq n:\mathcal{F}_i\geq\delta u_n}\frac{\mathcal{F}_i}{u_n}T_\delta^{i/n}(\Pi_n) - \max_{\varepsilon n\leq i\leq n}\frac{\mathcal{F}_i}{u_n}T_\delta^{i/n}(\Pi_n)\Big| \leq \delta\max_{\varepsilon n\leq i\leq n}T_\delta^{i/n}(\Pi_n)\leq\delta T_\delta^\varepsilon(\Pi_n).$$

Then, on $F_{n,\varepsilon,\delta}$, $T_\delta^\varepsilon(\Pi_n)$ converges in distribution to $\delta T_\delta^\varepsilon(\Pi)$ by the continuous mapping theorem and the fact that T_δ^ε is continuous in Π_n, as follows from the proof of Lemma 5.2 and Remark 5.3. So, as $\delta\downarrow 0$, $T_\delta^\varepsilon(\Pi)\xrightarrow{\mathbb{P}} T^\varepsilon(\Pi)$, as follows from the proof of Lemma 5.4, which implies that $\delta T_\delta^\varepsilon(\Pi)\xrightarrow{\mathbb{P}} 0$ as $\delta\downarrow 0$. As before, $\mathbb{P}(F_{n,\varepsilon,\delta})\to 1$ as $n\to\infty$ and then $\delta\downarrow 0$, so by intersecting the first probability on the right-hand side of (5.26) with $F_{n,\varepsilon,\delta}, F_{n,\varepsilon,\delta}^c$, as in (5.13), yields that it tends to zero as $n\to\infty$ and then $\delta\downarrow 0$. What remains is to show that the second probability on the right-hand side of (5.26) tends to zero as n tends to infinity, then $\delta\downarrow 0$ and finally $\varepsilon\downarrow 0$. We again use a similar argument as in (5.16) to find

$$\Big|\max_{\varepsilon n\leq i\leq n}\frac{\mathcal{F}_i}{u_n}T_\delta^{i/n}(\Pi_n) - \max_{\varepsilon n\leq i\leq n}\frac{\mathcal{F}_i}{u_n}T^{i/n}(\Pi_n)\Big| \leq \Big(\max_{i\in[n]}\frac{\mathcal{F}_i}{u_n}\Big)(T_\delta^\varepsilon(\Pi_n) - T^\varepsilon(\Pi_n)). \tag{5.27}$$

We show that the product of the maximum and $(T^\varepsilon_\delta(\Pi_n)-T^\varepsilon(\Pi_n))$ converges to zero in probability as first $n\to\infty$ and then $\delta\downarrow 0$. We can use the fact that $(T^\varepsilon_\delta(\Pi_n)-T^\varepsilon(\Pi_n))$ tends to zero in probability as $n\to\infty$ and then $\delta\downarrow 0$, as is shown in the proof of Lemma 5.4. To extend this result to the product of these two random processes, we introduce the events $B_{n,\delta}:=\{\max_{i\in[n]}\mathcal{F}_i/u_n\le \delta^{-\xi}\}$, for some $\xi\in(0,(2-\alpha)/2)$. Then, splitting the second probability on the right-hand side of (5.26) into two parts by using (5.27) and intersecting with the events $B_{n,\delta}$ and $B^c_{n,\delta}$, we obtain the upper bound

$$\mathbb{P}\bigg(\Big|\max_{\varepsilon n\le i\le n}\frac{\mathcal{F}_i}{u_n}T^{i/n}_\delta(\Pi_n)-\max_{\varepsilon n\le i\le n}\frac{\mathcal{F}_i}{u_n}T^{i/n}(\Pi_n)\Big|\ge \eta/4\bigg)\le \mathbb{P}\big(T^\varepsilon_\delta(\Pi_n)-T^\varepsilon(\Pi_n)\ge \eta\delta^\xi/4\big)\\ +\mathbb{P}\big(B^c_{n,\delta}\big)\,.$$

$\mathbb{P}(B^c_{n,\delta})$ converges to $\mathbb{P}(B^c_\delta)$, where $B_\delta:=\{Y\le\delta^{-\xi}\}$ and Y is the distributional limit of $\max_{i\in[n]}\mathcal{F}_i/u_n$. Then, as $\delta\downarrow 0$, $\mathbb{P}(B^c_\delta)\to 0$, as Y is almost surely finite. Following the steps of the argument in (5.10) through (5.11) with $\eta\delta^\xi/4$ instead of η, we find

$$\begin{aligned}\limsup_{n\to\infty}\mathbb{P}(|T^\varepsilon(\Pi_n)-T^\varepsilon_\delta(\Pi_n)|\ge\eta\delta^\xi/4)&\le\frac{4(1-\varepsilon)}{\eta(\alpha-2)}\delta^{-\xi}(1+\delta^{-\xi})^{-2}+\limsup_{n\to\infty}\mathbb{P}(\Pi_n(E^\xi_\delta)=0)\\&=\frac{4(1-\varepsilon)}{\eta(\alpha-2)}\delta^{-\xi}(1+\delta^{-\xi})^{-2}+\mathbb{P}(\Pi(E^\xi_\delta)=0),\end{aligned}$$

which tends to zero as $\delta\downarrow 0$. It thus follows that $P_1\to 0$ as $n\to\infty$ and then $\delta\downarrow 0$.

What remains, is to show that P_2 tends to zero as $n\to\infty,\varepsilon\downarrow 0$. This follows from a similar approach as in (5.17) through (5.23). Recall ℓ_k,h_k from (5.18). We then divide the set of indices $i\in[n]$ into subsets $A_{k,n}:=\{i\in[n]:i\in(2^{-(k+1)}n,2^{-k}n]\},0\le k\le\lfloor\log n/\log 2\rfloor$, and define the events $\mathcal{A}^{\mathcal{F}}_{k,n}:=\big\{\max_{i\in A_{k,n}}\mathcal{F}_i/u_n\in(\ell_k,h_k)\big\}$. Using (5.19), it readily follows that

$$\liminf_{n\to\infty}\mathbb{P}\big(\mathcal{A}^{\mathcal{F}}_{k,n}\big)\ge 1-2k^{-(1+\xi)}.$$

Hence, when setting $k_n:=\lfloor\log n/\log 2\rfloor$, and for any sufficiently large $K\in\mathbb{N}$,

$$\liminf_{n\to\infty}\mathbb{P}\bigg(\bigcap_{K\le k\le k_n}\mathcal{A}^{\mathcal{F}}_{k,n}\bigg)\ge 1-CK^{-\xi},\tag{5.28}$$

for some constant $C>0$, independent of K. Similar to (5.17), we write

$$\limsup_{\varepsilon\downarrow 0}\limsup_{n\to\infty}P_2=\limsup_{K\to\infty}\limsup_{n\to\infty}\mathbb{P}\bigg(\sup_{k\ge K}\sup_{i\in A_{k,n}}\frac{\mathcal{F}_i}{u_n}T^{i/n}(\Pi_n)\ge\eta/4\bigg).$$

Again, the idea is to replace the limit of ε to 0 by the limit of K to ∞ and the supremum over $k\ge K$. Now, by intersecting with a similar event to the one in (5.28), we find the upper bound

$$\begin{aligned}\limsup_{K\to\infty}\limsup_{n\to\infty}\mathbb{P}\bigg(\Big\{\sup_{k\ge K}\sup_{i\in A_{k,n}}\frac{\mathcal{F}_i}{u_n}T^{i/n}(\Pi_n)\ge\eta/4\Big\}\cap\Big(\bigcap_{\sqrt{K}\le k\le k_n}A^{\mathcal{F}}_{k,n}\Big)\bigg)\\+\mathbb{P}\bigg(\bigcup_{\sqrt{K}\le k\le k_n}\Big(A^{\mathcal{F}}_{k,n}\Big)^c\bigg).\end{aligned}\tag{5.29}$$

By (5.28), it follows that the double limit of the second probability equals zero, so we focus on the first probability. Following the approach in (5.21) and (5.22) and using a

Markov bound, we bound the first probability in (5.29) from above by

$$\begin{aligned}&\frac{4}{\eta}\mathbb{E}\Big[\sup_{k\geq K}\sup_{i\in A_{k,n}}\frac{\mathcal{F}_i}{u_n}T^{i/n}(\Pi_n)\mathbb{1}_{\cap_{\sqrt{K}\leq k\leq k_n}A^{\mathcal{F}}_{k,n}}\Big]\\&\leq\frac{4}{\eta}\mathbb{E}\Big[\sup_{k\geq K}\sup_{i\in A_{k,n}}\frac{h_k}{n}\Big(\sum_{j=i}^{2^{-\sqrt{K}}n}u_n/M_j+\sum_{j=2^{-\sqrt{K}}n}^{n}u_n/M_j\Big)\mathbb{1}_{\cap_{\sqrt{K}\leq k\leq k_n}A^{\mathcal{F}}_{k,n}}\Big],\end{aligned}\tag{5.30}$$

where we recall that $M_j := \max_{m\leq j}\mathcal{F}_m$. We then bound the maximum in the second sum from below by considering only the indices $m \leq 2^{-\sqrt{K}}n$ and using the events in the indicator to further bound the maximum from below by $\ell_{\sqrt{K}}$. The terms of the second sum then are independent of j, which yields the upper bound $n(\ell_{\sqrt{K}})^{-1}$. We rewrite the first sum, where we note that for $i \in A_{k,n}, i \geq 2^{-(k+1)}n$, and as before bound the maximum from below to find

$$\sum_{j=i}^{2^{-\sqrt{K}}n}(M_j/u_n)^{-1}\leq\sum_{j\geq\sqrt{K}}^{k+1}\sum_{p\in A_{j,n}}(\ell_{j+1})^{-1}\leq n\sum_{j\geq\sqrt{K}}^{k+1}2^{-(j+1)}(\ell_{j+1})^{-1}.$$

Since, for large j, we can bound $(\ell_j)^{-1}$ from above by $2^{j(1/(\alpha-1)+\zeta)}$ for some small ζ, we obtain the upper bound $Cn2^{(k+1)((2-\alpha)/(\alpha-1)+\zeta)}$, for some constant $C > 0$. Note that this upper bound, as well as the upper bound stated above for the second sum in (5.30) are deterministic. Hence, using both upper bounds and bounding the indicator in the expectation in (5.30) from above by 1 yields the upper bound

$$\begin{aligned}C_\eta\sup_{k\geq K}\sup_{i\in A_{k,n}}(h_k2^{(k+1)((2-\alpha)/(\alpha-1)+\zeta)}+(\ell_{\sqrt{K}})^{-1}h_k)&\leq C_\eta\sup_{k\geq K}2^{-(k+1)(1-\zeta)}k^\gamma+\ell_{\sqrt{K}}^{-1}h_k\\&=C_\eta 2^{-(K+1)(1-\zeta)}K^\gamma+\ell_{\sqrt{K}}^{-1}h_K,\end{aligned}$$

for some $\gamma > 0$ and where $C_\eta = (4/\eta)\max\{C, 1\}$. This bound no longer depends on n, and as we let K tend to infinity the bound tends to zero. This proves P_2 tends to zero with $n\to\infty$ and then $\epsilon\downarrow 0$. Combining this result with the convergence of P_1 to zero with $n\to\infty$ and then $\delta\downarrow 0$, it follows that the upper bound in (5.25) tends to zero, and therefore the two probabilities on the second line of the right-hand side of (5.13) tend to zero with $n\to\infty$, then $\delta\downarrow 0$ and finally $\varepsilon\downarrow 0$. Together with (5.24), this yields

$$\limsup_{n\to\infty}\mathbb{P}\Big(\max_{i\in[n]}\frac{\mathcal{F}_i}{u_n}T^{i/n}(\Pi_n)\in C\Big)\leq\mathbb{P}\Bigg(\sup_{(t,f)\in\Pi}fT^t(\Pi)\in C_\eta\Bigg).$$

Including the limit $\eta\downarrow 0$ finally yields, by the continuity of the probability measure,

$$\limsup_{n\to\infty}\mathbb{P}\Big(\max_{i\in[n]}\frac{\mathcal{F}_i}{u_n}T^{i/n}(\Pi_n)\in C\Big)\leq\mathbb{P}\Big(\sup_{(t,f)\in\Pi}fT^t(\Pi)\in C\Big),$$

and applying the Portmanteau lemma [18, Theorem 13.16] finishes the proof. □

6 Martingales and concentration

In this section we state and prove several important results, required for the proof of Theorem 2.7. As discussed in the overview of the proof of Theorem 2.7 in Section 3, to study the degree evolution we use particular martingales. Understanding the behaviour of these martingales is essential for describing the different phases in the behaviour of the degrees $(\mathcal{Z}_n(i))_{i\in\mathbb{N}}$ as stated in Theorem 2.7.

We devote this section to (i) proving several results regarding these martingales, which is required for studying the behaviour of the evolution of the degrees $(\mathcal{Z}_n(i))_{i\in\mathbb{N}}$ (in the *weak disorder regime*), (ii) as well as proving other important results regarding the behaviour of the maximum conditional mean degree, which determines the behaviour of the maximum degree in the *strong and extreme disorder regime*, which we deal with in two separate subsections. First, however, we formulate the following propositions which outline the behaviour of the maximum degree in the strong and extreme disorder regime:

Proposition 6.1 (Maximum mean degree in the strong and extreme disorder regime)**.** *Consider the three PAF models as in Definition 2.1. Let* Π *be a Poisson Point Process (PPP) on* $E := (0,1)\times(0,\infty)$ *with intensity measure* $\nu(\mathrm{d}t,\mathrm{d}x) := \mathrm{d}t\times(\alpha-1)x^{-\alpha}\mathrm{d}x$*, and let* $\theta_m := 1+\mathbb{E}[\mathcal{F}]/m$*. Then, for* $\alpha\in(2,1+\theta_m)$*,*

$$\max_{i\in[n]}\mathbb{E}_{\mathcal{F}}[\mathcal{Z}_n(i)/u_n]\xrightarrow{d}\max_{(t,f)\in\Pi}f(t^{-1/\theta_m}-1), \tag{6.1}$$

while for $\alpha\in(1,2)$*,*

$$\max_{i\in[n]}\mathbb{E}_{\mathcal{F}}[\mathcal{Z}_n(i)/n]\xrightarrow{d} m\max_{(t,f)\in\Pi}f\int_t^1\bigg(\int_E g\mathbb{1}_{\{u\leq s\}}\mathrm{d}\Pi(u,g)\bigg)^{-1}\mathrm{d}s. \tag{6.2}$$

Proposition 6.2 (Concentration in the strong and extreme disorder regime)**.** *Consider the three PAF models as in Definition 2.1. When* $\alpha\in(2,1+\theta_m)$*, for any* $\eta>0$*,*

$$\lim_{n\to\infty}\mathbb{P}\bigg(\Big|\max_{i\in[n]}\mathcal{Z}_n(i)-\max_{i\in[n]}\mathbb{E}_{\mathcal{F}}[\mathcal{Z}_n(i)]\Big|>\eta u_n\bigg)=0. \tag{6.3}$$

Similarly, when $\alpha\in(1,2)$*, for any* $\eta>0$*,*

$$\lim_{n\to\infty}\mathbb{P}\bigg(\Big|\max_{i\in[n]}\mathcal{Z}_n(i)-\max_{i\in[n]}\mathbb{E}_{\mathcal{F}}[\mathcal{Z}_n(i)]\Big|>\eta n\bigg)=0. \tag{6.4}$$

6.1 A family of martingales

To prove Propositions 6.1 and 6.2 and to understand the behaviour of the maximum degree in the *weak disorder regime*, we introduce a family of martingales and derive some of their properties. We define, for $k\in\mathbb{R}, n, n_0, m, m_0\in\mathbb{N}$ and $a,b>-1$ such that $a-b>-1$,

$$\begin{aligned} c_n^k(m) &:= \prod_{j=n_0}^{n-1}\prod_{\ell=1}^{m}\Big(1-\frac{k}{m_0+m(j-n_0)+k+(\ell-1)+S_j}\Big),\\ \widetilde{c}_n^k(m) &:= \prod_{j=n_0}^{n-1}\Big(1-\frac{k}{m_0+m(j-n_0)+k+S_j}\Big)^m, \qquad \binom{a}{b} := \frac{\Gamma(a+1)}{\Gamma(b+1)\Gamma(a-b+1)}, \end{aligned} \tag{6.5}$$

where we recall S_j from (2.1). For ease of writing, we omit the (m) in $c_n^k(m), \widetilde{c}_n^k(m)$ whenever there is no ambiguity. We can then formulate the following lemma:

Lemma 6.3 (Degree and fitness martingales)**.** *Let* $i\in\mathbb{N}, k\geq -\min(\mathcal{F}_i,1)$*. For the PAFRO model (*$m=1$*) and the PAFUD model with out-degree* $m\in\mathbb{N}$*, the random variable*

$$M_n^k(i) := c_n^k(m)\binom{\mathcal{Z}_n(i)+\mathcal{F}_i+(k-1)}{k}$$

is a martingale with respect to $\mathcal{G}_{n-1}$ *for* $n\geq i\vee n_0$*, under the conditional probability measure* $\mathbb{P}_{\mathcal{F}}(\cdot)$*. For the PAFFD model with out-degree* $m\in\mathbb{N}$*, the random variable*

$$\widetilde{M}_n^k(i) := \widetilde{c}_n^k(m)\binom{\mathcal{Z}_n(i)+\mathcal{F}_i+(k-1)}{k}$$

is a supermartingale (resp. submartingale) with respect to $\mathcal{G}_{n-1}$ for $n \geq i \vee n_0$, under the conditional probability measure $\mathbb{P}_{\mathcal{F}}(\cdot)$ when $k \geq 0$ (resp. $k \in (-\min(\mathcal{F}_i, 1), 0)$. Finally, for the PAFFD model, $M_n^1(i)$ is a martingale with respect to $\mathcal{G}_{n-1}$ for $n \geq i \vee n_0$ under the conditional probability measure $\mathbb{P}_{\mathcal{F}}(\cdot)$.

Proof. For ease of writing, let us define $X_n(i) := \mathcal{Z}_n(i) + \mathcal{F}_i$ and $\Delta X_n(i) := X_{n+1}(i) - X_n(i) = \Delta \mathcal{Z}_n(i)$. For the PAFRO model, we use $c_n^k(1)$, which, as defined in (6.5) for general $m \in \mathbb{N}$, is equal to

$$c_n^k(1) = \prod_{j=n_0}^{n-1} \Big(1 - \frac{k}{m_0 + (j - n_0) + k + S_j}\Big).$$

For the proof of for the PAFRO model, we omit the (1) in $c_n^k(1)$. We can write

$$\begin{aligned} \mathbb{E}_{\mathcal{F}}[M_{n+1}^k(i) \mid \mathcal{G}_n] &= c_{n+1}^k \mathbb{E}_{\mathcal{F}}\Big[\binom{X_{n+1}(i) + (k-1)}{k} \;\Big|\; \mathcal{G}_n\Big] \\ &= c_{n+1}^k \mathbb{E}_{\mathcal{F}}\Big[\binom{X_n(i) + (k-1)}{k} \frac{\Gamma(X_{n+1}(i) + k)}{\Gamma(X_n(i) + k)} \frac{\Gamma(X_n(i))}{\Gamma(X_{n+1}(i))} \;\Big|\; \mathcal{G}_n\Big] \\ &= c_{n+1}^k \binom{X_n(i) + (k-1)}{k} \mathbb{E}_{\mathcal{F}}\Big[1 + \Delta X_n(i) \frac{k}{X_n(i)} \;\Big|\; \mathcal{G}_n\Big], \end{aligned} \tag{6.6}$$

as $\Delta X_n(i)$ is either 0 or 1. Then, taking the expected value of $\Delta X_n(i)$ yields

$$\mathbb{E}_{\mathcal{F}}[M_{n+1}^k(i) \mid \mathcal{G}_n] = c_{n+1}^k \binom{X_n(i) + (k-1)}{k} \Big(1 + \frac{X_n(i)}{m_0 + (n - n_0) + S_n} \frac{k}{X_n(i)}\Big) = M_n^k(i),$$

as $c_{n+1}^k(1 + k/(m_0 + (n - n_0) + S_n)) = c_n^k$. Note that the conditional mean of $M_n^k(i)$ is finite almost surely as well. For the PAFFD model with out-degree $m \in \mathbb{N}$, we can follow the same steps to find

$$\begin{aligned} \mathbb{E}_{\mathcal{F}}[\widetilde{M}_{n+1}^k(i) \mid \mathcal{G}_n] &= \widetilde{c}_{n+1}^k \binom{X_n(i) + (k-1)}{k} \mathbb{E}_{\mathcal{F}}\Big[\frac{\Gamma(X_{n+1}(i) + k)}{\Gamma(X_n(i) + k)} \frac{\Gamma(X_n(i))}{\Gamma(X_{n+1}(i))} \;\Big|\; \mathcal{G}_n\Big] \\ &= \widetilde{c}_{n+1}^k \binom{X_n(i) + (k-1)}{k} \mathbb{E}_{\mathcal{F}}\Big[\prod_{\ell=0}^{\Delta X_n(i)-1} \frac{X_n(i) + k + \ell}{X_n(i) + \ell} \;\Big|\; \mathcal{G}_n\Big] \\ &\leq \widetilde{c}_{n+1}^k \binom{X_n(i) + (k-1)}{k} \mathbb{E}_{\mathcal{F}}\Big[\Big(\frac{X_n(i) + k}{X_n(i)}\Big)^{\Delta X_n(i)} \;\Big|\; \mathcal{G}_n\Big], \end{aligned} \tag{6.7}$$

where we use Gamma function's properties in the second line and note that $x \mapsto (x+k)/x$ is decreasing in x for $k \geq 0$ in the last step. For $k \in (-\min(\mathcal{F}_i, 1), 0)$ the upper bound becomes a lower bound, as $x \mapsto (x + k)/x$ is decreasing in x in that case. Conditionally on $\mathcal{G}_n$, the number of edges vertex $n + 1$ connects to i is a binomial random variable with m trials and success probability $X_n(i)/\sum_{j=1}^n X_n(j)$, so

$$\mathbb{E}_{\mathcal{F}}\Big[\Big(\frac{X_n(i) + k}{X_n(i)}\Big)^{\Delta X_n(i)} \;\Big|\; \mathcal{G}_n\Big] = \Big(1 + \frac{k}{\sum_{j=1}^n X_n(j)}\Big)^m,$$

where we use that a random variable $X \sim \mathrm{Bin}(m, p)$ has probability generating function $\mathbb{E}\big[z^X\big] = (pz + (1-p))^m$, $z \in \mathbb{R}$. Then, recalling that for the PAFFD model $\sum_{i=1}^n X_n(i) = m_0 + m(n - n_0) + S_n$ yields the result. For the PAFUD model, we require a few more steps. As the connection of the i^{th} edge of vertex $n + 1$ is dependent on the connection of edges $1, \ldots, i - 1$, we iteratively condition on $\mathcal{G}_{n,j}$, $j = m - 1, m - 2, \ldots, 0$, the graph with n vertices where the $n+1^{\text{st}}$ vertex has connected j of its half-edges to the vertices $1,\dots,n$. More precisely, letting $X_{n,j} := \mathcal{Z}_{n,j}(i)+\mathcal{F}_i$, we write

$$\begin{aligned}\mathbb{E}_{\mathcal{F}}[M^k_{n+1}(i)\mid \mathcal{G}_n] &= c^k_{n+1}\mathbb{E}_{\mathcal{F}}\Big[\mathbb{E}\Big[\binom{X_{n+1,0}(i)+(k-1)}{k}\,\Big|\,\mathcal{G}_{n,m-1}\Big]\Big|\mathcal{G}_n\Big]\\ &= c^k_{n+1}\mathbb{E}_{\mathcal{F}}\Big[\mathbb{E}\Big[\binom{X_{n,m-1}(i)+\mathbb{1}_{n+1,m,i}+(k-1)}{k}\,\Big|\,\mathcal{G}_{n,m-1}\Big]\Big|\mathcal{G}_n\Big],\end{aligned}$$

where $\mathbb{1}_{n+1,m,i}$ is the indicator of the event that the m^{th} half-edge of vertex $n+1$ connects with vertex i. Now, as in (6.6), we write this as

$$\mathbb{E}_{\mathcal{F}}[M^k_{n+1}(i)\mid \mathcal{G}_n] = c^k_{n+1}\mathbb{E}_{\mathcal{F}}\Big[\binom{X_{n,m-1}(i)+(k-1)}{k}\Big(1+k\frac{\mathbb{E}[\mathbb{1}_{n+1,m,i}\mid \mathcal{G}_{n,m-1}]}{X_{n,m-1}(i)}\Big)\Big|\mathcal{G}_n\Big].$$

By the definition of the PAFUD model, the mean of the indicator equals $X_{n,m-1}(i)/\sum_{j=1}^n X_{n,m-1}(j) = X_{n,m-1}(i)/(m_0+m(n-n_0)+(m-1)+S_n)$. Hence, we obtain

$$\begin{aligned}&\mathbb{E}_{\mathcal{F}}[M^k_{n+1}(i)\mid \mathcal{G}_n]\\ &\quad= c^k_{n+1}\Big(1+\frac{k}{m_0+m(n-n_0)+(m-1)+S_n}\Big)\mathbb{E}_{\mathcal{F}}\Big[\binom{X_{n,m-1}(i)+(k-1)}{k}\Big|\mathcal{G}_n\Big],\end{aligned}$$

which, when iteratively following the same steps by conditioning on $\mathcal{G}_{n,j}$ for $j=m-2,\dots,0$, yields the required result. Finally, we prove that $M^1_n(i)$ is a martingale in the PAFFD model. We repeat the steps in (6.7), but note that as $k=1$, we can omit the inequality and obtain

$$\mathbb{E}_{\mathcal{F}}[M^1_{n+1}(i)\,|\,\mathcal{G}_n] = c^1_{n+1}(m)X_n(i)(1+\mathbb{E}_{\mathcal{F}}[\Delta X_n(i)\,|\,\mathcal{G}_n]/X_n(i)).$$

As before, we note that $\Delta X_n(i)$ is a binomial random variable with mean $mX_n(i)/\sum_{j=1}^n X_n(j)$. Thus,

$$\mathbb{E}_{\mathcal{F}}[M^1_{n+1}(i)\,|\,\mathcal{G}_n] = c^1_{n+1}(m)X_n(i)\Big(1+\frac{m}{m_0+m(n-n_0)+S_n}\Big) = c^n_n(m)X_n(i) = M^1_n(i),$$

which finishes the proof. □

From Lemma 6.3, we immediately conclude that the (super)martingales $M^k_n(i), \widetilde{M}^k_n(i)$ converge almost surely, as they are non-negative, to some random variables $\xi^k_i, \widetilde{\xi}^k_i$, respectively. To distil from this an understanding of the behaviour of the evolution of the degrees $(\mathcal{Z}_n(i))_{i\in\mathbb{N}}$, we study the growth rate of the normalising sequences $c^k_n, \widetilde{c}^k_n$:

Lemma 6.4. *Consider the sequences $c^k_n, \widetilde{c}^k_n$ in* (6.5) *and recall $\theta_m := 1+\mathbb{E}[\mathcal{F}]/m$. If $\mathbb{E}\big[\mathcal{F}^{1+\varepsilon}\big]<\infty$ for some $\varepsilon>0$, then for any $k\in\mathbb{R}, m\in\mathbb{N}$,*

$$c^k_n(m)n^{k/\theta_m} \xrightarrow{a.s.} c_k(m), \qquad \widetilde{c}^k_n(m)n^{k/\theta_m} \xrightarrow{a.s.} \widetilde{c}_k(m), \tag{6.8}$$

for some almost surely finite random variables $c_k(m), \widetilde{c}_k(m)$. When the fitness distribution satisfies Assumption 2.3 with $\alpha\in(1,2)$, for any $k\in\mathbb{R}, m\in\mathbb{N}$,

$$c^k_n \xrightarrow{a.s.} \underline{c}_k(m), \qquad \widetilde{c}^k_n \xrightarrow{a.s.} \underline{\widetilde{c}}_k(m), \tag{6.9}$$

for some almost surely finite random variables $\underline{c}_k(m), \underline{\widetilde{c}}_k(m)$ (again omitting the (m) whenever there is no ambiguity). Furthermore, the following upper and lower bounds *hold almost surely for* $c_n^k(m)$ *when* $\mathbb{E}\big[\mathcal{F}^{1+\varepsilon}\big]<\infty$ *for some* $\varepsilon>0$ *(they hold for* $\widetilde{c}_n^k(m)$ *as well). For* $n_0+1\le i\le n$*,*

$$\begin{aligned}\frac{c_i^k(m)}{c_n^k(m)}\Big(\frac{i}{n}\Big)^{k/\theta_m} &\le \exp\bigg\{\frac{k}{\theta_m}\log\Big(\frac{i}{n}\frac{n-(n_0+1)}{(i-(n_0+1))\vee 1}\Big)\\ &\qquad+\frac{mk}{\mathbb{E}[\mathcal{F}]}\sum_{j=i}^{\infty}\frac{|S_j/j-\mathbb{E}[\mathcal{F}]\,|}{m_0+m(j-n_0)+S_j}\bigg\},\\ \frac{c_i^k(m)}{c_n^k(m)}\Big(\frac{i}{n}\Big)^{k/\theta_m} &\ge 1-\frac{mk}{\mathbb{E}[\mathcal{F}]}\sum_{j=i}^{n-1}\frac{|S_j/j-\mathbb{E}[\mathcal{F}]\,|}{m_0+m(j-n_0)+S_j}-\frac{m}{2}\sum_{j=i}^{n-1}\Big(\frac{k}{S_j}\Big)^2\\ &\qquad-\frac{m_0+\mathbb{E}[\mathcal{F}]\,n_0+(m-1)}{\theta_m^2}\frac{\pi^2}{6i}-\frac{1}{\theta_m((i-(n_0+1))\vee 1)}.\end{aligned} \tag{6.10}$$

Proof. We only prove the results for $c_n^k(1)$, as the proofs for $m>1$ and $\widetilde{c}_n^k(m)$ follow similarly. For ease of writing, let $\theta:=\theta_1$. We start by proving (6.8). We can write

$$\begin{aligned}c_n^k n^{k/\theta} &= \exp\bigg\{-\sum_{j=n_0}^{n-1}\log\bigg(1+\frac{k}{m_0+j-n_0+S_j}\bigg)+\frac{k}{\theta}\log n\bigg\}\\ &=\exp\bigg\{-\sum_{j=n_0}^{n_0+\lceil 2|k|\rceil}\log\bigg(1+\frac{k}{m_0+j-n_0+S_j}\bigg)-\sum_{j=n_0+\lceil 2|k|\rceil+1}^{n-1}\frac{k}{j\theta}\\ &\qquad-\sum_{j=n_0+\lceil 2|k|\rceil+1}^{n-1}\frac{k}{j\theta}\frac{(\mathbb{E}[\mathcal{F}]-S_j/j)-(m_0-n_0)/j}{(m_0-n_0)/j+1+S_j/j}\\ &\qquad+\sum_{j=n_0+\lceil 2|k|\rceil+1}^{n-1}\sum_{\ell=2}^{\infty}(-1)^{\ell}\frac{1}{\ell}\Big(\frac{k}{m_0+j-n_0+S_j}\Big)^{\ell}+\frac{k}{\theta}\log n\bigg\},\end{aligned} \tag{6.11}$$

where we apply a Taylor expansion on the logarithmic terms in the sum for $j\ge n_0+\lceil 2|k|\rceil+1$. The second sum and the last term balance, their sum converges to some finite value depending on k and γ, where γ is the Euler-Mascheroni constant. We now show the almost sure absolute convergence of the third sum in the second line of (6.11). This is implied by the almost sure convergence of

$$\sum_{j=1}^{n}\frac{1}{j^2}|S_j-j\mathbb{E}[\mathcal{F}]\,|.$$

We prove this by showing that the mean of this sum converges. Let $\varepsilon>0$ such that the $(1+\varepsilon)^{\text{th}}$ moment of the $\mathcal{F}_i$ exists. Using Hölder's inequality, we obtain

$$\sum_{j=1}^{n}\mathbb{E}\big[|S_j-j\mathbb{E}[\mathcal{F}]\,|/j^2\big]\le\sum_{j=1}^{n}\frac{1}{j^2}\mathbb{E}\big[|S_j-j\mathbb{E}[\mathcal{F}]\,|^{1+\varepsilon}\big]^{1/(1+\varepsilon)}.$$

Now, we use a specific case of the Marcinkiewicz-Zygmund inequality [14, Proposition 3.8.2], which states that for $q\in[1,2]$ and i.i.d. X_i with $\mathbb{E}[X_1]=0,\mathbb{E}[|X_1|^q]<\infty$, there exists a constant c_q such that

$$\mathbb{E}\bigg[\Big|\sum_{i=1}^{j}X_i\Big|^q\bigg]\le c_q j\mathbb{E}[|X_1|^q]\,. \tag{6.12}$$

Thus, if we set $X_i:=\mathcal{F}_i-\mathbb{E}[\mathcal{F}]$, it follows that

$$\sum_{j=1}^{n}\frac{1}{j^2}\mathbb{E}\big[|S_j-j\mathbb{E}[\mathcal{F}]\,|^{1+\varepsilon}\big]^{1/(1+\varepsilon)}\le c_{1+\varepsilon}\mathbb{E}\big[|\mathcal{F}-\mathbb{E}[\mathcal{F}]\,|^{1+\varepsilon}\big]^{1/(1+\varepsilon)}\sum_{j=1}^{n}j^{-(2-1/(1+\varepsilon))},$$

which converges, as $\varepsilon > 0$. Finally, taking the absolute value of the double sum in (6.11) yields the upper bound

$$\sum_{j=n_0+\lceil 2|k|\rceil+1}^{n-1} \sum_{\ell=2}^{\infty} \frac{1}{\ell}\Big(\frac{|k|}{m_0+j-n_0+S_j}\Big)^{\ell} \leq \sum_{\ell=2}^{\infty} \sum_{j=\lceil 2|k|\rceil+1}^{\infty} \frac{|k|^{\ell}}{j^{\ell}} \leq |k| \sum_{\ell=2}^{\infty}\sum_{i=2}^{\infty} i^{-\ell}.$$

In the first step, we first bound $m_0+j-n_0+S_j$ from below by $j-n_0$ and then take all terms where $ik < j \leq (i+1)k$, $i \geq 2$, and bound them from below by $i|k|$, which yields the same upper bound $|k|$ times in the third step. The right-hand side equals

$$|k|\sum_{\ell=2}^{\infty}(\zeta(\ell)-1) = |k|,$$

where ζ is the Riemann zeta function, which thus proves the almost sure convergence of the double sum. This proves (6.8). For proving (6.9) we use a different approach. Namely, we prove that $-\log c_n^k$ converges almost surely, which yields the desired result as well. To that end, let $M_j := \max_{i\leq j}\mathcal{F}_i$. Then, we write

$$-\log c_n^k = \sum_{j=n_0}^{n-1} \log\Big(1+\frac{k}{m_0+j-n_0+S_j}\Big) \leq \sum_{j=1}^{J}\frac{k}{M_j} + k\sum_{j=J+1}^{n} j^{-1/(\alpha-1)+\varepsilon}, \tag{6.13}$$

where we use (4.33) in the last step to conclude that, by the Borel-Cantelli lemma, there exists an almost surely finite random index J such that for all $j \geq J$, $M_j \geq j^{1/(\alpha-1)-\varepsilon}$, for some small $\varepsilon \in (0,(2-\alpha)/(\alpha-1))$, as well as that $\log(1+x) \leq x$ for all $x > -1$. It therefore follows that the upper bound on the right-hand side of (6.13) converges as n tends to infinity almost surely, and therefore so does c_n^k, since $-\log c_n^k$ is non-negative and increasing. We now turn to the bounds in (6.10). Rather than using a Taylor expansion as in (6.11), we simply use that $\log(1+x) \leq x$, to obtain

$$\frac{c_i^k}{c_n^k}\Big(\frac{i}{n}\Big)^{k/\theta} \leq \exp\Big\{k(E(n)-E(i)) + k\sum_{j=i}^{n-1}\frac{S_j - j\mathbb{E}[\mathcal{F}]}{(m_0+j\theta-n_0)(m_0+j-n_0+S_j)}\Big\}, \tag{6.14}$$

where

$$E(n) := \sum_{j=n_0}^{n-1}\frac{1}{m_0+j\theta-n_0} - \frac{1}{\theta}\log n.$$

We rewrite $E(n)$ to find

$$\begin{aligned} E(n) = &\Big(\sum_{j=0}^{n-(n_0+1)}\frac{1}{m_0+\mathbb{E}[\mathcal{F}]\,n_0+j\theta} - \sum_{j=1}^{n-(n_0+1)}\frac{1}{j\theta}\Big)\\ &+\Big(\sum_{j=1}^{n-(n_0+1)}\frac{1}{j\theta} - \frac{1}{\theta}\log(n-(n_0+1))\Big) + \frac{1}{\theta}\log(1-(n_0+1)/n), \end{aligned} \tag{6.15}$$

where we note that the first and second term are decreasing and the final term is increasing in n. Hence, we obtain the upper bound for all $n_0+1 \leq i \leq n$,

$$E(n)-E(i) \leq \frac{1}{\theta}\log\Big(\frac{i}{n}\frac{n-(n_0+1)}{(i-(n_0+1))\vee 1}\Big).$$

Using this inequality and taking the absolute value of the terms in the sum in (6.14), yields the upper bound

$$\exp\Big\{\frac{k}{\theta}\log\Big(\frac{i}{n}\frac{n-(n_0+1)}{(i-(n_0+1))\vee 1}\Big) + \frac{k}{\mathbb{E}[\mathcal{F}]}\sum_{j=i}^{\infty}\frac{|S_j/j-\mathbb{E}[\mathcal{F}]\,|}{m_0+j-n_0+S_j}\Big\},$$

as required. Similarly, we find a lower bound of the same form. As $\log(1+x) \geq x - x^2/2$ for $x \geq 0$, $\exp\{-x\} \geq 1-x$ for $x \in \mathbb{R}$, we find

$$\begin{aligned}\frac{c_i^k}{c_n^k}\Big(\frac{i}{n}\Big)^{k/\theta} &\geq \exp\Big\{-k\sum_{j=i}^{n-1}\frac{|S_j/j-\mathbb{E}[\mathcal{F}]\,|}{\mathbb{E}[\mathcal{F}]\,(m_0+j-n_0+S_j)} - \frac{1}{2}\sum_{j=i}^{n-1}\Big(\frac{k}{m_0+j-n_0+S_j}\Big)^2 \\ &\qquad + k(E(n)-E(i))\Big\} \\ &\geq 1-k\sum_{j=i}^{n-1}\frac{|S_j/j-\mathbb{E}[\mathcal{F}]\,|}{\mathbb{E}[\mathcal{F}]\,(m_0+j-n_0+S_j)} - \frac{1}{2}\sum_{j=i}^{n-1}\Big(\frac{k}{m_0+j-n_0+S_j}\Big)^2 \\ &\quad + k(E(n)-E(i)).\end{aligned} \tag{6.16}$$

Using (6.15) and the fact that $\sum_{j=1}^{n-1}\frac{1}{j} - \log n$ is non-decreasing, we obtain the lower bound

$$\begin{aligned}&1-k\sum_{j=i}^{n-1}\frac{|S_j/j-\mathbb{E}[\mathcal{F}]\,|}{\mathbb{E}[\mathcal{F}]\,(m_0+j-n_0+S_j)} - \frac{1}{2}\sum_{j=i}^{n-1}\Big(\frac{k}{S_j}\Big)^2 - \frac{m_0+\mathbb{E}[\mathcal{F}]\,n_0}{\theta^2}\sum_{j=i-n_0}^{n-(n_0+1)}\frac{1}{j^2} \\ &\quad + \frac{1}{\theta(n-(n_0+1))} - \frac{1}{\theta((i-(n_0+1))\vee 1)} \\ &\geq 1-k\sum_{j=i}^{n-1}\frac{|S_j/j-\mathbb{E}[\mathcal{F}]\,|}{\mathbb{E}[\mathcal{F}]\,(m_0+j-n_0+S_j)} - \frac{1}{2}\sum_{j=i}^{n-1}\Big(\frac{k}{S_j}\Big)^2 - \frac{m_0+\mathbb{E}[\mathcal{F}]\,n_0}{\theta^2}\frac{\pi^2}{6i} \\ &\quad - \frac{1}{\theta((i-(n_0+1))\vee 1)},\end{aligned}$$

which finishes the proof. □

We now prove two results which are used later on to prove parts of Theorem 2.7 (in the weak disorder regime). First, we show that the almost sure limits of certain (super)martingales in Lemma 6.3 do not have an atom at zero:

Lemma 6.5. *For $k \geq 1$, consider the martingales $M_n^k(i)$ for the PAFRO and PAFUD models and $\widetilde{M}_n^k(i)$ for the PAFFD model as in Lemma 6.3 and their almost sure limits $\xi_i^k, \widetilde{\xi}_i^k$, respectively. Then, the $\xi_i^k, \widetilde{\xi}_i^k$ do not have an atom at zero.*

Proof. We first focus on the martingales M_n^k for the PAFRO and PAFUD models. Let $\varepsilon > 0$. We can write,

$$\begin{aligned}\mathbb{P}_{\mathcal{F}}\big(\xi_i^k < \varepsilon\big) &= \lim_{n\to\infty}\mathbb{P}_{\mathcal{F}}\bigg(c_n^k\binom{\mathcal{Z}_n(i)+\mathcal{F}_i+(k-1)}{k} < \varepsilon\bigg) \\ &\leq \lim_{n\to\infty}\mathbb{P}_{\mathcal{F}}\big(c_n^k(\mathcal{Z}_n(i)+\mathcal{F}_i)^k < \varepsilon\Gamma(k+1)\big),\end{aligned} \tag{6.17}$$

since $x^k \leq \Gamma(x+k)/\Gamma(x)$ for $k \geq 1, x > 0$, by [16, Theorem 1]. Now, take $p \in (-\min(\mathcal{F}_i,1)/k, 0)$. The goal is to raise both sides to the power p and use a Markov bound. We first, however, need some other inequalities to obtain useful expressions. Using the concavity of $\log x$ and noting that $x+pk$ is a weighted average of x and $x+k$ when $p \in (0,1)$ and $x+k$ is a weighted average of x and $x+pk$ when $p \geq 1$, we obtain, for all $x, k \geq 0$,

$$\Big(1-\frac{k}{x+k}\Big)^p \geq 1-\frac{pk}{x+pk} \text{ when } p\in(0,1), \Big(1-\frac{k}{x+k}\Big)^p \leq 1-\frac{pk}{x+pk} \text{ when } p \geq 1. \tag{6.18}$$

From the first inequality, we also immediately obtain, for $p \in (-1,0), k \geq 0, x \geq k|p|$,

$$\Big(1-\frac{k}{x+k}\Big)^p \leq 1-\frac{pk}{x+pk}. \tag{6.19}$$

It thus follows that, when $p \in (-\min(\mathcal{F}_i,1)/k,0)$, $(c_n^k)^p \le c_n^{kp}$, as $\mathcal{F}_i > k|p|$. Also, from [25] it follows that for all $x \ge 0, s \in (0,1)$,

$$x^s \ge \frac{\Gamma(x+s)}{\Gamma(x)}.$$

Hence, since $\Gamma(x)/\Gamma(x+s)$ is decreasing in x for $s \ge 0$, when $p \in (-1,0), x \ge |p|$,

$$x^p \le \frac{\Gamma(x+p)}{\Gamma(x)}, \tag{6.20}$$

so that, combining both (6.19) and (6.20) in (6.17) with $p \in (-\min(\mathcal{F}_i/k,1/k),0)$, yields

$$\begin{aligned} \mathbb{P}_{\mathcal{F}}\big(c_n^k(\mathcal{Z}_n(i)+\mathcal{F}_i)^k < \varepsilon\Gamma(k+1)\big) &\le \mathbb{P}_{\mathcal{F}}\big(M_n^{kp}(i) \ge \varepsilon^p\Gamma(k+1)^p/\Gamma(kp+1)\big) \\ &\le \mathbb{E}_{\mathcal{F}}[M_n^{kp}(i)](\varepsilon\Gamma(k+1))^{|p|}\Gamma(kp+1) \\ &= M_{i\vee n_0}^{kp}(i)\varepsilon^{|p|}\Gamma(k+1)^{|p|}\Gamma(pk+1), \end{aligned} \tag{6.21}$$

which is finite almost surely and tends to zero with ε almost surely. We can thus first take the limit of n to infinity, and then let ε tend to zero. Hence, almost surely,

$$\mathbb{P}_{\mathcal{F}}\big(\xi_i^k = 0\big) = \lim_{\varepsilon\downarrow 0}\mathbb{P}_{\mathcal{F}}\big(\xi_i^k < \varepsilon\big) = 0,$$

and thus $\mathbb{P}(\xi_i^1 = 0) = 0$, by the dominated convergence theorem. For the PAFFD model, an altered argument is required, since $\widetilde{M}_n^k(i)$ is a submartingale for negative k, as follows from Lemma 6.3 so that the final steps in (6.21) no longer work. Rather, we only follow the same steps for $\widetilde{\xi}_i^k$ in (6.17). Then, let us define, for a large constant $C > 0$, $\eta \in (0, \mathbb{E}[\mathcal{F}]/(\mathbb{E}[\mathcal{F}]+m))$ and a large integer $N \ge i \vee n_0$, the stopping time $T_N := \inf\{n \ge N : \mathcal{Z}_n(i) \ge Cn^{1-\eta}\}$. We aim to show that we can construct a sequence $\hat{c}_n^k$, to be defined later, such that

$$\hat{M}_{T_N\wedge n}^k(i) := \hat{c}_{T_N\wedge n}^k \binom{\mathcal{Z}_{T_N\wedge n}(i)+\mathcal{F}_i+(k-1)}{k}$$

is a supermartingale for $k \in (-\min(\mathcal{F}_i,1),0)$ for the PAFFD model. First, recall the computations in (6.7). We notice that the product in the second line contains terms which are positive but less than 1 when $k \in (-\min(\mathcal{F}_i,1),0)$. Therefore, the product decreases as the number of terms increases, so that we can bound the expected value from above by $1 + k\mathbb{P}(\Delta\mathcal{Z}_n(i) \ge 1 \,|\, \mathcal{G}_n)/(\mathcal{Z}_n(i)+\mathcal{F}_i)$. If we define

$$\hat{c}_n^k := \prod_{j=n_0}^{n-1}\Big(1 - \frac{kma_j}{m_0+m(j-n_0)+S_j+kma_j}\Big),\ a_n := 1 - \frac{m-1}{2}\frac{Cn^{-\eta}+\mathcal{F}_i/n}{(m_0+m(n-n_0)+S_n)/n},$$

we obtain

$$\begin{aligned} &\mathbb{E}_{\mathcal{F}}\big[\hat{M}_{T_N\wedge(n+1)}^k(i)\mathbb{1}_{\{T_N \ge n+1\}}\,\big|\,\mathcal{G}_n\big] \\ &\quad\le \hat{M}_n^k(i)\Big(1 - \frac{kma_n}{m_0+m(n-n_0)+S_n+kma_n}\Big)\Big(1 + k\frac{\mathbb{P}(\Delta\mathcal{Z}_n(i) \ge 1 \,|\, \mathcal{G}_n)}{\mathcal{Z}_n(i)+\mathcal{F}_i}\Big)\mathbb{1}_{\{T_N \ge n+1\}}. \end{aligned}$$

We now bound $\mathbb{P}(\Delta\mathcal{Z}_n(i) \ge 1 \,|\, \mathcal{G}_n)$ from below, using that $1-(1-x)^m \ge mx - m(m-1)x^2/2$ for all $x \in (0,1), m \in \mathbb{N}$. Then, on $\{T_N \ge n+1\}$, we can bound $\mathcal{Z}_n(i)$ from above by $Cn^{1-\eta}$, which yields the upper bound

$$\begin{aligned} &\hat{M}_n^k(i)\Big(1 - \frac{kma_n}{m_0+m(n-n_0)+S_n+kma_n}\Big)\Big(1 + \frac{kma_n}{m_0+m(n-n_0)+S_n}\Big)\mathbb{1}_{\{T_N \ge n+1\}} \\ &\qquad = \hat{M}_n^k(i)\mathbb{1}_{\{T_N \ge n+1\}} = \hat{M}_{T_N\wedge n}^k(i)\mathbb{1}_{\{T_N \ge n+1\}}. \end{aligned}$$

Finally, as the event $\{T_N \le n\}$ is $\mathcal{G}_n$ measurable,

$$\mathbb{E}_{\mathcal{F}}\big[M^k_{T_N\wedge(n+1)}(i)\mathbb{1}_{\{T_N\le n\}}\,\big|\,\mathcal{G}_n\big] = \hat{M}^k_{T_N}(i)\mathbb{1}_{\{T_N\le n\}} = \hat{M}^k_{T_N\wedge n}(i)\mathbb{1}_{\{T_N\le n\}}.$$

Together with the computations above, this yields

$$\mathbb{E}_{\mathcal{F}}\big[\hat{M}^k_{T_N\wedge(n+1)}(i)\,\big|\,\mathcal{G}_n\big] \le \hat{M}^k_{T_N\wedge n}(i),$$

which shows indeed that $\hat{M}^k_{T_N\wedge n}(i)$ is a supermartingale for $k\in(-\min(\mathcal{F}_i,1),0)$. It also follows relatively easily, following similar steps as in the proof of Lemma 6.4, that $\hat{c}^k_n n^{k/\theta_m}\xrightarrow{a.s.}\hat{c}_k$ for some random variable $\hat{c}_k$ as n tends to infinity. So, we can then write, for $k\ge 1, p\in(-\min(\mathcal{F}_i/k,1/k),0)$, continuing the steps in (6.17) and using (6.19) and (6.20) as in (6.21),

$$\mathbb{P}_{\mathcal{F}}(\widetilde{\xi}^k_i<\varepsilon)\le\lim_{n\to\infty}\mathbb{P}_{\mathcal{F}}((c^{kp}_n/\hat{c}^{kp}_n)\hat{M}^{kp}_n(i)\ge\varepsilon^p\Gamma(k+1)^p/\Gamma(kp+1)).$$

We now intersect with the event $\{T_N\ge n+1\}$ and its complement to obtain the upper bound

$$\begin{aligned}&\lim_{n\to\infty}\mathbb{P}_{\mathcal{F}}\Big(\{(c^{kp}_n/\hat{c}^{kp}_n)\hat{M}^{kp}_n(i)>\varepsilon^p\Gamma(k+1)^p/\Gamma(kp+1)\}\cap\{T_N\ge n+1\}\Big)+\mathbb{P}_{\mathcal{F}}(T_N\le n)\\ &\quad\le\lim_{n\to\infty}\mathbb{P}_{\mathcal{F}}\Big((c^{kp}_n/\hat{c}^{kp}_n)\hat{M}^{kp}_{T_N\wedge n}(i)>\varepsilon^p\Gamma(k+1)^p/\Gamma(kp+1)\Big)+\mathbb{P}_{\mathcal{F}}(T_N\le n)\,.\end{aligned}$$

Using the Markov inequality for the first probability and because $\hat{M}^{kp}_{T_N\wedge n}(i)$ is a supermartingale since $kp\in(-\min(\mathcal{F}_i,1),0)$, we find the upper bound

$$\begin{aligned}&\lim_{n\to\infty}(c^{kp}_n/\hat{c}^{kp}_n)\varepsilon^{|p|}\mathbb{E}_{\mathcal{F}}[\hat{M}^{kp}_{T_N\wedge n}(i)]\Gamma(k+1)^{|p|}\Gamma(kp+1)+\mathbb{P}_{\mathcal{F}}(T_N\le n)\\ &\quad\le(c_{kp}/\hat{c}_{kp})\varepsilon^{|p|}\mathbb{E}_{\mathcal{F}}[\hat{M}^{kp}_N(i)]\Gamma(k+1)^{|p|}\Gamma(kp+1)+\lim_{n\to\infty}\mathbb{P}_{\mathcal{F}}(T_N\le n)\,.\end{aligned}\tag{6.22}$$

We note that the first term tends to zero with ε. For the second probability we write, for some s^{th} moment bound, with $s>(\mathbb{E}[\mathcal{F}]/(\mathbb{E}[\mathcal{F}]+m)-\eta)^{-1}$,

$$\begin{aligned}\mathbb{P}_{\mathcal{F}}(T_N\le n)&\le\sum_{j=N}^n\mathbb{P}_{\mathcal{F}}\big((\mathcal{Z}_j(i)+\mathcal{F}_i)^s\ge C^sj^{s(1-\eta)}\big)\\ &\le\frac{\Gamma(k+1)}{C^s}\sum_{j=N}^n(\widetilde{c}^s_j)^{-1}j^{-s(1-\eta)}\mathbb{E}_{\mathcal{F}}[\widetilde{M}^s_j(i)].\end{aligned}$$

Using the upper bound for $c^s_{n_0}/c^s_n=1/c^s_n$ in (6.10), we find the upper bound

$$C_{k,s}A\widetilde{M}^s_{i\vee n_0}(i)\sum_{j=N}^n j^{s(1/\theta_m-(1-\eta))}\le\widetilde{C}_{k,s}A\widetilde{M}^s_{i\vee n_0}(i)N^{1-s(\mathbb{E}[\mathcal{F}]/(\mathbb{E}[\mathcal{F}]+m)-\eta)},$$

where A equals the upper bound in (6.10) with $i=n_0$. This upper bound is independent of n, so we find, combining this with (6.22),

$$\lim_{\varepsilon\downarrow 0}\mathbb{P}_{\mathcal{F}}(\widetilde{\xi}^k_i<\varepsilon)\le\widetilde{C}_{k,s}A\widetilde{M}^s_{i\vee n_0}(i)N^{1-s(\mathbb{E}[\mathcal{F}]/(\mathbb{E}[\mathcal{F}]+m)-\eta)},$$

where the right-hand side tends to zero almost surely as N tends to infinity, by the choice of s. Thus, it follows that $\lim_{\varepsilon\downarrow 0}\mathbb{P}_{\mathcal{F}}(\widetilde{\xi}^k_i<\varepsilon)=0$ for all $k\ge 1$. Again, using the dominated convergence theorem finally yields the required result. □

As a final result describing the behaviour of the martingales $M^k_n(i)$, we show that, for particular values of k, these martingales are small when i is large.

Lemma 6.6. *Consider the martingales (resp. supermartingales) $M_n^k(i)$ (resp. $\widetilde{M}_n^k(i)$) as in Lemma 6.3. Let $M := \sup\{s \geq 1 : \mathbb{E}[\mathcal{F}^s] < \infty\}$ and assume that $M > \theta_m$. Then, for all $m \in \mathbb{N}, k \in (\theta_m, M)$, almost surely*

$$\lim_{i\to\infty} \sup_{n\geq n_0\vee i} M_n^k(i) = 0, \qquad \lim_{i\to\infty} \sup_{n\geq n_0\vee i} \widetilde{M}_n^k(i) = 0. \tag{6.23}$$

Proof. We note that the first result is implied if, for any $\varepsilon > 0$,

$$\mathbb{P}\Big(\sup_{n\geq i\vee n_0} M_n^k(i) \geq \varepsilon \text{ for infinitely many } i\Big) = 0,$$

and similarly for $\widetilde{M}_n^k(i)$. We now use the 'good' event $E_\ell(\delta) := \{|S_j/j - \mathbb{E}[\mathcal{F}]\,| \leq \delta\ \forall j \geq \ell\}$, where we take $\delta > 0$ sufficiently small such that $k \in (\theta_m(1+\delta), M)$. That is, we intersect with $E_\ell(\delta)$ and $E_\ell(\delta)^c$. By writing i.o. for 'infinitely often', we find

$$\begin{aligned}\mathbb{P}\Big(\sup_{n\geq i\vee n_0} M_n^k(i) \geq \varepsilon \text{ i.o.}\Big) &\leq \mathbb{P}\Big(\Big\{\sup_{n\geq i\vee n_0} M_n^k(i) \geq \varepsilon \text{ i.o.}\Big\} \cap E_\ell(\delta)\Big) + \mathbb{P}(E_\ell(\delta)^c)\\ &= \mathbb{P}\Big(\mathbb{1}_{E_\ell(\delta)} \sum_{i=1}^\infty \mathbb{1}_{A_i} = \infty\Big) + \mathbb{P}(E_\ell(\delta)^c)\,,\end{aligned} \tag{6.24}$$

where $A_i := \{\sup_{n\geq i\vee n_0} M_n^k(i) \geq \varepsilon\}$. We now show that the first probability on the right-hand side equals 0 for every $\ell \in \mathbb{N}$, by showing the sum of indicators has a finite mean. We write

$$\mathbb{E}\Big[\mathbb{1}_{E_\ell(\delta)} \sum_{i=1}^\infty \mathbb{1}_{A_i}\Big] = \mathbb{E}\Big[\mathbb{1}_{E_\ell(\delta)} \mathbb{E}_{\mathcal{F}}\Big[\sum_{i=1}^\infty \mathbb{1}_{A_i}\Big]\Big], \tag{6.25}$$

and first deal with the conditional expectation. We apply Doob's martingale inequality [22, Theorem II 1.7] to the events A_i to find

$$\mathbb{P}_{\mathcal{F}}(A_i) = \lim_{N\to\infty} \mathbb{P}_{\mathcal{F}}\Big(\sup_{i\vee n_0\leq n\leq N} M_n^k(i) \geq \varepsilon\Big) \leq \lim_{N\to\infty} \frac{1}{\varepsilon}\mathbb{E}_{\mathcal{F}}[M_N^k(i)] = \frac{1}{\varepsilon}\mathbb{E}_{\mathcal{F}}[M_{i\vee n_0}^k(i)], \tag{6.26}$$

where the first step holds by the monotonicity of the events $\{\sup_{i\vee n_0\leq n\leq N} M_n^k(i) \geq \varepsilon\}$.

Doob's martingale inequality holds for submartingales only, though. However, we can still prove the same upper bound for $\widetilde{M}_n^k(i)$, but a different technique is required. We define the stopping time $\tau_\varepsilon := \inf\{n \geq i \vee n_0 \,|\, \widetilde{M}_n^k(i) \geq \varepsilon\}$. Then, for any $N \in \mathbb{N}$,

$$\mathbb{P}_{\mathcal{F}}\Big(\sup_{i\vee n_0\leq n\leq N} \widetilde{M}_n^k(i) \geq \varepsilon\Big) = \mathbb{P}_{\mathcal{F}}(\tau_\varepsilon \leq N) = \mathbb{P}_{\mathcal{F}}\Big(\mathbb{1}_{\{\tau_\varepsilon\leq N\}} \widetilde{M}_{\tau_\varepsilon}^k(i) \geq \varepsilon\Big),$$

so that using Markov's inequality yields the upper bound

$$\frac{1}{\varepsilon}\mathbb{E}_{\mathcal{F}}[\mathbb{1}_{\{\tau_\varepsilon\leq N\}}\widetilde{M}_{\tau_\varepsilon}^k(i)] \leq \frac{1}{\varepsilon}\Big(\mathbb{E}_{\mathcal{F}}[\mathbb{1}_{\{\tau_\varepsilon\leq N\}}\widetilde{M}_{\tau_\varepsilon}^k(i)] + \mathbb{E}_{\mathcal{F}}[\mathbb{1}_{\{\tau_\varepsilon> N\}}\widetilde{M}_N^k(i)]\Big) = \frac{1}{\varepsilon}\mathbb{E}_{\mathcal{F}}[\widetilde{M}_{\tau_\varepsilon\wedge N}^k(i)],$$

see also [22, Exercise 1.25, Chapter II]. We now use the optional sampling theorem [26, Theorem 10.10], which yields the required upper bound. Again, by monotonicity and taking N to infinity we obtain the same result. Using (6.26) in (6.25) and recalling $M_n^k(i)$ from Lemma 6.3 yields the upper bound

$$\mathbb{E}\Big[\mathbb{1}_{E_\ell(\delta)} \sum_{i=1}^\infty \varepsilon^{-1} c_{i\vee n_0}^k \binom{\mathcal{Z}_{i\vee n_0}(i) + \mathcal{F}_i + (k-1)}{k}\Big].$$

Note that, for $i > n_0$, $\mathcal{Z}_{i\vee n_0}(i) = 0$ and for $i \in [n_0]$, $\mathcal{Z}_{i\vee n_0}(i) = \mathcal{Z}_{n_0}(i) \leq \sum_{j=1}^{n_0} \mathcal{Z}_{n_0}(i) = m_0$. Also, for $i \geq \ell \vee n_0$ and on $E_\ell(\delta)$, we can bound $c_{i\vee n_0}^k$ from above by $Ci^{-k/(\theta_m(1+\delta))}$ for some large constant $C>0$. For $i\in[(\ell\vee n_0)-1]$, we can just bound $c^k_{i\vee n_0}(i)$ from above by 1. This yields the upper bound

$$C\sum_{i=\ell\vee n_0}^{\infty}\mathbb{E}\Big[\mathbb{1}_{E_\ell(\delta)}i^{-k/(\theta_m(1+\delta))}\binom{\mathcal{F}_i+(k-1)+m_0}{k}\Big]+\sum_{i=1}^{(\ell\vee n_0)-1}\mathbb{E}\Big[\binom{\mathcal{F}_i+(k-1)+m_0}{k}\Big]$$
$$\le \widetilde{C}(1+\mathbb{E}\big[\mathcal{F}^k\big])\sum_{i=\ell\vee n_0}^{\infty}i^{-k/(\theta_m(1+\delta))}+\widetilde{C}(1+\mathbb{E}\big[\mathcal{F}^k\big])(\ell\vee n_0),$$

which is finite by the choice of k and δ. We note that we can indeed bound the mean of $\binom{\mathcal{F}+(k-1)+m_0}{k}$ by a constant times 1 plus the k^{th} moment of $\mathcal{F}$. Namely, using the asymptotics of the Gamma function,

$$\begin{aligned}\mathbb{E}\Big[\binom{\mathcal{F}+(k-1)+m_0}{k}\Big]&=\int_0^\infty\binom{x+(k-1)+m_0}{k}\mu(\mathrm{d}x)\\&\le\int_0^{x^*}\binom{x+(k-1)+m_0}{k}\mu(\mathrm{d}x)+C_1\int_{x^*}^\infty x^k\mu(\mathrm{d}x)\\&\le C_2(1+\mathbb{E}\big[\mathcal{F}^k\big]),\end{aligned}$$

with $C_2:=\max\{C_1,\int_0^{x^*}\binom{x+(k-1)+m_0}{k}\mu(\mathrm{d}x)\}$ and x^* such that for $x\ge x^*$, $\binom{x+(k-1)+m_0}{k}\le C_1x^k$. It follows that the mean in (6.25) is finite and thus that the first probability on the right-hand side of (6.24) equals 0. Hence,

$$\mathbb{P}\Big(\sup_{n\ge i\vee n_0}M_n^k(i)\ge\varepsilon\text{ i.o.}\Big)\le\mathbb{P}(E_\ell(\delta)^c)\,,$$

which tends to 0 as $\ell\to\infty$ by the strong law of large numbers, and so we obtain (6.23). $\square$

6.2 The maximum conditional mean degree in the strong and extreme disorder regime

We now use the martingales studied in the previous subsection, specifically Lemmas 6.3 and 6.4, as well as the results attained in Section 5 to prove Propositions 6.1 and 6.2.

First, though, we state a final result from [1], which provides conditions such that the maximum of a double array converges to a certain limit:

Proposition 6.7. *[1, Proposition 3.1] Let $\{a_{n,i}:i\in[n]\}_{n\ge1}$ be a double array of non-negative numbers such that*

1. *For all $i\ge1,\lim_{n\to\infty}a_{n,i}=a_i<\infty$,*
2. *$\sup_{n\ge1}a_{n,i}\le b_i<\infty$,*
3. *$\lim_{i\to\infty}b_i=0$,*
4. *For $i\ne j$, $a_i\ne a_j$.*

Then,

- *$\max_{i\in[n]}a_{n,i}\to\max_{i\ge1}a_i$, as $n\to\infty$.*
- *In addition, there exist I_0 and N_0 such that $\max_{i\in[n]}a_{n,i}=a_{n,I_0}$ for all $n\ge N_0$.*

We now prove Proposition 6.1:

Proof of Proposition 6.1. The focus of the proof is on the PAFUD model. The proof for the PAFRO model follows by setting $m=1$, the proof for the PAFFD model follows in the same way, as we only look at the mean of $M_n^1(i)$, which by Lemma 6.3 is a martingale for both the PAFUD *and* PAFFD model.

We start by proving (6.1). Take $\alpha\in(2,1+\theta_m)$. Using Lemma 6.3, it directly follows that

$$\mathbb{E}_{\mathcal{F}}[\mathcal{Z}_n(i)]=(c_n^1(m))^{-1}\mathbb{E}_{\mathcal{F}}[M_n^1(i)]-\mathcal{F}_i=\frac{c^1_{i\vee n_0}(m)}{c_n^1(m)}\mathcal{Z}_{i\vee n_0}(i)+\mathcal{F}_i\Big(\frac{c^1_{i\vee n_0}(m)}{c_n^1(m)}-1\Big). \quad (6.27)$$

Note that for $i\geq n_0$ the first term on the right-hand side equals zero. We can then construct the inequalities

$$\max_{i\in[n]}\frac{\mathcal{F}_i}{u_n}\Big(\frac{c^1_{i\vee n_0}}{c_n^1}-1\Big)\leq \max_{i\in[n]}\mathbb{E}_{\mathcal{F}}[\mathcal{Z}_n(i)/u_n]\leq \max_{i\in[n]}\frac{\mathcal{F}_i}{u_n}\Big(\frac{c^1_{i\vee n_0}}{c_n^1}-1\Big)+\frac{m_0}{u_nc_n^1}.$$

By Lemma 6.4, the last term on the right-hand side tends to zero almost surely, as $\alpha-1<\theta_m$. That is, since $u_n=\widetilde{\ell}(n)n^{1/(\alpha-1)}$ for some slowly-varying function $\widetilde{\ell}$, and $c_n^1(m)n^{1/\theta_m}$ converges almost surely, the fact that $\alpha-1<\theta_m$ yields that $u_nc_n^1$ diverges to ∞ almost surely.

By the reverse triangle inequality, it follows that for $x,y\in\mathbb{R}_+^n$,

$$|\max_{i\in[n]}x_i-\max_{i\in[n]}y_i|=|\|x\|_\infty-\|y\|_\infty|\leq\|x-y\|_\infty=\max_{i\in[n]}|x_i-y_i|. \quad (6.28)$$

So, as $c^1_{i\vee n_0}=c_i^1$ for all $i\geq n_0$,

$$\Big|\max_{i\in[n]}\frac{\mathcal{F}_i}{u_n}\Big(\frac{c^1_{i\vee n_0}}{c_n^1}-1\Big)-\max_{i\in[n]}\frac{\mathcal{F}_i}{u_n}\Big(\frac{c^1_{i}}{c_n^1}-1\Big)\Big|\leq \max_{i\in[n]}\frac{\mathcal{F}_i}{u_n}\frac{c_i^1-c^1_{i\vee n_0}}{c_n^1}=\max_{i<n_0}\frac{\mathcal{F}_i}{u_n}\frac{c_i^1-c^1_{n_0}}{c_n^1},$$

which again tends to zero almost surely by Lemma 6.4, as it is a maximum over a finite number of terms. Therefore, assuming the limits exist, it follows that

$$\lim_{n\to\infty}\max_{i\in[n]}\mathbb{E}_{\mathcal{F}}[\mathcal{Z}_n(i)/u_n]=\lim_{n\to\infty}\max_{i\in[n]}\frac{\mathcal{F}_i}{u_n}\Big(\frac{c_i^1}{c_n^1}-1\Big) \quad (6.29)$$

almost surely. We now show that

$$\Big|\max_{i\in[n]}\frac{\mathcal{F}_i}{u_n}\Big(\frac{c_i^1}{c_n^1}-1\Big)-\max_{i\in[n]}\frac{\mathcal{F}_i}{u_n}\Big(\Big(\frac{i}{n}\Big)^{-1/\theta_m}-1\Big)\Big|\xrightarrow{\mathbb{P}}0. \quad (6.30)$$

Using (6.28) we find

$$\Big|\max_{i\in[n]}\frac{\mathcal{F}_i}{u_n}\Big(\frac{c_i^1}{c_n^1}-1\Big)-\max_{i\in[n]}\frac{\mathcal{F}_i}{u_n}\Big(\Big(\frac{i}{n}\Big)^{-1/\theta_m}-1\Big)\Big|\leq\max_{i\in[n]}\frac{\mathcal{F}_i}{u_n}\Big(\frac{n}{i}\Big)^{1/\theta_m}\Big|\frac{c_i^1}{c_n^1}\Big(\frac{i}{n}\Big)^{-1/\theta_m}-1\Big|.$$

Then, let $\eta\in(1,(\alpha-2)\wedge 1)$ and let $(\varepsilon_n)_{n\in\mathbb{N}}$ be a sequence such that $\varepsilon_n:=n^{-\beta}$, with $\beta\in(0,\theta_m\eta/(1+(1+\theta_m)\eta))$. We split the maximum into two parts: indices i which are at most ε_nn and at least ε_nn and deal with these separately. (Note that $\beta<1$ and thus $\varepsilon_nn\to\infty$.) We first define, for $A\subseteq[n]$ and $\delta>0$,

$$E_A:=\Big\{\max_{i\in A}\frac{\mathcal{F}_i}{u_n}\Big(\frac{n}{i}\Big)^{1/\theta_m}\Big|\frac{c_i^1}{c_n^1}\Big(\frac{i}{n}\Big)^{1/\theta_m}-1\Big|>\delta\Big\}. \quad (6.31)$$

This yields

$$\mathbb{P}\big(E_{[n]}\big)\leq\mathbb{P}\big(E_{[\varepsilon_nn]}\big)+\mathbb{P}\big(E_{[n]\setminus[\varepsilon_nn]}\big). \quad (6.32)$$

We first investigate the latter probability. We write,

$$\mathbb{P}\big(E_{[n]\setminus[\varepsilon_n n]}\big) \le \mathbb{P}\bigg(\Big(\max_{i>\varepsilon_n n}\frac{\mathcal{F}_i}{u_n}\Big)\varepsilon_n^{-1/\theta_m}\max_{i>\varepsilon_n n}\bigg|\frac{c_i^1}{c_n^1}\Big(\frac{i}{n}\Big)^{1/\theta_m}-1\bigg|>\delta\bigg), \tag{6.33}$$

where we bound the $(n/i)^{1/\theta_m}$, as in the definition of E_A in (6.31), from above by $\varepsilon_n^{-1/\theta_m}$ and take the maximum over the fitness variables and the absolute value separately. Since the number of terms in $\max_{i>\varepsilon_n n}\mathcal{F}_i/u_n$ is asymptotically n, that is, $(n-\varepsilon_n n)/n = 1-\varepsilon_n = 1-o(1)$, it follows that the first maximum on the right-hand side converges in distribution. For the second maximum in (6.33), when $i\ge\varepsilon_n n$, the terms in the absolute value should be small due to the almost sure convergence of $c_n^1 n^{1/\theta_m}$ and $c_i^1 i^{1/\theta_m}$ because of Lemma 6.4 (note that $i>\varepsilon_n n$ so that i tends to infinity with n). We show a slightly stronger result, namely that

$$\varepsilon_n^{-1/\theta_m}\max_{i>\varepsilon_n n}\bigg|\frac{c_i^1}{c_n^1}\Big(\frac{i}{n}\Big)^{1/\theta_m}-1\bigg|\xrightarrow{\mathbb{P}}0.$$

To prove this, we use the upper and lower bound in (6.10). The upper bound, when considering $\varepsilon_n n\le i\le n$, is largest for $i=\varepsilon_n n$. Thus, we have a uniform upper bound for all $\varepsilon_n n\le i\le n$,

$$\frac{c_i^1}{c_n^1}\Big(\frac{i}{n}\Big)^{1/\theta_m} \le \exp\bigg\{\frac{k}{\theta_m}\log\Big(\varepsilon_n\frac{n-(n_0+1)}{\varepsilon_n n-(n_0+1)}\Big)+\frac{mk}{\mathbb{E}[\mathcal{F}]}\sum_{j=\varepsilon_n n}^{\infty}\frac{|S_j/j-\mathbb{E}[\mathcal{F}]\,|}{m_0+m(j-n_0)+S_j}\bigg\}.$$

For n large, the denominator in the sum can be bounded from below by $mj/2$ and the term in the logarithm can be bounded from above by $1+2(n_0+1)/(\varepsilon_n n)$. Hence, we obtain the upper bound

$$\frac{c_i^1}{c_n^1}\Big(\frac{i}{n}\Big)^{1/\theta_m} \le \exp\bigg\{\frac{k}{\theta_m}\log\Big(1+\frac{2(n_0+1)}{\varepsilon_n n}\Big)+\frac{2k}{\mathbb{E}[\mathcal{F}]}\sum_{j=\varepsilon_n n}^{\infty}\frac{|S_j/j-\mathbb{E}[\mathcal{F}]\,|}{j}\bigg\}.$$

Similarly, the lower bound in (6.16) is largest when $i=n-1$ (note that the second maximum in (6.33) is never attained at $i=n$, so we can ignore this case), from which we obtain

$$\max_{i\ge\varepsilon_n n}\bigg(\frac{c_i^1}{c_n^1}\Big(\frac{i}{n}\Big)^{1/\theta_m}-1\bigg) \ge -\frac{m_0+\mathbb{E}[\mathcal{F}]\,n_0+(m-1)}{\theta_m^2}\frac{\pi^2}{6(n-1)}-\frac{1}{\theta_m(n-(n_0+2))}\ge -\frac{C}{n},$$

for some constant $C>0$. It then follows that, as $\varepsilon_n^{-1/\theta_m}=n^{\beta/\theta_m}\ge 1$, $a(\mathrm{e}^x-1)\le \mathrm{e}^{ax}-1$ for all $x\in\mathbb{R}$ when $a\ge 1$,

$$\begin{aligned}
&\varepsilon_n^{-1/\theta_m}\max_{i\ge\varepsilon_n n}\bigg|\frac{c_i^1}{c_n^1}\Big(\frac{i}{n}\Big)^{1/\theta_m}-1\bigg| \\
&\quad\le \max\bigg\{\frac{C}{n^{1-\beta/\theta_m}},\exp\bigg\{\frac{k}{\theta_m}\log\bigg(\Big(1+\frac{2(n_0+1)}{n^{1-\beta}}\Big)^{n^{\beta/\theta_m}}\bigg) \\
&\qquad+\frac{2k}{\mathbb{E}[\mathcal{F}]}n^{\beta/\theta_m}\sum_{j=\varepsilon_n n}^{\infty}\frac{|S_j/j-\mathbb{E}[\mathcal{F}]\,|}{j}\bigg\}-1\bigg\}.
\end{aligned}\tag{6.34}$$

Clearly, the first argument tends to zero, as $\beta<\theta_m$. What remains to prove is that the second argument of the maximum on the right-hand side of (6.34) converges to zero in probability. The first term in the exponent tends to zero, as $1-\beta>\beta/\theta_m$ by the choice of β. For the second term, using Markov's inequality, for any $\delta>0$,

$$\begin{aligned}
\mathbb{P}\bigg(n^{\beta/\theta_m}\sum_{j=\varepsilon_n n}^{\infty}\frac{|S_j/j-\mathbb{E}[\mathcal{F}]\,|}{j}>\delta\bigg) &\le \delta^{-1}n^{\beta/\theta_m}\sum_{j=\varepsilon_n n}^{\infty}j^{-2}\mathbb{E}[|S_j-j\mathbb{E}[\mathcal{F}]\,|] \\
&\le \delta^{-1}n^{\beta/\theta_m}\sum_{j=\varepsilon_n n}^{\infty}j^{-2}\mathbb{E}\big[|S_j-j\mathbb{E}[\mathcal{F}]\,|^{1+\eta}\big]^{1/(1+\eta)},
\end{aligned}$$

where we note that $\eta \in (0, (\alpha-2)\wedge 2)$, such that we can apply the Marcinkiewicz-Zygmund inequality as in (6.12). This yields, for some constant $C > 0$, the upper bound

$$Cn^{\beta/\theta_m} \sum_{j=\varepsilon_n n}^{\infty} j^{-2+1/(1+\eta)} \leq \widetilde{C} n^{\beta(\eta/(1+\eta)+1/\theta_m)-\eta/(1+\eta)},$$

which tends to zero by the choice of β. It now follows that the right-hand side of (6.34) tends to zero in probability. This implies, using Slutsky's theorem [24], that for any $\delta > 0$,

$$\lim_{n\to\infty} \mathbb{P}\big(E_{[n]\setminus[\varepsilon_n n]}\big) = 0. \tag{6.35}$$

For the first probability on the right-hand side of (6.32), we show that $\max_{i\leq\varepsilon_n n}(\mathcal{F}_i/u_n)(n/i)^{1/\theta_m}$ tends to zero in probability when n tends to infinity and that $\max_{i\leq\varepsilon_n n}|(c_i^1/c_n^1)(i/n)^{1/\theta_m} - 1|$ converges almost surely. We focus on the former first. The claim is proved by using the Poisson Point Process (PPP) weak limit. Recall Π_n in (5.1) and its weak limit Π. We write

$$\Pi_n = \sum_{i=1}^{n} \delta_{(i/n,\mathcal{F}_i/u_n)} \Rightarrow \sum_{i\geq 1} \delta_{(t_i,f_i)} =: \Pi \quad \text{in } M_p(E), \tag{6.36}$$

where δ is a Dirac measure, and Π is a PPP on $(0,1)\times(0,\infty)$ with intensity measure $\nu(\mathrm{d}t,\mathrm{d}x) := \mathrm{d}t \times (\alpha-1)x^{-\alpha}\mathrm{d}x$ [21, Corollary 4.19]. We now define Π' to be the PPP on $\mathbb{R}_+$ obtained from mapping points $(t,f)\in\Pi$ to ft^{-1/θ_m} and let Π'_ε be the restriction of Π' to points (t,f) such that $t\leq\varepsilon$. More formally,

$$\Pi' := \sum_{(t,f)\in\Pi} \delta_{(ft^{-1/\theta_m})}, \qquad \Pi'_\varepsilon := \sum_{(t,f)\in\Pi} \mathbb{1}_{\{t\leq\varepsilon\}} \delta_{(ft^{-1/\theta_m})}.$$

Now, we fix an arbitrary $\delta, \eta > 0$. Then, we can find an $\varepsilon > 0$ sufficiently small, such that

$$\begin{aligned}\mathbb{P}\bigg(\max_{(t,f)\in\Pi:t\leq\varepsilon} ft^{-1/\theta_m} > \delta\bigg) &= 1 - \mathbb{P}(\Pi'_\varepsilon((\delta,\infty)) = 0)\\ &= 1 - \exp\Big\{\int_0^\varepsilon \int_{\delta t^{1/\theta_m}}^\infty (\alpha-1)f^{-\alpha}\mathrm{d}f\mathrm{d}t\Big\}\\ &= 1 - \exp\Big\{-\frac{\theta_m}{\theta_m-(\alpha-1)}\delta^{-(\alpha-1)}\varepsilon^{(\theta_m-(\alpha-1))/\theta_m}\Big\}\end{aligned} \tag{6.37}$$

is at most $\eta/2$. Due to (6.36) and the continuous mapping theorem, any continuous functional T of Π_n converges in distribution to $T(\Pi)$. We use this to compare the law of $\max_{i\leq\varepsilon n}(\mathcal{F}_i/u_n)(i/n)^{-1/\theta_m}$ and $\max_{(t,f)\in\Pi:t\leq\varepsilon} ft^{-1/\theta_m}$ by defining, for $\varepsilon\in(0,1]$, the functional T_ε, such that $T_\varepsilon(\Pi) := \max_{(t,f)\in\Pi:t\leq\varepsilon} ft^{-1/\theta_m}$. Let $M_k := \{\Pi \in M_p(E) \mid T_\varepsilon(\Pi) < k\}, k\in\mathbb{N}$. Then, on M_k, T_ε is continuous, and thus T_ε is continuous on $\cup_{k\in\mathbb{N}} M_k$. Since the point processes Π with intensity measure ν as described above are such that $T_\varepsilon(\Pi)$ is finite almost surely, as follows from (6.37), $\Pi\in M_k$ for some $k\in\mathbb{N}$ and thus T_ε is continuous with respect to Π almost surely for any $\varepsilon\in(0,1]$. It follows that, for δ,η fixed, ε chosen such that (6.37) holds and n sufficiently large,

$$\mathbb{P}\bigg(\max_{i\in[\varepsilon n]} \frac{\mathcal{F}_i}{u_n}(i/n)^{-1/\theta_m} > \delta\bigg) \leq \mathbb{P}\bigg(\max_{(t,f)\in\Pi:t\leq\varepsilon} ft^{-1/\theta_m} > \delta\bigg) + \eta/2 < \eta.$$

As ε_n decreases monotonically, $\varepsilon_n < \varepsilon$ for n sufficiently large. Hence, it follows that for n large,

$$\mathbb{P}\bigg(\max_{i\in[\varepsilon_n n]} \frac{\mathcal{F}_i}{u_n}(i/n)^{-1/\theta_m} > \delta\bigg) \leq \mathbb{P}\bigg(\max_{i\leq\varepsilon n} \frac{\mathcal{F}_i}{u_n}(i/n)^{-1/\theta_m} > \delta\bigg) < \eta. \tag{6.38}$$

We therefore can conclude that $\max_{i\in[\varepsilon_n n]}(\mathcal{F}_i/u_n)(i/n)^{1/\theta_m} \xrightarrow{\mathbb{P}} 0$ as $n\to\infty$, as η is arbitrary. We now show that $\max_{i\leq \varepsilon_n n}|(c_i^1/c_n^1)(i/n)^{1/\theta_m}-1|$ converges almost surely. Because of Lemma 6.4, $c_n^k n^{k/\theta_m} \xrightarrow{a.s.} c_k$, so that for each fixed $i\in\mathbb{N}$, $|(c_i^1/c_n^1)(i/n)^{1/\theta_m} - 1| \xrightarrow{a.s.} |c_i^1 i^{1/\theta_m}/c_1 - 1| =: A_i$. Note that it follows from the proof of Lemma 6.4 that $c_k>0$ almost surely (and thus for c_1 in particular), so that $A_i<\infty$ almost surely for all $i\in\mathbb{N}$. Also, $A_i\neq A_j$ almost surely for all $i\neq j$. Using the lower and upper bound in (6.10), we obtain for every $i\geq n_0+1$ fixed and $n\geq i$,

$$
\begin{aligned}
&\sup_{n\geq i}\left|\frac{c_i^1}{c_n^1}\Big(\frac{i}{n}\Big)^{1/\theta_m}-1\right|\\
&\quad\leq \max\Bigg\{\frac{mk}{\mathbb{E}[\mathcal{F}]}\sum_{j=i}^\infty \frac{|S_j/j-\mathbb{E}[\mathcal{F}]\,|}{m_0+m(j-n_0)+S_j}+\frac{m}{2}\sum_{j=i}^\infty\Big(\frac{k}{S_j}\Big)^2+\frac{m_0+\mathbb{E}[\mathcal{F}]\,n_0+(m-1)}{\theta_m^2}\frac{\pi^2}{6i}\\
&\qquad\qquad +\frac{1}{\theta_m((i-(n_0+1))\vee 1)},\\
&\qquad\qquad \exp\Bigg\{\frac{k}{\theta_m}\log\Big(\frac{i}{(i-(n_0+1))\vee 1}\Big)+\frac{mk}{\mathbb{E}[\mathcal{F}]}\sum_{j=i}^\infty\frac{|S_j/j-\mathbb{E}[\mathcal{F}]\,|}{m_0+j-n_0+S_j}\Bigg\}-1\Bigg\}\\
&\quad=: B_i.
\end{aligned}
$$

As the sums in the maximum are almost surely finite for all $i\in\mathbb{N}$, as follows from the proof of Lemma 6.4 and the strong law of large numbers, $\lim_{i\to\infty}B_i=0$ almost surely. Thus, combining the above steps with Lemma 6.7, we conclude that as $n\to\infty$,

$$
\max_{i\in[n]}\left|\frac{c_i^1}{c_n^1}\Big(\frac{i}{n}\Big)^{1/\theta_m}-1\right| \xrightarrow{a.s.} \sup_{i\geq 1}A_i,
$$

and there exist almost surely finite random variables I,N, such that the maximum is almost surely attained at index $i=I$ for all $n\geq N$. It thus follows that the maximum converges almost surely to an almost surely finite limit A_I. We can now conclude that, as $\varepsilon_n n\to\infty$,

$$
\max_{i\leq \varepsilon_n n}\left|\frac{c_i^1}{c_n^1}\Big(\frac{i}{n}\Big)^{1/\theta_m}-1\right| \xrightarrow{a.s.} \sup_{i\geq 1}A_i = A_I,
$$

which, together with (6.38), yields

$$
\max_{i\leq \varepsilon_n n}\frac{\mathcal{F}_i}{u_n}\Big(\frac{i}{n}\Big)^{-1/\theta_m}\max_{i\leq \varepsilon_n n}\left|\frac{c_i^1}{c_n^1}\Big(\frac{i}{n}\Big)^{1/\theta_m}-1\right| \xrightarrow{\mathbb{P}} 0.
$$

Combining this with (6.32) and (6.35), we obtain (6.30). By a similar argument as before, we find,

$$
\max_{i\in[n]}\frac{\mathcal{F}_i}{u_n}\Big(\Big(\frac{i}{n}\Big)^{-1/\theta_m}-1\Big) \xrightarrow{d} \max_{(t,f)\in\Pi} f(t^{-1/\theta_m}-1). \tag{6.39}
$$

Thus, combining (6.29), (6.30) and (6.39) and applying Slutsky's theorem [24], we arrive at the desired result.

We now prove (6.2) and so we let $\alpha\in(1,2)$. An important result is stated in Proposition 5.1. By the construction of Π_n in (5.1) and the definition of T^ε in (5.2), it follows that

$$
\frac{\mathcal{F}_i}{u_n}T^{i/n}(\Pi_n)=\frac{\mathcal{F}_i}{u_n}\int_{i/n}^1\Big(\int_E f\mathbb{1}_{\{t\leq s\}}\mathrm{d}\Pi_n(t,f)\Big)^{-1}\mathrm{d}s=\frac{\mathcal{F}_i}{u_n}\frac{1}{n}\sum_{j=i}^n\frac{u_n}{S_j}=\frac{\mathcal{F}_i}{n}\sum_{j=i}^n\frac{1}{S_j},
$$

as for $s\in[j/n,(j+1)/n)$ the integrand is constant. Hence, by Proposition 5.1, what remains is to prove that

$$
\left|\max_{i\in[n]}\mathbb{E}_{\mathcal{F}}[\mathcal{Z}_n(i)/n]-\max_{i\in[n]}\frac{\mathcal{F}_i}{n}\sum_{j=i}^n m/S_j\right| \xrightarrow{\mathbb{P}} 0. \tag{6.40}
$$

Recall the result in (6.29) regarding the limit of the maximum conditional mean. The above is therefore implied by the following two statements:

$$\begin{aligned} &\Big|\max_{i\in[n]}\frac{\mathcal{F}_i}{n}\Big(\frac{c_i^1}{c_n^1}-1\Big)-\max_{i\in[n]}\frac{\mathcal{F}_i}{n}\sum_{j=i}^n m/(m_0+m(j-n_0)+S_j)\Big| \overset{\mathbb{P}}{\longrightarrow} 0,\\ &\Big|\max_{i\in[n]}\frac{\mathcal{F}_i}{n}\sum_{j=i}^n m/(m_0+m(j-n_0)+S_j)-\max_{i\in[n]}\frac{\mathcal{F}_i}{n}\sum_{j=i}^n m/S_j\Big| \overset{\mathbb{P}}{\longrightarrow} 0. \end{aligned} \tag{6.41}$$

We start by proving the first line of (6.41). Let us write $Z_j := m_0 + m(j-n_0)+S_j$. By (6.28), it follows that

$$\Big|\max_{i\in[n]}\frac{\mathcal{F}_i}{n}(c_i^1/c_n^1-1)-\max_{i\in[n]}\frac{\mathcal{F}_i}{n}\sum_{j=i}^n m/Z_j\Big| \leq \max_{i\in[n]}\frac{\mathcal{F}_i}{n}\bigg((c_i^1/c_n^1-1)-\sum_{j=i}^n m/Z_j\bigg),$$

as the terms within the brackets on the right-hand side are a.s. positive. Then, we further bound the expression on the right-hand side from above by splitting the maximum into two parts, as

$$\begin{aligned} \max_{i\in[n]}\frac{\mathcal{F}_i}{n}\bigg((c_i^1/c_n^1-1)-\sum_{j=i}^n m/Z_j\bigg) \leq{}& \max_{i\in[i_n]}\frac{\mathcal{F}_i}{n}\bigg((c_i^1/c_n^1-1)-\sum_{j=i}^n m/Z_j\bigg)\\ &+\max_{i_n\leq i\leq n}\frac{\mathcal{F}_i}{n}\bigg((c_i^1/c_n^1-1)-\sum_{j=i}^n m/Z_j\bigg), \end{aligned} \tag{6.42}$$

where i_n is strictly increasing and tends to infinity with n. We first investigate the second maximum, by bounding the terms within the brackets. Namely, recalling the definition of c_n^1 in (6.5) and applying the inequality $1-x\leq \mathrm{e}^{-x}$ for all $x\in\mathbb{R}$ to c_i^1/c_n^1 yields

$$(c_i^1/c_n^1-1)-\sum_{j=i}^n m/Z_j \leq \exp\bigg\{\sum_{j=i}^n m/Z_j\bigg\}-1-\sum_{j=i}^n m/Z_j=\sum_{k=2}^\infty\bigg(\sum_{j=i}^n m/Z_j\bigg)^k.$$

Now, fix $\varepsilon>0$. By (4.33) there exists an almost surely finite random variable J such that for all $j\geq J$, $M_j\geq j^{1/(\alpha-1)-\varepsilon}$, with $M_j=\max_{k\leq j}\mathcal{F}_k$. So, on $\{i\geq J\}$, $Z_j\geq j^{1/(\alpha-1)-\varepsilon}$ for all $j\geq i$. This yields the upper bound

$$\sum_{k=2}^\infty m i^{-k((2-\alpha)/(\alpha-1)-\varepsilon)}\leq C i^{-2((2-\alpha)/(\alpha-1)-\varepsilon)}, \tag{6.43}$$

for some constant $C>0$, as we can bound an exponentially decaying sum by a constant times its first term. It follows, on $i_n\geq J$, which holds with high probability, and by (6.43), that

$$\max_{i_n\leq i\leq n}\frac{\mathcal{F}_i}{n}\bigg((c_i^1/c_n^1-1)-\sum_{j=i}^n m/Z_j\bigg)\leq C i_n^{-2((2-\alpha)/(\alpha-1)-\varepsilon)}\frac{u_n}{n}\max_{i_n\leq i\leq n}\frac{\mathcal{F}_i}{u_n}, \tag{6.44}$$

which tends to zero in probability when $i_n^{-2((2-\alpha)/(\alpha-1)-\varepsilon)}u_n/n=o(1)$, that is, when $i_n=n^\rho$, with $\rho\in(1/2,1)$. On the other hand, when considering the first maximum in (6.42), we find

$$\max_{i\in[i_n]}\frac{\mathcal{F}_i}{n}\bigg((c_i^1/c_n^1-1)-\sum_{j=i}^n m/Z_j\bigg)\leq (1/c_n^1)\frac{u_{i_n}}{n}\max_{i\leq i_n}\frac{\mathcal{F}_i}{u_{i_n}}, \tag{6.45}$$

where we bound the terms inside the brackets on the left-hand side by omitting all negative terms and by noting that $c_i^1 \leq 1$ for all i. The right-hand side of (6.45) converges to zero in probability when $u_{i_n}/n = o(1)$, that is, when $i_n = n^\rho$ with $\rho < \alpha - 1$, since c_n^1 converges almost surely for $\alpha \in (1,2)$ by Lemma 6.4. We conclude that for $\alpha \in (3/2,2)$ we can find a $\rho \in (1/2, \alpha-1)$ such that both maxima tend to zero in probability. When $\alpha \in (1,3/2]$, however, such a ρ cannot be found and more work is required to prove the desired result. In this case, we split the maximum into $K = K(\alpha) < \infty$ maxima, as follows: Let $A_{i,n} := \mathcal{F}_i/n, B_{i,n} := (c_i^1/c_n^1 - 1) - \sum_{j=i}^n m/Z_j$. Then, we define $i_n^k := n^{\rho_k}$, $k = 0,1,\ldots,K$, with $\rho_0 = 0, \rho_K = 1$, and

$$\rho_k := \frac{\alpha-1}{2}\frac{c^k-1}{c-1}, \qquad k \in \{1,2,\ldots,K-1\},$$

where $c := 2(2-\alpha) - 2\varepsilon(\alpha-1) \neq 1$. Note that ρ_k is strictly increasing in k, independent of $c<1$ or $c>1$. We now write

$$\max_{i\in[n]} A_{i,n}B_{i,n} \leq \sum_{k=0}^{K-1} \max_{i_n^k \leq i \leq i_n^{k+1}} A_{i,n} \max_{i_n^k \leq i \leq i_n^{k+1}} B_{i,n}. \tag{6.46}$$

We first deal with the $k=0$ term. As in (6.45), since $\rho_1 < \alpha - 1$, $\max_{i_n^0 \leq i \leq i_n^1} A_{i,n} \max_{i_n^0 \leq i \leq i_n^1} B_{i,n}$ tends to zero in probability. For $k = 1,\ldots,K-2$, following the same steps that lead to the bound in (6.44), we obtain

$$\max_{i_n^k \leq i \leq i_n^{k+1}} A_{i,n} \max_{i_n^k \leq i \leq i_n^{k+1}} B_{i,n} \leq C_k (i_n^k)^{-2((2-\alpha)/(\alpha-1)-\varepsilon)} \frac{u_{i_n^{k+1}}}{n} \max_{i_n^k \leq i \leq i_n^{k+1}} \frac{\mathcal{F}_i}{u_{i_n^{k+1}}},$$

for some constant $C_k > 0$. This upper bound tends to zero in probability when

$$\rho_{k+1} < \alpha - 1 + (2(2-\alpha) - 2\varepsilon(\alpha-1))\rho_k = (\alpha-1) + c\rho_k \tag{6.47}$$

is satisfied. By the definition of ρ_k, this holds when

$$\frac{c^{k+1}-1}{c-1} - 2 < c\frac{c^k-1}{c-1} \ \Leftrightarrow -1 + \sum_{j=1}^k c^j < \sum_{j=1}^k c^j,$$

which is indeed the case. Finally, for $k = K-1$, again using the similar bound as in (6.44), we find that the final term of the sum in (6.46) converges to zero in probability when $\rho_{K-1} \in (1/2,1)$. What remains to show, is that for all $\alpha \in (1,3/2]$ there does exist a finite K such that $\rho_{K-1} \in (1/2,1)$. We distinguish two cases: $\alpha = 3/2$ and $\alpha \in (1,3/2)$. For the first case, $c<1$ for any choice of ε. This implies that $\rho_k \to 1/(4\varepsilon)$ as k tends to infinity, so taking $\varepsilon < 1/2$ suffices. For $\alpha \in (1,3/2)$, we can choose ε sufficiently small, such that $c>1$, so that ρ_k diverges. In both cases there therefore exists a K such that $\rho_k > 1/2$ for all $k \geq K-1$. Thus, in both cases, we can define $K := \inf\{k \in \mathbb{N} \,|\, \rho_k > 1/2\} + 1$. The only issue left to address regarding K, is that it is possible that $\rho_{K-1} > 1$. However, in that case we can simply choose $\rho_{K-1} = a$, for any $a \in (1/2,1)$, since $\rho_{K-2} \leq 1/2 < a$ by the definition of K, and decreasing ρ_{K-1} does not violate the constraint in (6.47) for $k = K-2$. We hence obtain the first line in (6.41).

The proof for the second line in (6.41) follows similarly. First, by letting $i = i(n)$ tend to infinity with n, we bound, conditionally on $\{i \geq J\}$,

$$\bigg|\sum_{j=i}^n \frac{m}{S_j} - \sum_{j=i}^n \frac{m}{Z_j}\bigg| \leq C\sum_{j=i}^n j/M_j^2 \leq C\sum_{j=i}^n j^{1-2/(\alpha-1)+\varepsilon} \leq \widetilde{C} i^{-2((2-\alpha)/(\alpha-1)-\varepsilon/2)}, \tag{6.48}$$

for some constant $C \geq m + m_0$. We note that this bound is similar to the upper bound for $(c_i^1/c_n^1 - 1) - \sum_{j=i}^n 1/(j + S_j/m)$ in (6.43). Also, both sums on the left-hand side of (6.48) converge almost surely, as $\alpha \in (1,2)$. Thus, a similar approach, with the same indices $i_n^0, \dots, i_n^K$ can be used to obtain the desired result. Combining both statements in (6.41) and using the triangle inequality and the continuous mapping theorem proves (6.40), which together with Proposition 5.1 finishes the proof. □

We now prove Proposition 6.2:

Proof of Proposition 6.2. The focus of the proof is on the PAFUD model, for which we use the martingales $M_n^k(i)$. The proof for the PAFRO model follows by setting $m = 1$, and for the PAFFD model it follows in a similar fashion, where all upper bounds still hold when the supermartingale $\widetilde{M}_n^k(i)$ is to be used. We prove (6.3) first. Applying (6.28), a p^{th} moment bound for some $p > 1$ to be determined later, using Markov's inequality and Hölder's inequality yields

$$\begin{aligned}
&\mathbb{P}_{\mathcal{F}}(|\max_{i\in[n]} \mathcal{Z}_n(i) - \max_{i\in[n]} \mathbb{E}_{\mathcal{F}}[\mathcal{Z}_n(i)]| > \eta u_n) \\
&\quad\le \mathbb{P}_{\mathcal{F}}\Big(\max_{i\in[n]} |\mathcal{Z}_n(i) - \mathbb{E}_{\mathcal{F}}[\mathcal{Z}_n(i)]| > \eta u_n\Big) \\
&\quad\le \frac{1}{(\eta u_n)^p}\sum_{i=1}^n \mathbb{E}_{\mathcal{F}}\big[|\mathcal{Z}_n(i) - \mathbb{E}_{\mathcal{F}}[\mathcal{Z}_n(i)]|^p\big] \\
&\quad\le \frac{1}{(\eta u_n)^p}\sum_{i=1}^n \mathbb{E}_{\mathcal{F}}\big[|\mathcal{Z}_n(i) - \mathbb{E}_{\mathcal{F}}[\mathcal{Z}_n(i)]|^{2k}\big]^{p/(2k)},
\end{aligned} \tag{6.49}$$

where $k > p/2$ is an integer. As $\mathcal{Z}_n(i) - \mathbb{E}_{\mathcal{F}}[\mathcal{Z}_n(i)] = (\mathcal{Z}_n(i) + \mathcal{F}_i) - \mathbb{E}_{\mathcal{F}}[\mathcal{Z}_n(i) + \mathcal{F}_i]$ and $2k$ is even, we find, using Hölder's and Jensen's inequality and setting $X_n(i) := \mathcal{Z}_n(i) + \mathcal{F}_i$,

$$\begin{aligned}
\mathbb{E}_{\mathcal{F}}\big[|\mathcal{Z}_n(i) - \mathbb{E}_{\mathcal{F}}[\mathcal{Z}_n(i)]|^{2k}\big] &= \sum_{j=0}^{2k}\binom{2k}{j}\mathbb{E}_{\mathcal{F}}[X_n(i)^j](-1)^j\mathbb{E}_{\mathcal{F}}[X_n(i)]^{2k-j} \\
&= \sum_{j=0}^{k}\binom{2k}{2j}\mathbb{E}_{\mathcal{F}}[X_n(i)^{2j}]\mathbb{E}_{\mathcal{F}}[X_n(i)]^{2k-2j} \\
&\quad - \sum_{j=1}^{k}\binom{2k}{2j-1}\mathbb{E}_{\mathcal{F}}[X_n(i)^{2j-1}]\mathbb{E}_{\mathcal{F}}[X_n(i)]^{2k-(2j-1)} \\
&\le \sum_{j=0}^{k}\binom{2k}{2j}\mathbb{E}_{\mathcal{F}}[X_n(i)^{2k}] - \sum_{j=1}^{k}\binom{2k}{2j-1}\mathbb{E}_{\mathcal{F}}[X_n(i)]^{2k}.
\end{aligned}$$

Using that

$$\sum_{j=0}^{2k}\binom{2k}{j} = 2^{2k}, \qquad \sum_{j=0}^{2k}\binom{2k}{j}(-1)^j = 0,$$

it follows that both sums in the last line of (6.49) equal 2^{2k-1}. We can thus bound (6.49) from above by

$$\frac{2^{2k-1}}{(\eta u_n)^p}\sum_{i=1}^n(\mathbb{E}_{\mathcal{F}}\big[(\mathcal{Z}_n(i) + \mathcal{F}_i)^{2k}\big] - \mathbb{E}_{\mathcal{F}}[\mathcal{Z}_n(i) + \mathcal{F}_i]^{2k})^{p/(2k)}. \tag{6.50}$$

We now aim to bound the $2k^{\text{th}}$ moment of $\mathcal{Z}_n(i) + \mathcal{F}_i$. Since, for $x \ge 0, k \in \mathbb{N}$, $x^{2k} \le \prod_{j=1}^{2k}(x + (j-1)) = \binom{x+(2k-1)}{2k}(2k)!$, it follows from Lemma 6.3 that

$$\mathbb{E}_{\mathcal{F}}\big[(\mathcal{Z}_n(i) + \mathcal{F}_i)^{2k}\big] \le (c_n^{2k})^{-1}(2k)!\mathbb{E}_{\mathcal{F}}[M_n^{2k}(i)] = \frac{c_{i\vee n_0}^{2k}}{c_n^{2k}}(2k)!\binom{\mathcal{Z}_{i\vee n_0}(i) + \mathcal{F}_i + 2k - 1}{2k}.$$

We note that this inequality would still hold for the PAFFD model, when using the supermartingales $\widetilde{M}_n^k(i)$ and the sequences $\widetilde{c}_n^k(i)$. We thus obtain the upper bound

$$\mathbb{E}_{\mathcal{F}}\big[(\mathcal{Z}_n(i)+\mathcal{F}_i)^{2k}\big] \leq \frac{c_{i\vee n_0}^{2k}}{c_n^{2k}}(\mathcal{Z}_{i\vee n_0}+\mathcal{F}_i)^{2k}+\frac{c_{i\vee n_0}^{2k}}{c_n^{2k}}P_{2k-1}(\mathcal{Z}_{i\vee n_0}(i)+\mathcal{F}_i),$$

where $P_{2k-1}(x)=(2k)!\binom{x+2k-1}{2k}-x^{2k}$ is a polynomial of degree $2k-1$. Using (6.27), we find

$$\begin{aligned}\mathbb{E}_{\mathcal{F}}\big[(\mathcal{Z}_n(i)+\mathcal{F}_i)^{2k}\big]-\mathbb{E}_{\mathcal{F}}[\mathcal{Z}_n(i)+\mathcal{F}_i]^{2k} &\leq \Big(\frac{c_{i\vee n_0}^{2k}}{c_n^{2k}}-\Big(\frac{c_{i\vee n_0}^{1}}{c_n^{1}}\Big)^{2k}\Big)(\mathcal{Z}_{i\vee n_0}(i)+\mathcal{F}_i)^{2k}\\ &\quad+\frac{c_{i\vee n_0}^{2k}}{c_n^{2k}}P_{2k-1}(\mathcal{Z}_{i\vee n_0}(i)+\mathcal{F}_i).\end{aligned} \tag{6.51}$$

Using the definition of c_n^k in (6.5) yields, for all $1\leq r\leq n$,

$$\begin{aligned}\frac{c_r^{2k}}{c_n^{2k}}&=\prod_{j=r\vee n_0}^{n-1}\prod_{\ell=1}^{m}\Big(1+\frac{2k}{m_0+m(j-n_0)+(\ell-1)+S_j}\Big)\\ &\leq \prod_{j=r\vee n_0}^{n-1}\prod_{\ell=1}^{m}\Big(1+\frac{1}{m_0+m(j-n_0)+(\ell-1)+S_j}\Big)^{2k}=\Big(\frac{c_r^1}{c_n^1}\Big)^{2k}.\end{aligned} \tag{6.52}$$

Therefore, using this in (6.51) we obtain an upper bound that contains powers of $\mathcal{F}_i$ of order at most $2k-1$. This is the essential step to proving concentration holds. Namely, in (6.50), this upper bound yields an expression with powers of $\mathcal{F}_i$ of order at most $p(1-1/2k)$, which is just slightly less than p. The aim is, for every value of $\alpha>2$, to find values p,k such that the $p(1-1/2k)^{\text{th}}$ moment of $\mathcal{F}$ exists and such that the entire expression in (6.50) still tends to zero.

Let us write

$$P_{2k-1}(x)=\sum_{\ell=0}^{2k-1}C_\ell x^\ell,$$

for non-negative constants C_ℓ. Combining (6.51) and (6.52) in (6.50), bounding $\mathcal{Z}_{i\vee n_0}(i)$ from above by m_0 and recalling that $p/(2k)<1$, results in the upper bound

$$\frac{2^{2k}}{(\eta u_n)^p}\sum_{i=1}^{n}\Big(\frac{c_i^{2k}}{c_n^{2k}}\Big)^{p/(2k)}\sum_{\ell=0}^{2k-1}\widetilde{C}_\ell^{p/(2k)}\mathcal{F}_i^{\ell p/2k}, \tag{6.53}$$

where the $\widetilde{C}_\ell>0$ are constants. We focus on the term where $\ell=2k-1$, as this is the boundary case. All other cases follow analogously. For the first n_0 terms, we can bound c_i^{2k} from above by $(i/n_0)^{-p/\theta_m}$. For $n_0+1\leq i\leq n$, we use (6.10) to bound c_i^{2k}/c_n^{2k} from above. This yields for all terms, for some constant $C>0$,

$$\begin{aligned}&\frac{\widetilde{C}_{2k-1}^{p/(2k)}2^{2k}}{(\eta u_n)^p}\Bigg(\exp\Big\{\frac{mp}{\mathbb{E}[\mathcal{F}]}\sum_{j=n_0}^{\infty}\frac{|S_j/j-\mathbb{E}[\mathcal{F}]\,|}{j-n_0+S_j}-C\Big\}\vee 1\Bigg)\sum_{i=1}^{n}\Big(\frac{nn_0}{i}\Big)^{p/\theta_m}\mathcal{F}_i^{p(1-1/2k)}\\ &\leq C_{k,p,\theta_m}\exp\Big\{\frac{mp}{\mathbb{E}[\mathcal{F}]}\sum_{j=n_0}^{\infty}\frac{|S_j/j-\mathbb{E}[\mathcal{F}]\,|}{j-n_0+S_j}\Big\}\frac{n^{p/\theta_m}}{u_n^p}\sum_{i=1}^{n}i^{-p/\theta_m}\mathcal{F}_i^{p(1-1/2k)},\end{aligned} \tag{6.54}$$

for some constant C_{k,p,θ_m}. In the last line, the exponential term is almost surely finite, as follows from the proof of Lemma 6.4. We now show that the fraction multiplied by the sum converges to zero in mean when p and k are chosen in a specific way. That is, for $\alpha>2$, set $p:=(1+\varepsilon)(\alpha-1)$, where $\varepsilon\in(0,1/(\alpha+1))$ and set $k:=\lceil p/2\rceil$. First note that $2k > p$, which was required for the Hölder inequality used in (6.49). We now show that the $p(1-1/(2k))^{\text{th}}$ moment of the fitness distribution exists. For this to hold, $\alpha - 1 > p(1-1/(2k))$ needs to be satisfied, or, equivalently,

$$k < \frac{p}{2(p-(\alpha-1))} = \frac{1+\varepsilon}{2\varepsilon},$$

and, as $\varepsilon \in (0, 1/(\alpha+1))$,

$$\frac{1+\varepsilon}{2\varepsilon} - \frac{p}{2} = \frac{1+\varepsilon}{2}(1/\varepsilon - (\alpha-1)) > 1+\varepsilon.$$

It follows that, indeed,

$$(1+\varepsilon)/(2\varepsilon) > p/2 + 1 + \varepsilon > \lceil p/2 \rceil = k.$$

Hence, taking the mean, we obtain

$$\frac{n^{p/\theta_m}}{u_n^p} \sum_{i=1}^n i^{-p/\theta_m} \mathbb{E}\Big[\mathcal{F}_i^{p(1-1/(2k))}\Big] \le C \frac{n^{p/\theta_m}}{u_n^p} n^{(1-p/\theta_m)\vee 0},$$

with $C > 0$ a constant. This tends to zero with n, as $u_n = n^{1/(\alpha-1)}\widetilde{\ell}(n)$ for some slowly-varying function $\widetilde{\ell}(n)$, and both $p > \alpha - 1$ and $\theta_m > \alpha - 1$ hold. So, the last expression in (6.54) consists of an almost surely finite random variable (the exponential term) and a term that converges to zero mean, which implies that the entire expression converges to zero in probability. The same argument holds also for all other values of ℓ in (6.53). Thus, as n tends to infinity,

$$\mathbb{P}_{\mathcal{F}}\Big(|\max_{i\in[n]} \mathcal{Z}_n(i) - \max_{i\in[n]} \mathbb{E}_{\mathcal{F}}[\mathcal{Z}_n(i)]| > \eta u_n\Big) \xrightarrow{\mathbb{P}} 0. \tag{6.55}$$

As this conditional probability measure is bounded from above by one, it follows from the dominated convergence theorem and (6.55) that (6.3) holds.

We now prove (6.4), so let $\alpha \in (1,2)$. A different approach is required, so we write, using (6.28), a union bound and Chebyshev's inequality,

$$\begin{aligned}
\mathbb{P}_{\mathcal{F}}\Big(|\max_{i\in[n]} \mathcal{Z}_n(i) - \max_{i\in[n]} \mathbb{E}_{\mathcal{F}}[\mathcal{Z}_n(i)]| > \eta n\Big) &\le \mathbb{P}_{\mathcal{F}}\Big(\max_{i\in[n]} |\mathcal{Z}_n(i) - \mathbb{E}_{\mathcal{F}}[\mathcal{Z}_n(i)]| > \eta n\Big) \\
&\le \sum_{i=1}^n \mathbb{P}_{\mathcal{F}}\big(|M_n^1(i) - \mathbb{E}_{\mathcal{F}}[M_n^1(i)]| \ge \eta n c_n^1\big) \\
&\le (\eta n c_n^1)^{-2} \sum_{i=1}^n \mathrm{Var}_{\mathcal{F}}(M_n^1(i)).
\end{aligned} \tag{6.56}$$

We now use the martingale property to split the variance in the variance of martingale increments. To this end, we need to introduce some notation. Recall that $\mathcal{Z}_{n,j}(i)$ is the degree of i in $\mathcal{G}_{n,j}$, the graph with n vertices where the $n+1^{\text{st}}$ vertex has connected j half-edges with the first n vertices. Now, let us write

$$\begin{aligned}
c_{n,j}^1(m) &:= \prod_{r=n_0}^{n-1} \prod_{\ell=1}^{j} \Big(1 - \frac{1}{m_0 + m(r-n_0) + (\ell-1) + 1 + S_r}\Big), \\
M_{n,j}^1(i) &:= c_{n-1,j}^1(m)(\mathcal{Z}_{n-1,j}(i) + \mathcal{F}_i).
\end{aligned}$$

If we let $M_\ell := M_{n,j}^1(i)$, where $n \ge n_0, j \in [m]$ are such that $mn + (j-1) = \ell$, it follows from the proof of Lemma 6.3 that M_ℓ is a martingale for the PAFRO and PAFUD model. Hence, we can then write the conditional variance of $M_n^1(i)$ as in (6.56) as

$$\mathrm{Var}_{\mathcal{F}}(M_n^1(i)) = \sum_{k=i+1\vee n_0+1}^{n} \sum_{j=1}^{m} \mathrm{Var}_{\mathcal{F}}(\Delta M_{k,j}^1(i)), \tag{6.57}$$

where $\Delta M^1_{k,j}(i) := M^1_{k,j}(i) - M^1_{k,j-1}(i)$, and where we note that $M^1_{k,0}(i) = M^1_{k-1,m}(i) = M^1_k(i)$ for all $k = i \vee n_0, \ldots, n$. We then obtain

$$\begin{aligned}&\mathrm{Var}_{\mathcal{F}}(\Delta M^1_{k,j}(i))\\ &\quad= (c^1_{k,j-1})^2\mathbb{E}_{\mathcal{F}}\Big[\Big(\mathbb{1}_{k,j,i} - \frac{\mathcal{Z}_{k-1,j-1}(i)+\mathcal{F}_i+\mathbb{1}_{k,j,i}}{m_0+m((k-1)-n_0)+(j-1)+1+S_{k-1}}\Big)^2\Big],\end{aligned} \tag{6.58}$$

where $\mathbb{1}_{k,j,i}$ is the indicator of the event that vertex k connects its j^{th} half-edge to vertex i. We rewrite this to find the upper bound

$$\begin{aligned}\mathrm{Var}_{\mathcal{F}}(\Delta M^1_{k,j}(i)) &\leq \mathbb{E}_{\mathcal{F}}\Big[\Big(\mathbb{1}_{k,j,i} - \frac{\mathcal{Z}_{k-1,j-1}(i)+\mathcal{F}_i}{m_0+m((k-1)-n_0)+(j-1)+S_{k-1}}\Big)^2\Big]\\ &= \mathbb{E}_{\mathcal{F}}\big[\mathrm{Var}(\mathbb{1}_{k,j,i}\,|\,\mathcal{G}_{k-1,j-1})\big]\\ &\leq \mathbb{E}_{\mathcal{F}}\Big[\frac{\mathcal{Z}_{k-1,j-1}(i)+\mathcal{F}_i}{m_0+m((k-1)-n_0)+(j-1)+S_{k-1}}\Big].\end{aligned} \tag{6.59}$$

Combining this with (6.56) and (6.57) and switching summations yields

$$\mathbb{P}_{\mathcal{F}}\Big(|\max_{i\in[n]}\mathcal{Z}_n(i) - \max_{i\in[n]}\mathbb{E}_{\mathcal{F}}[\mathcal{Z}_n(i)]| > \eta n\Big) \leq (\eta n c^1_n)^{-2}mn,$$

This final expression tends to zero almost surely, as c^1_n converges almost surely when $\alpha \in (1,2)$, as follows from Lemma 6.4. For the PAFFD model, we can use similar steps. We construct $\widetilde{M}_\ell := \widetilde{M}^1_{n,j}(i)$ as above, with $\widetilde{M}^1_{n,j} := \widetilde{c}^1_{n-1,j}(m)(\mathcal{Z}_{n-1,j}(i)+\mathcal{F}_i)$, and

$$\widetilde{c}_{n,j}(m) := \prod_{r=n_0}^{n-1}\Big(1 - \frac{1}{m_0+m(r-n_0)+S_r}\Big)^j.$$

It again follows from the proof of Lemma 6.3 that $\widetilde{M}_\ell$ is a supermartingale, thus yielding (6.57) for $\widetilde{M}_n(i)$. Then, all further steps can be applied for the PAFFD model as well, where the equality in (6.57) becomes an upper bound and the denominator of the fractions in (6.58) and (6.59) changes to $m_0 + m((k-1)-n_0) + S_{k-1}$.

For the PAFRO model, an adapted final step is required, as the conditional moments in (6.59) do not sum to one (when summing over i from 1 to $k-1$). Rather, we set m to 1 and follow the same steps up to (6.59). Then, we obtain by switching the summations,

$$\mathbb{P}_{\mathcal{F}}\Big(|\max_{i\in[n]}\mathcal{Z}_n(i) - \max_{i\in[n]}\mathbb{E}_{\mathcal{F}}[\mathcal{Z}_n(i)]| > \eta n\Big) \leq (\eta n c^1_n)^{-2}\sum_{k=n_0+1}^{n}\sum_{i=1}^{k-1}\frac{\mathbb{E}_{\mathcal{F}}[\mathcal{Z}_{k-1}(i)+\mathcal{F}_i]}{m_0+((k-1)-n_0)}.$$

Now, in the same spirit as the steps from (4.5) through (4.6), we obtain the upper bound

$$(\eta n c^1_n)^{-2}\sum_{k=n_0+1}^{n}\sum_{i=1}^{k-1}\frac{(m_0+\mathcal{F}_i)(m_0+k-n_0)}{(m_0+i\vee n_0-n_0)(m_0+(k-1)-n_0)} =: (c^1_n)^{-2}Q_n,$$

where, in n the last step, we separate this upper bound into a product of two quantities. That is, we consider $(c^1_n)^{-2}$ and the rest of the terms, Q_n. Since c^1_n converges almost surely when $\alpha \in (1,2)$, it follows that $(c^1_n)^{-2}$ does too. Then, it remains to show that Q_n converges to zero in mean. Hence, taking the mean with respect to the fitness random variables yields

$$\mathbb{E}[Q_n] \leq \frac{2}{(\eta n)^2}\sum_{k=n_0+1}^{n}\sum_{i=1}^{k-1}\frac{m_0+\mathbb{E}[\mathcal{F}]}{m_0+i\vee n_0-n_0} \leq \frac{1}{(\eta n)^2}\sum_{k=n_0+1}^{n}(C_1+C_2\log k) \leq \frac{\widetilde{C}_1+\widetilde{C}_2\log n}{\eta^2 n},$$

which proves that Q_n does indeed converge to zero in mean. We thus also obtain for the PAFRO model that

$$\mathbb{P}_{\mathcal{F}}\Big(|\max_{i\in[n]}\mathcal{Z}_n(i)-\max_{i\in[n]}\mathbb{E}_{\mathcal{F}}[\mathcal{Z}_n(i)]|>\eta n\Big)\overset{\mathbb{P}}{\longrightarrow}0.$$

Finally, like the argument made above (6.56), applying the dominated convergence theorem proves (6.4) for all three models, which concludes the proof. □

7 Proof of the maximum degree growth theorem

In this section, we use the results from Section 6 to prove Theorem 2.7.

Proof of Theorem 2.7. We start by proving (i) and (ii). This directly follows from Lemmas 6.3 and 6.4. As discussed after Lemma 6.3, the martingales (resp. supermartingales) $M_n^k(i)$ (resp. $\widetilde{M}_n^k(i)$) converge almost surely to ξ_i^k (resp. $\widetilde{\xi}_i^k$). Also, for the PAFFD model, $M_n^1(i)$ converges almost surely to ξ_i^1 as well. By these two lemmas, $c_n^1\mathcal{Z}_n(i)=M_n^1(i)-c_n^1\mathcal{F}_i$ converges almost surely to ξ_i^1 for the PAFRO and PAFUD models, $\widetilde{c}_n^1\mathcal{Z}_n(i)=\widetilde{M}_n^1(i)-\widetilde{c}_n^1\mathcal{F}_i$ converges almost surely to $\widetilde{\xi}_i^1$ for the PAFFD model and $c_n^1(m)n^{1/\theta_m}$ and $\widetilde{c}_n^1(m)n^{1/\theta_m}$ converge almost surely to $c_1,\widetilde{c}_1$, respectively, when $\mathbb{E}\big[\mathcal{F}^{1+\varepsilon}\big]<\infty$ for some $\varepsilon>0$. Hence, we can set $\xi_i:=(c_1)^{-1}\xi_i^1$ for the PAFRO (note $m=1$) and the PAFUD model, and $\xi_i:=(\widetilde{c}_1)^{-1}\widetilde{\xi}_i^1$ for the PAFFD model. Since c_1 and $\widetilde{c}_1$ are finite almost surely, it follows directly from Lemma 6.5 that ξ_i has no atom at zero for all $i\in\mathbb{N}$ for any of the three models.

When $\alpha\in(1,2)$, we note that $c_n^1\overset{a.s.}{\longrightarrow}\underline{c}_1$ without the need of rescaling and thus (2.9) follows with $\mathcal{Z}_\infty(i):=\xi_i^1/\underline{c}_1-\mathcal{F}_i$, as $\mathcal{Z}_n(i)=M_n^1(i)/c_n^1-\mathcal{F}_i$, for the PAFRO and PAFUD models and $\mathcal{Z}_\infty(i):=\widetilde{\xi}_i^1/\widetilde{\underline{c}}_1-\mathcal{F}_i$ for the PAFFD model.

We now prove (iii). From the second inequality in (6.18) we obtain $(c_n^1)^k\le c_n^k$ when $k\ge 1$. Furthermore, from [16, Theorem 1] it follows that $x^k\le\Gamma(x+k)/\Gamma(x)$ for all $x>0,k\ge1$. Hence, $(c_n^1\mathcal{Z}_n(i))^k\le c_n^k(\mathcal{Z}_n(i)+\mathcal{F}_i)^k\le M_n^k(i)\Gamma(k+1)$ for $k\ge1$. Recall M from Lemma 6.6. Clearly, $M>\theta_m$ when $\mathbb{E}\big[\mathcal{F}^{\theta_m+\varepsilon}\big]<\infty$ for some $\varepsilon>0$. So, if we let $k\in(\theta_m,M)$, Lemma 6.6 yields

$$\lim_{i\to\infty}\sup_{n\ge i}c_n^1\mathcal{Z}_n(i)=0\text{ almost surely.}$$

It then follows from Lemma 6.7, as $c_n^1\mathcal{Z}_n(i)\overset{a.s.}{\longrightarrow}\xi_i^1$ and $\xi_i^1\ne\xi_j^1$ almost surely for $i\ne j$,

$$\max_{i\in[n]}n^{-1/\theta_m}\mathcal{Z}_n(i)=(n^{1/\theta_m}c_n^1)^{-1}\max_{i\in[n]}c_n^1\mathcal{Z}_n(i)\overset{a.s.}{\longrightarrow}(c_1)^{-1}\sup_{i\ge1}\xi_i^1=\sup_{i\ge1}\xi_i,\quad\text{and}\quad I_n\overset{a.s.}{\longrightarrow}I,$$

for some almost surely finite random variable I. The same approach with $\widetilde{M}_n^k(i)$ holds for the PAFFD model.

We now turn to the convergence of $\max_{i\in[n]}\mathcal{Z}_n(i)/u_n$ and $\max_{i\in[n]}\mathcal{Z}_n(i)/n$ as in (iv) and (v), respectively. This follows immediately by applying Slutsky's theorem to the results in Propositions 6.1 and 6.2. For the convergence of I_n/n as in (2.11) and (2.12), we let $0\le a<b\le1$ and define, using $z(t,f):=f(t^{-1/\theta_m}-1)$, the random variables

$$Q_\ell(a):=\max_{(t,f)\in\Pi:0<t<a}z(t,f),\quad Q(a,b):=\max_{(t,f)\in\Pi:a<t<b}z(t,f),\quad Q_r(b):=\max_{(t,f)\in\Pi:b<t<1}z(t,f),$$

and events

$$\begin{aligned}M_n(a,b)&:=\Big\{\max_{an<i<bn}\mathcal{Z}_n(i)/u_n>\big(\max_{1\le i\le an}\mathcal{Z}_n(i)/u_n\vee\max_{bn\le i\le n}\mathcal{Z}_n(i)/u_n\big)\Big\},\\ M(a,b)&:=\Big\{Q(a,b)>Q_\ell(a)\vee Q_r(b)\Big\}.\end{aligned}\tag{7.1}$$

We can then conclude, for $\alpha \in (2, 1+\theta_m)$,

$$\lim_{n\to\infty} \mathbb{P}(I_n/n \in (a,b)) = \lim_{n\to\infty} \mathbb{P}(M_n(a,b)) = \mathbb{P}(M(a,b)), \tag{7.2}$$

since it follows from the proof of Propositions 6.1 and 6.2 that the vector $(\mathcal{Z}_n(i)/u_n)_{i\in[n]}$ converges in distribution when $\alpha \in (2, 1+\theta_m)$. Now, by the fact that Π is a PPP with intensity measure $\nu(\mathrm{d}t \times \mathrm{d}x) = \mathrm{d}t \times (\alpha-1)x^{-\alpha}\mathrm{d}x$, we find

$$\mathbb{P}(Q(a,b) \le x) = \exp\left\{-\int_a^b \int_{x(t^{-1/\theta_m}-1)^{-1}}^\infty (\alpha-1)s^{-\alpha}\mathrm{d}s\mathrm{d}t\right\} = \exp\{-g(a,b)x^{-(\alpha-1)}\}, \tag{7.3}$$

where $g(a,b) := \int_a^b (t^{-1/\theta_m}-1)^{\alpha-1}\mathrm{d}t < \infty$ for all $0 \le a \le b \le 1$. Similarly, using the independence property of PPPs,

$$\mathbb{P}(Q_\ell(a) \vee Q_r(b) \le x) = \exp\{-(g(0,a)+g(b,1))x^{-(\alpha-1)}\}. \tag{7.4}$$

Combining (7.3) and (7.4) in (7.2) by conditioning on $Q_\ell(a) \vee Q_r(b)$, we obtain

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}(I_n/n \in (a,b)) &= 1 - \int_0^\infty (\alpha-1)x^{-\alpha}(g(0,a)+g(b,1))\exp\{-g(0,1)x^{-(\alpha-1)}\}\mathrm{d}x\\ &= \frac{g(a,b)}{g(0,1)}.\end{aligned}$$

Then, using the variable transform $s = x^{1/\theta_m}$ yields

$$g(a,b) = \theta_m \int_{a^{1/\theta_m}}^{b^{1/\theta_m}} s^{(\theta_m-(\alpha-1))-1}(1-s)^{\alpha-1}\mathrm{d}s = \frac{\Gamma(\theta_m-(\alpha-1))\Gamma(\alpha)}{\Gamma(\theta_m)}\mathbb{P}\big(B^{\theta_m} \in (a,b)\big),$$

where Γ is the Gamma function, from which it follows that $I \overset{d}{=} B^{\theta_m}$, with B a Beta$(\theta_m - (\alpha-1), \alpha)$ random variable. Via a similar approach, redefining $M_n(a,b)$ and $M(a,b)$ accordingly for $\alpha \in (1,2)$, we can show I_n/n converges in distribution when $\alpha \in (1,2)$, though it is not possible to find an explicit expression for the law of I'. Finally, we address the joint convergence of $(I_n/n, \max_{i\in[n]} \mathcal{Z}_n(i)/u_n)$. We let $0 < c < d < \infty$ and define the events

$$E_n(a,b,c,d) =: \Big\{\max_{an<i<bn} \mathcal{Z}_n(i)/u_n \in (c,d)\Big\}, \quad E(a,b,c,d) := \Big\{Q(a,b) \in (c,d)\Big\}. \tag{7.5}$$

We can then write, using these events and the events in (7.1) and letting $A := (a,b)\times(c,d)$,

$$\mathbb{P}\Big((I_n/n, \max_{i\in[n]} \mathcal{Z}_n(i)/u_n) \in A\Big) = \mathbb{P}(M_n(a,b) \cap E_n(a,b,c,d)),$$

which converges to $\mathbb{P}(M(a,b) \cap E(a,b,c,d))$ as n tends to infinity by the same argument as provided for the limit in (7.2). Again, by conditioning on $Q_\ell(a) \vee Q_r(b)$ and using (7.4), we find

$$\begin{aligned}&\mathbb{P}(M(a,b) \cap E(a,b,c,d))\\ &= \mathbb{P}(E(a,b,c,d))\mathbb{P}(Q_\ell(a) \vee Q_r(b) \le c)\\ &\quad + \int_c^d \mathbb{P}(E(a,b,x,d))\,(\alpha-1)x^{-\alpha}(g(0,a)+g(b,1))\exp\{-(g(0,a)+g(b,1))x^{-(\alpha-1)}\}\mathrm{d}x.\end{aligned}$$

Using (7.3), (7.4) and (7.5) the first term on the right-hand side equals

$$\begin{aligned}&(\exp\{-g(a,b)d^{-(\alpha-1)}\} - \exp\{-g(a,b)c^{-(\alpha-1)}\})\exp\{-(g(0,a)+g(b,1))c^{-(\alpha-1)}\}\\ &\quad = \exp\{-g(a,b)d^{-(\alpha-1)} - (g(0,a)+g(b,1))c^{-(\alpha-1)}\} - \exp\{-g(0,1)c^{-(\alpha-1)}\}.\end{aligned} \tag{7.6}$$

For the second term, we realise we can write

$$\mathbb{P}(E(a,b,x,d)) = \mathbb{P}(Q(a,b) \in (x,d)) = \mathbb{P}(Q(a,b) \le d)) - \mathbb{P}(Q(a,b) \le x)\,,$$

so that we can split the integral into two parts. The first part, using (7.3) and (7.4), becomes

$$\begin{aligned}\mathbb{P}(Q(a,b) \le d) &\int_c^d (\alpha-1)x^{-\alpha}(g(0,a)+g(b,1))\exp\{-(g(0,a)+g(b,1))x^{-(\alpha-1)}\}\mathrm{d}x \\ &= \exp\{-g(0,1)d^{-(\alpha-1)}\} - \exp\{-g(a,b)d^{-(\alpha-1)} - (g(0,a)+g(b,1))c^{-(\alpha-1)}\},\end{aligned} \tag{7.7}$$

and the second part equals

$$\begin{aligned}\int_c^d &\mathbb{P}(Q(a,b) \le x)(\alpha-1)x^{-\alpha}(g(0,a)+g(b,1))\exp\{-(g(0,a)+g(b,1))x^{-(\alpha-1)}\}\mathrm{d}x \\ &= \int_c^d (\alpha-1)x^{-\alpha}(g(0,a)+g(b,1))\exp\{-g(0,1)x^{-(\alpha-1)}\}\mathrm{d}x \\ &= \Big(1-\frac{g(a,b)}{g(0,1)}\Big)(\exp\{-g(0,1)d^{-(\alpha-1)}\} - \exp\{-g(0,1)c^{-(\alpha-1)}\}).\end{aligned} \tag{7.8}$$

Combining (7.6), (7.7) and (7.8), yields as n tends to infinity,

$$\begin{aligned}\mathbb{P}\Big((I_n/n, \max_{i\in[n]} \mathcal{Z}_n(i)/u_n) \in A\Big) &\to \frac{g(a,b)}{g(0,1)}(\exp\{-g(0,1)d^{-(\alpha-1)}\} - \exp\{-g(0,1)c^{-(\alpha-1)}\}) \\ &= \mathbb{P}(I \in (a,b))\,\mathbb{P}\Big(\max_{(t,f)\in\Pi} f(t^{-1/\theta_m}-1) \in (c,d)\Big),\end{aligned}$$

where the final step regarding the law of the maximum of the PPP, a Fréchet distribution with shape parameter $\alpha-1$ and scale parameter $g(0,1)^{1/(\alpha-1)} = (\Gamma(\theta_m-(\alpha-1))\Gamma(\alpha)/\Gamma(\theta_m))^{1/(\alpha-1)}$, follows from a similar argument as in (7.3). As before, redefining the events in (7.1) and (7.5) accordingly and using the same steps yields the joint convergence of $(I_n/n, \max_{i\in[n]} \mathcal{Z}_n(i)/n)$ when $\alpha \in (1,2)$, which concludes the proof of Theorem 2.7. □

References

[1] K. B. Athreya, A. P. Ghosh, and S. Sethuraman. Growth of preferential attachment random graphs via continuous-time branching processes. *Proc. Indian Acad. Sci. Math. Sci.*, 118(3):473–494, 2008.

[2] A. L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[3] S. Bhamidi. Universal techniques to analyze preferential attachment trees: Global and local analysis. *In preparation*, 2007.

[4] G. Bianconi and A. L. Barabási. Competition and multiscaling in evolving networks. *EPL (Europhysics Letters)*, 54(4):436, 2001.

[5] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular variation*, volume 27 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1989.

[6] B. Bollobás, O. Riordan, J. Spencer, and G. Tusnády. The degree sequence of a scale-free random graph process. *Random Structures Algorithms*, 18(3):279–290, 2001.

[7] C. Borgs, J. Chayes, C. Daskalakis, and S. Roch. First to market is not everything: an analysis of preferential attachment with fitness. In *STOC'07—Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, pages 135–144. ACM, New York, 2007.

[8] M. Deijfen, H. van den Esker, R. van der Hofstad, and G. Hooghiemstra. A preferential attachment model with random initial degrees. *Ark. Mat.*, 47(1):41–72, 2009.

[9] S. Dereich. Preferential attachment with fitness: unfolding the condensate. *Electron. J. Probab.*, 21, 2016.

[10] S. Dereich, C. Mailler, and P. Mörters. Nonextensive condensation in reinforced branching processes. *Ann. Appl. Probab.*, 27(4):2539–2568, 2017.

[11] S. Dereich and P. Mörters. Random networks with sublinear preferential attachment: degree evolutions. *Electron. J. Probab.*, 14:1222–1267, 2009.

[12] S. Dereich and M. Ortgiese. Robust analysis of preferential attachment models with fitness. *Combin. Probab. Comput.*, 23(3):386–411, 2014.

[13] G. Ergün and G. Rodgers. Growing random networks with fitness. *Physica A: Statistical Mechanics and its Applications*, 303(1-2):261–272, 2002.

[14] A. Gut. *Probability: a graduate course*. Springer Texts in Statistics. Springer, New York, second edition, 2013.

[15] R. van der Hofstad. *Random graphs and complex networks. Vol. 1*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2017.

[16] G. J. O. Jameson. Inequalities for gamma function ratios. *Amer. Math. Monthly*, 120(10):936–940, 2013.

[17] K. Joag-Dev and F. Proschan. Negative association of random variables, with applications. *Ann. Statist.*, 11(1):286–295, 1983.

[18] A. Klenke. *Probability theory*. Universitext. Springer, London, second edition, 2014.

[19] C. Mailler, P. Mörters, and A. Senkevich. Competing growth processes with random growth rates and random birth times. *Preprint, arXiv:1909.07690*, 2019.

[20] T. F. Móri. The maximum degree of the Barabási-Albert random tree. *Combin. Probab. Comput.*, 14(3):339–348, 2005.

[21] S. I. Resnick. *Extreme values, regular variation and point processes*. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2008. Reprint of the 1987 original.

[22] D. Revuz and M. Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, Berlin, third edition, 1999.

[23] D. Sénizergues. Geometry of weighted recursive and affine preferential attachment trees. *Preprint, arXiv:1904.07115*, 2019.

[24] A. W. van der Vaart. *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, 1998.

[25] J. G. Wendel. Note on the gamma function. *Amer. Math. Monthly*, 55:563–564, 1948.

[26] D. Williams. *Probability with martingales*. Cambridge Mathematical Textbooks. Cambridge University Press, 1991.

Acknowledgments. We would like to thank the referees who carefully read the paper for their suggestions, which helped to improve the quality and the readability of the paper.

2.2 Conclusion

We presented a comprehensive and self-contained analysis of the degree distribution and maximum degree of three variants of preferential attachment models with additive fitness. In this presentation we outlined three different phases or regimes, for which different behaviour can be observed.

Theorem 2.4 states the convergence of the degree-weighted fitness measure, the empirical fitness-in-degree distribution and the empirical degree distribution in the weak and strong disorder regime, as well as at the critical point between these two regimes. Theorem 2.6 further investigates the asymptotic behaviour of the limiting degree distribution in the weak and strong disorder regimes and at the critical point between these two regimes, showing it is a power law with a different exponent depending on the particular regime. Importantly, it shows that the power-law exponent remains 'unaffected' by the additive fitness (in comparison to the affine preferential attachment models discussed in the introduction of the chapter) in the weak disorder regime, whereas the additive fitness solely determines the power-law exponent in the strong disorder regime. The behaviour at the critical point between the weak and strong regime, which is different from the behaviour in either of the regimes, begs the question what the maximum degree behaves like at this critical point.

The behaviour of the maximum degree in each of the three regimes is presented in Theorem 2.7. The contrast with the behaviour of the degrees of fixed vertices, as shown in Equations (2.8) and (2.9), provides clear insight in the differences between the three regimes. In the extreme disorder regime, fixed vertices only attain a finite degree almost surely, whereas the maximum degree vertex establishes connections with a strictly positive proportion of all vertices in the graph and thus attracts a strictly positive proportion of all edges in the graph. Though this behaviour can be compared to the super-linear preferential attachment models studied in, among others, [120], the main difference is that the location of the maximum degree keeps changing, and in fact is of order n, due to the random environment. In the super-linear models, however, persistence occurs.

In the strong disorder regime the location of the maximum degree grows linearly in n as well, but the maximum degree grows sub-linear (though still faster than the degree of any fixed vertex). Finally, in the weak disorder regime, the maximum degree grows at the same rate as any fixed vertex and, as is the case for the affine preferential attachment models discussed in the introduction of this chapter, persistence occurs, where there exists a fixed vertex that attains the maximum degree for all but finitely many steps. This is the most notable difference between the weak disorder regime and the strong and extreme disorder regime and shows the effect of the random environment most clearly.

An indirect comparison can be made between the three regimes of the phase transition this model undergoes and the three phases observed in the Bianconi-Barabási model as investigated in [23]. Roughly speaking, the weak, strong and extreme disorder regimes can be compared to the first-mover-advantage, fit-get-richer and innovation-pays-off phases, though the underlying conditions giving rise to these phases as well as the behaviour observed in the two models is different.

To conclude, we have shown that the introduction of additive fitness to the preferential attachment model allows for richer behaviour, which can serve as a more natural and more detailed explanation of the underlying mechanisms governing real-world networks.

Chapter 3

The maximal degree in random recursive graphs with random weights

In this chapter we consider weighted random graphs, a more general form of the weighted recursive tree model defined in [71]. We investigate its degree distribution and its maximum degree. Depending on the distribution of the vertex-weights, we either use similar techniques to those developed in Chapter 2, or we adopt techniques used in [44] for random recursive trees and directed acyclic graphs. The following preprint, which is joint work with Marcel Ortgiese, is available on the arXiv [96].

3.1 Outline of the article

The weighted recursive graph model, as defined in the tree case in [71], is comparable to the models discussed in the previous chapter. Again, we consider a graph process $(\mathcal{G}_n)_{n\in\mathbb{N}}$, starting with a single node and no edges, in which vertices enter the graph one by one and connect to $m \in \mathbb{N}$ predecessors when they enter the graph. Also, every vertex $i \in \mathbb{N}$ is assigned an i.i.d. positive fitness W_i, which we refer to as its vertex-weight from now on. A new vertex n chooses each of the m predecessors to connect to with a probability proportional to the vertex-weight of the predecessor. As in Chapter 2, we also allow for a 'Bernoulli' model in which the new vertex connects to each predecessor with a probability proportional to its vertex-weight and where connections are negatively correlated. As in the previous chapter, we are interested in the behaviour of the empirical degree distribution and the maximum degree distribution.

This model relates to the previously-investigated preferential attachment models with additive fitness in the sense that this model omits the feedback effect of the increasing degree. As shown in the Chapter 2, the fitness in these preferential attachment models is only able to influence the overall behaviour of the graph in a significant way when its distribution is sufficiently heavy-tailed. This is reflected in the three regimes discussed there. The fitness affects the evolution of the graph in a very subtle way compared to the much stronger preferential attachment mechanism that is at the heart of the model, so that only heavy-tailed distributions have a significant impact. Here, we study a model in which the preferential attachment mechanism is omitted, to allow for the subtleties of the fitness/vertex-weight distribution to come through.

As a result, we are able to identify a large number of classes of vertex-weight distributions for which different behaviour of the degree distribution and maximum degree can be observed. The most distinct difference in the behaviour of the maximum degree can be observed between almost surely bounded and almost surely unbounded vertex-weights. In the latter case, the fact that vertices with arbitrarily large vertex-weights can appear implies the vertex-weights have a direct influence on the behaviour of the maximum degree. Using techniques including point process convergence and extreme value theory, we are able to obtain the behaviour of the maximum degree for many classes of vertex-weight distributions with unbounded support.

In the case of almost surely bounded vertex-weights, we observe that the behaviour of the maximum degree and the degree distribution is similar to the random recursive tree (and the directed acyclic graph, its multigraph counterpart), which can be interpreted as a weighted recursive graph where all vertex-weights are almost surely one. As the weights are bounded, their influence on the graph process is significantly different from the case where the weights are unbounded. Using *only* the fact that the vertex-weights are bounded, we are able to obtain the first order behaviour of the maximum degree, adapting techniques developed in [44] for the random recursive tree and directed acyclic graph.

The techniques used in the analysis of the WRG model are robust, in the sense that the results for the tree case ($m = 1$) follow directly from the analysis of the multigraph case. This holds when the vertex-weights are unbounded as well as when the vertex-weights are bounded. Though other techniques, such as embedding the graph process in a continuous-time branching process, can also be used to analyse the tree case, these techniques generally do not allow for an extension to the multigraph case.

In Section 2 we present the main results regarding the degree distribution and the maximum degree. Section 4 is then devoted to proving the results related to the degree distribution. Sections 5 and 6 provide the necessary technical results required to prove the results related to the maximum degree, which is done in Section 7.

Appendix 6B: Statement of Authorship

This declaration concerns the article entitled:									
The maximal degree in random recursive graphs with random weights									
Publication status (tick one)									
Draft manuscript		**Submitted**		**In review**	x	**Accepted**		**Published**	
Publication details (reference)	Preprint: arXiv:2007.05438 Authors: Bas Lodewijks, Marcel Ortgiese								
Copyright status (tick the appropriate statement)									
I hold the copyright for this material	x	Copyright is retained by the publisher, but I have been given permission to replicate the material here							
Candidate's contribution to the paper (provide details, and also indicate as a percentage)	The candidate predominantly executed the… Formulation of ideas: 80%. The candidate drove the formulation of ideas to go into this paper for a large part. Design of methodology: 80%. The candidate has largely contributed to the development of the theoretical methodology in this paper, working out most proofs in detail to a large extent. Experimental work: N/A Presentation of data in journal format: 80%. The candidate has written the initial draft, after which adjustments and changes were proposed by the other authors.								
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.								
Signed	Bas Lodewijks	**Date**	01/07/2021						

THE MAXIMAL DEGREE IN RANDOM RECURSIVE GRAPHS WITH RANDOM WEIGHTS

BAS LODEWIJKS AND MARCEL ORTGIESE

Abstract. We study a generalisation of the random recursive tree (RRT) model and its multigraph counterpart, the uniform directed acyclic graph (DAG). Here, vertices are equipped with a random vertex-weight representing initial inhomogeneities in the network, so that a new vertex connects to one of the old vertices with a probability that is proportional to their vertex-weight. We first identify the asymptotic degree distribution of a uniformly chosen vertex for a general vertex-weight distribution. For the maximal degree, we distinguish several classes that lead to different behaviour: For bounded vertex-weights we obtain results for the maximal degree that are similar to those observed for RRTs and DAGs. If the vertex-weights have unbounded support, then the maximal degree has to satisfy the right balance between having a high vertex-weight and being born early.

For vertex-weights in the Fréchet maximum domain of attraction the first order behaviour of the maximal degree is random, while for those in the Gumbel maximum domain of attraction the leading order is deterministic. Surprisingly, in the latter case, the second order is random when considering vertices in a compact window in the optimal region, while it becomes deterministic when considering all vertices.

1. Introduction

A random recursive tree (RRT) is a growing random tree model in which one starts with a single vertex, denoted as the root, and for $n \geq 2$, adds a vertex n which is then connected to a vertex chosen uniformly at random among the vertices $\{1, \ldots, n-1\}$. Since the selection is uniform, this model is also known as the uniform attachment tree or uniform random recursive tree. Its multigraph counterpart known as uniform directed acyclic graphs (DAGs or uniform DAGs) was introduced by Devroye and Lu in [8] and allows for an incoming vertex to connect to k predecessors. The RRT was first introduced by Na and Rapoport in 1970 [23] and has since attracted a wealth of interest, uncovering the behaviour of many of its properties, including, among others: the number of leaves, profile of the tree, height of the tree, vertex degrees and the size of sub-trees. [27] and the more recent [9] provide good surveys on the topic.

In this paper we study a more general model, the weighted recursive graph (WRG), which can be interpreted as a random recursive tree (or uniform DAG) in a random environment. Here, we assign to every vertex a random, independent non-negative vertex-weight and incoming vertices are connected to predecessors not uniformly at random but with a probability proportional to the vertex-weights. The tree case of this model, the weighted recursive tree (WRT), was originally introduced by Borovkov and Vatutin in [6] (with a deterministic weight for the first vertex), as well as by Hiesmayr and Işlak in [15]. Another type of WRT was introduced by Borovkov and Vatutin in [5], where the vertex-weights have a specific product-form. These *weighted* recursive tree models have received far less attention overall, though it allows for much more diverse behaviour.

Recent work on weighted recursive trees includes [21], [26] and [24] where the profile and height of the tree are analysed as well as vertex degrees. Additionally, Iyer [18] and Iyer

Date: August 27, 2021.

Key words and phrases. Weighted recursive graph, Weighted random recursive tree, Random recursive graph, Uniform DAG, Maximum degree, Degree distribution, Random environment.

and Fountoulakis [11] study degree distributions of many weighted growing tree models and the weighted recursive tree is a particular example.

In what follows we first analyse the degree distribution of a uniformly chosen vertex and the behaviour of the maximum degree in WRGs, which recovers and extends results on the degree distribution of RRTs and WRTs as well as the maximum degree in RRTs. Degree distributions in RRTs have been studied in [12, 20, 22] and [23] and, as mentioned above, Iyer (and Fountoulakis) study the degree distributions of a very general class of weighted growing trees in [18] and [11]. In this paper we extend the results on degree distributions in random recursive trees and weighted recursive trees to their multigraph counterparts, the DAG and WRG models, respectively.

Szymański [28] was the first to obtain results on the growth rate of the maximum degree in RRTs. These results were later extended in [8] and finer properties of high degrees were analysed in [13], [1] and [10]. Recently, Banerjee and Bhamidi [3] studied the occurrence of persistence in growing random networks. Here persistence means that there exists a vertex in the network whose degree is maximal for all but finitely many steps. Also, [3] presents results describing the growth rate of the location of the maximum degree (the index of vertices attaining the maximum degree) in RRTs. In WRTs, the behaviour of degrees and the maximum degree has received attention from Sénizergues in [26], where the vertex-weights need not be i.i.d. random variables but can satisfy more general conditions. In particular cases, it is shown that these graphs are equivalent in law to preferential attachment models with additive fitness (PAF), also studied in [19].

Here, we extend and generalise the results of Devroye and Lu in [8] to WRGs and analyse the growth of the maximal degree for a broad range of vertex-weights distributions. Moreover, we identify the location of the maximal site in many cases, a result which was shown (among others) for constant weight models in [3].

Our methods are related to the analysis of the preferential attachment with additive fitness carried out in [19]. For these preferential attachment models, the attachment probabilities are proportional to the degree plus a random weight (fitness). In these models, we distinguish three different regimes: first of all a *weak disorder regime*, where the preferential attachment mechanism dominates (and there is persistence). This is closely related to the work of [26], which in turn corresponds to a WRT where the partial sums of the weights is at most of order n^γ for $\gamma \in (0,1)$. Moreover, in [19] we identify a *strong and extreme disorder regime* where the influence of the random weights takes over, which appears when the distribution of the weights is sufficiently heavy-tailed.

For WRGs there is no preferential attachment component for the vertex-weights to compete with so that the influence of the random weights is more immediate and already appears for less heavy-tailed weights. More precisely, for the maximal degree we distinguish three regimes: for bounded weights the system behaves similarly to a RRT, whereas for unbounded weights that are in the domain of attraction of a Gumbel distribution the maximal degree grows faster and we can identify the time when the maximizing vertex comes into the system. Finally, in the case when the weights are in the domain of attraction of a Fréchet distribution, the behaviour is similar to (but not exactly the same as) the preferential attachment with additive fitness in the strong and extreme disorder regimes and the leading asymptotics of the maximal degree is random and we identify the limit as a functional of a Poisson point process. In contrast to the latter, in the case of Gumbel weights, the first order growth of the degrees is deterministic and by comparison to the case of the maximum of i.i.d. weights, it would be natural to conjecture that the second order is random. We confirm this observation for two special sub-cases of Gumbel weights. However, this result is only true when we consider a compact window in the region of indices that should correspond to the maximal one. Finally, we identify the true second order and, somewhat surprisingly, it is also deterministic. This behaviour comes from the fact that we have to consider a much larger optimal window than initially suspected.

Our results for the degree distribution follow by adjusting the proofs in [19], as at least on the level of degree distributions the WRG model is essentially a simpler model compared to the PAF models. The results for the maximum degree in the case of bounded weights follow with similar techniques as in [8], which can be extended to WRGs. The main contribution of this paper is to understand the first and second order asymptotics of the maximal degree in the case when the weights are unbounded and satisfy suitable regularity assumptions. For unbounded weights, the system is driven by the competition between the benefit of being an old vertex and so having time to accumulate a high degree and the benefit of being a young vertex with a large weight. To understand the first order of growth (as well as the second order when considering a compact window), we show concentration of the degrees around the conditional means (when conditioning on the weights) and then in a second step analyse the conditional mean degree using extreme value theory (similarly as in [19]). However, as the weights have a more immediate impact, the results become more dependent on the exact distribution of weights chosen and thus require more intricate calculations. Finally, to obtain the true second order asymptotics in the Gumbel case, we can no longer rely on the elegant tools of convergence to Poisson processes from extreme value theory and instead have to carefully keep track of errors made in the corresponding approximations.

Notation. Throughout the paper we use the following notation: we let $\mathbb{N} := \{1, 2, \ldots\}$ be the natural numbers, set $\mathbb{N}_0 := \{0, 1, \ldots\}$ to include zero and let $[t] := \{i \in \mathbb{N} : i \leq t\}$ for any $t \geq 1$. For $x \in \mathbb{R}$, we let $\lceil x \rceil := \inf\{n \in \mathbb{Z} : n \geq x\}$ and $\lfloor x \rfloor := \sup\{n \in \mathbb{Z} : n \leq x\}$. Moreover, for sequences $(a_n, b_n)_{n\in\mathbb{N}}$ we say that $a_n = o(b_n), a_n \sim b_n, a_n = \mathcal{O}(b_n)$ if $\lim_{n\to\infty} a_n/b_n = 0, \lim_{n\to\infty} a_n/b_n = 1$ and if there exist constants $C > 0, n_0 \in \mathbb{N}$ such that $a_n \leq C b_n$ for all $n \geq n_0$, respectively. For random variables $X, (X_n)_{n\in\mathbb{N}}$ we denote $X_n \xrightarrow{d} X, X_n \xrightarrow{\mathbb{P}} X$ and $X_n \xrightarrow{a.s.} X$ for convergence in distribution, probability and almost sure convergence of X_n to X, respectively. Also, we write $X_n = o_{\mathbb{P}}(1)$ if $X_n \xrightarrow{\mathbb{P}} 0$. Throughout, we denote by $(W_i)_{i\in\mathbb{N}}$ i.d.d. random variables and use the conditional probability measure $\mathbb{P}_W(\cdot) := \mathbb{P}(\,\cdot\,|(W_i)_{i\in\mathbb{N}})$ and conditional expectation $\mathbb{E}_W[\cdot] := \mathbb{E}\,[\,\cdot\,|(W_i)_{i\in\mathbb{N}}]$.

2. Definitions and main results

The Weighted Recursive Graph (WRG) model is a growing random graph model that is a generalisation of the random recursive tree (RRT) and the uniform directed acyclic graph (DAG) models in which vertices are assigned (random) weights and new vertices connect with existing vertices with a probability proportional to the vertex-weights.

We define the WRG model as follows:

Definition 2.1 (Weighted Recursive Graph)**.** Let $(W_i)_{i\geq 1}$ be a sequence of i.i.d. copies of a non-negative random variable W such that $\mathbb{P}(W > 0) = 1$, let $m \in \mathbb{N}$ and set

$$S_n := \sum_{i=1}^{n} W_i.$$

We construct the *Weighted Recursive Graph* as follows:

1) Initialise the graph with a single vertex 1, the root, and assign to the root a vertex-weight W_1. We let $\mathcal{G}_1$ denote this graph. .
2) For $n \geq 1$, introduce a new vertex $n+1$ and assign to it the vertex-weight W_{n+1} and m half-edges. Conditionally on $\mathcal{G}_n$, independently connect each half-edge to some vertex $i \in [n]$ with probability W_i/S_n. Let $\mathcal{G}_{n+1}$ denote this graph.

We treat $\mathcal{G}_n$ as a directed graph, where edges are directed from new vertices towards old vertices.

Remark 2.2. (i) Note that the edge connection probabilities remain unchanged if we multiply each weight by the same constant. In particular, we may assume, without loss of generality, that $x_0 : \sup\{x \in \mathbb{R} \,|\, \mathbb{P}(W \leq x) < 1\}$ is either 1 (when the weights are almost surely bounded) or ∞ (when the weights are almost surely unbounded. Similarly, in the latter case of unbounded vertex-weights we can assume that $\mathbb{E}[W] = 1$.

(ii) It is possible to extend the definition of the WRG to the case of *random out-degree* and the results presented in this paper still hold. Namely, we can allow that vertex $n+1$ connects to *every* vertex $i \in [n]$ independently with probability W_i/S_n. At the start of sections dedicated to proving the results we present below, we discuss why the results hold for the random out-degree model as well.

To formulate our results we need to assume that the distribution of the weights is sufficiently regular, allowing us to control their extreme value behaviour.

Assumption 2.3 (Vertex-weight distributions)**.** The vertex-weights $W, (W_i)_{i\in\mathbb{N}}$ satisfy one of the following conditions:

(**Bounded**) The vertex-weights are almost surely bounded, i.e. $x_0 := \sup\{x \in \mathbb{R} \,|\, \mathbb{P}(W \leq x) < 1\} < \infty$. Without loss of generality, we can assume that $x_0 = 1$.
Within this class, we can further identify vertex-weight distributions that belong to the Weibull maximum domain of attraction (MDA) and Gumbel MDA.

(**Gumbel**) The vertex-weights follow a distribution that belongs to the Gumbel maximum domain of attraction (MDA) such that $x_0 = \infty$. Without loss of generality, $\mathbb{E}[W] = 1$. This implies that there exist sequences $(a_n, b_n)_{n\in\mathbb{N}}$, such that

$$\frac{\max_{i\in[n]} W_i - b_n}{a_n} \xrightarrow{d} \Lambda,$$

where Λ is a Gumbel random variable.
Within this class, we further distinguish the following three (non-exhaustive) sub-classes:

- (**SV**) $b_n \sim \ell(\log n)$ where ℓ is an increasing function that is slowly-varying at infinity, i.e. $\lim_{x\to\infty} \ell(cx)/\ell(x) = 1$ for all $c > 0$.
- (**RV**) There exist $a, c_1, \tau > 0$, and $b \in \mathbb{R}$ such that
$$\mathbb{P}(W \geq x) \sim a x^b \mathrm{e}^{-(x/c_1)^\tau} \quad \text{as } x \to \infty.$$
- (**RaV**) There exist $a, c_1 > 0, b \in \mathbb{R}$, and $\tau > 1$ such that
$$\mathbb{P}(W \geq x) \sim a(\log x)^b \mathrm{e}^{-(\log(x)/c_1)^\tau} \quad \text{as } x \to \infty.$$

(**Fréchet**) The vertex-weights follow a distribution that belongs to the Fréchet MDA. Without loss of generality, $\mathbb{E}[W] = 1$ (given that $\mathbb{E}[W] < \infty$ is satisfied). This implies that there exists a non-negative function $\ell(x)$ that is slowly-varying at infinity and some $\alpha > 1$, such that

$$\mathbb{P}(W \geq x) = \ell(x) x^{-(\alpha-1)}.$$

Moreover, if we let $u_n := \sup\{t \in \mathbb{R} : \mathbb{P}(W \geq t) \geq 1/n\}$,

$$\max_{i\in[n]} W_i/u_n \xrightarrow{d} \Phi_{\alpha-1},$$

where $\Phi_{\alpha-1}$ is a Fréchet random variable with exponent $\alpha - 1$.

Remark 2.4. Note that [29] shows (with a slight error in the paper in that the $\log a$ term below is a $\log \tau$ in [29]) that if the weight distribution satisfies the assumption (**RV**), then we can choose

$$a_n = c_2(\log n)^{1/\tau - 1}, \qquad b_n = c_1(\log n)^{1/\tau} + a_n((b/\tau)\log\log n + b\log c_1 + \log a), \qquad (2.1)$$

for the same constants as above and $c_2 := c_1/\tau$. Moreover, in the **(RaV)** sub-case, we can choose

$$\begin{aligned} b_n &= \exp\{c_1(\log n)^{1/\tau} + c_2(\log n)^{1/\tau-1}((b/\tau)\log\log n + b\log c_1 + \log a)\}, \\ a_n &= c_2(\log n)^{1/\tau-1}b_n, \end{aligned} \tag{2.2}$$

In particular, the three sub-cases in the **(Gumbel)** case, **(SV)**, **(RV)** and **(RaV)**, can be distinguished as $b_n = g(\log n), a_n = \widetilde{g}(\log n)$, with $g, \widetilde{g}$ slowly-varying, regularly-varying and rapidly-varying functions at infinity, respectively. Note that in all cases, a_n and b_n itself are slowly varying at infinity. In the **(RV)** sub-case, we very often use the asymptotic equivalence for b_n, that is, $b_n \sim c_1(\log n)^{1/\tau}$. Moreover, in the **(RaV)** sub-case, we can think of b_n as $\exp\{(\log n)^{1/\tau}\ell(\log n)\}$ and a_n as $c_2(\log n)^{1/\tau-1}b_n$.

Furthermore, as noted in the assumption, we recall that the three sub-cases within the **(Gumbel)** case do not cover all possible distributions in the Gumbel MDA with unbounded support. As an example, $\mathbb{P}(W \geq x) = a\exp\{-(x/c_1)^\tau + (x/c_1)^{\tau-1}\}$ as $x \to \infty$ is a distribution within the Gumbel MDA, but does not satisfy any of the three sub-cases. On the other hand the **(Bounded)** and **(Fréchet)** classes are exhaustive.

We now present the results for the degree distribution and the maximum degree in the WRG model. In comparison to the preferential attachment with additive fitness (PAF) models as studied in [19], vertex-weights with a distribution with a 'thin' tail, i.e. distributions with exponentially decaying tails or bounded support, now can also exert their influence on the behaviour of the system.

Throughout, we write

$$\mathcal{Z}_n(i) := \text{ in-degree of vertex } i \text{ in } \mathcal{G}_n.$$

We prefer to work with the in-degree as it then is easier to (in principle) generalize our methods to graphs with random out-degree. Obviously, if the out-degree is fixed, we can recover the results for the degree from our results for $\mathcal{Z}_n(i)$.

The first result deals with the degree distribution of the WRG model. Let us first introduce the following measures and quantities:

$$\Gamma_n := \frac{1}{n}\sum_{i=1}^n \mathcal{Z}_n(i)\delta_{W_i}, \quad \Gamma_n^{(k)} := \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{\mathcal{Z}_n(i)=k\}}\delta_{W_i}, \quad p_n(k) := \Gamma_n^{(k)}([0,\infty)),$$

where δ is a Dirac measure, and which correspond to the empirical weight distribution of a vertex sampled weighted by its in-degree, then the joint empirical vertex-weight and in-degree distribution and finally the empirical degree distribution. We can then formulate the following theorem:

Theorem 2.5 (Degree distribution in WRGs)**.** *Consider the WRG model in Definition 2.1 and suppose that the vertex-weights have finite mean and denote their distribution by μ. Then, for any $k \in \mathbb{N}_0$, almost surely as $n \to \infty$,*

$$\Gamma_n \to \Gamma, \qquad \Gamma_n^{(k)} \to \Gamma^{(k)}, \quad \text{and} \quad p_n(k) \to p_k, \tag{2.3}$$

where the first two statements hold with respect to the topology of weak convergence and the limits are given as

$$\Gamma(\mathrm{d}x) := \frac{xm}{\mathbb{E}[W]}\mu(\mathrm{d}x), \qquad \Gamma^{(k)}(\mathrm{d}x) = \frac{\mathbb{E}[W]/m}{\mathbb{E}[W]/m + x}\left(\frac{x}{\mathbb{E}[W]/m+x}\right)^k \mu(\mathrm{d}x), \tag{2.4}$$

and

$$p_k = \int_0^\infty \frac{\mathbb{E}[W]/m}{\mathbb{E}[W]/m + x}\left(\frac{x}{\mathbb{E}[W]/m+x}\right)^k \mu(\mathrm{d}x). \tag{2.5}$$

Finally, let the vertex-weight distribution be a power law as in the **(Fréchet)** *case of Assumption 2.3 with $\alpha \in (1,2)$, such that there exists an $x_l > 0$ with $\mu(x_l,\infty) = 1$, i.e. the vertex-weights are bounded away from zero almost surely. Let U_n be a vertex selected* *uniformly at random from* $[n]$*, let* $\varepsilon > 0$ *and let* $E_n := \{\mathcal{Z}_n(U_n) = 0\}$*. Then, for all* n *sufficiently large,*

$$\mathbb{P}(E_n) \geq 1 - Cn^{-((2-\alpha)\wedge(\alpha-1))/\alpha+\varepsilon}, \tag{2.6}$$

for some constant $C > 0$*.*

An important question regarding the degree distribution p_k, as in (2.5), is its asymptotic behaviour as $k \to \infty$. As it turns out, the answer depends on the particular choice of the distribution of the random variable W. Before presenting a theorem dedicated to the asymptotic behaviour of the limiting degree distribution p_k, we first introduce the following general lemma, which allows us to distinguish several different sub-cases of bounded weights.

Lemma 2.6. *Let* W *be a non-negative random variable such that* $x_0 := \sup\{x > 0 : \mathbb{P}(W \leq x) < 1\} < \infty$*. Then, the distribution of* W *belongs to the Weibull (resp. Gumbel) MDA if and only if* $(x_0 - W)^{-1}$ *is a non-negative random variable with a distribution with unbounded support that belongs to the Fréchet (resp. Gumbel) MDA.*

We are now ready to present the results on the asymptotics of p_k.

Theorem 2.7 (Asymptotic behaviour of p_k)**.** *Consider the WRG with vertex-weights* $(W_i)_{i\in\mathbb{N}}$*, which are i.i.d. copies of a non-negative random variable* W *such that* $\mathbb{E}[W] < \infty$*. We consider the different cases of Assumption 2.3.*

(**Bounded**) *Let* $\theta_m := 1 + \mathbb{E}[W]/m$ *and recall that* $x_0 = \sup\{x > 0 : \mathbb{P}(W \leq x) < 1\} = 1$*.*

- *When* W *belongs to the Weibull MDA with parameter* $\alpha > 1$*, for all* $k > m/\mathbb{E}[W]$*,*

$$\underline{L}(k)k^{-(\alpha-1)}\theta_m^{-k} \leq p_k \leq \overline{L}(k)k^{-(\alpha-1)}\theta_m^{-k}, \tag{2.7}$$

  where $\underline{L}, \overline{L}$ *are slowly varying at infinity.*

- *When* W *belongs to the Gumbel MDA,*
  - (i) *If* $(1-W)^{-1}$ *satisfies the* (**RV**) *sub-case with parameter* $\tau > 0$*, then for* $\gamma := 1/(\tau+1)$*,*

$$p_k = \exp\Big\{ -(1+o(1))\frac{\tau^\gamma}{1-\gamma}\Big(\frac{(1-\theta_m^{-1})k}{c_1}\Big)^{1-\gamma}\Big\}\theta_m^{-k}. \tag{2.8}$$

  - (ii) *If* $(1-W)^{-1}$ *satisfies the* (**RaV**) *sub-case with parameter* $\tau > 1$*,*

$$p_k = \exp\Big\{-\Big(\frac{\log k}{c_1}\Big)^\tau\Big(1-\tau(\tau-1)\frac{\log\log k}{\log k}+\frac{K_{\tau,c_1,\theta_m}}{\log k}(1+o(1))\Big)\Big\}\theta_m^{-k}, \tag{2.9}$$

    where $K_{\tau,c_1,\theta_m} := \tau\log(\mathrm{e}c_1^\tau(1-\theta_m^{-1})/\tau)$*.*

- *When* W *has an atom at* x_0*, i.e.* $q_0 = \mathbb{P}(W = x_0) > 0$*, set*

$$\begin{aligned} s_k &:= \inf\{x \in (0,1) : \exp\{-(1-\theta_m^{-1})(1-x)k\} \leq \mathbb{P}(W \in (x,1))\},\\ r_k &:= \exp\{-(1-\theta_m^{-1})(1-s_k)k\}. \end{aligned} \tag{2.10}$$

  Then,

$$p_k = q_0(1-\theta_m^{-1})\theta_m^{-k}(1+\mathcal{O}(r_k)). \tag{2.11}$$

(**Gumbel**) (i) *If* W *satisfies the* (**RV**) *sub-case with parameter* τ*, then for* $\gamma := 1/(\tau+1)$*,*

$$p_k = \exp\Big\{ - \frac{\tau^\gamma}{1-\gamma}\Big(\frac{k}{c_1 m}\Big)^{1-\gamma}(1+o(1))\Big\}. \tag{2.12}$$

(ii) *If* W *satisfies the* (**RaV**) *sub-case with parameter* $\tau > 1$*,*

$$p_k = \frac{1}{k}\exp\Big\{-\Big(\frac{\log(k/m)}{c_1}\Big)^\tau\Big(1-\tau(\tau-1)\frac{\log\log(k/m)}{\log(k/m)}+\frac{K_{\tau,c_1}}{\log(k/m)}(1+o(1))\Big)\Big\}, \tag{2.13}$$

where $K_{\tau,c_1} := \tau\log(\mathrm{e}c_1^\tau/\tau)$*.*

(**Fréchet**) *When* $\alpha > 2$,

$$\underline{\ell}(k)k^{-\alpha} \leq p_k \leq \overline{\ell}(k)k^{-\alpha}, \tag{2.14}$$

where $\overline{\ell}, \underline{\ell}$ *are slowly varying at infinity.*

Remark 2.8. We observe that the (**Gumbel**)-(**SV**) sub-case is not covered in Theorem 2.7. This is due to the fact that this case only specifies the behaviour of the first order asymptotic growth of the maximum of n vertex-weights, b_n, and provides no details of the underlying distribution of the vertex-weights. As a result, we are not able to obtain a precise asymptotic expression for p_k, though precise statements about its maximum degree can still be observed, as follows later in this section.

The asymptotic behaviour of the degree distribution p_k in Theorem 2.7 allows for a non-rigorous estimation of the size of the maximum degree in $\mathcal{G}_n$. As is the case for the degree distribution, the behaviour of the maximum degree in the WRG model is highly dependent on the underlying distribution of the vertex-weights as well, and on a heuristic level one would expect the size of the maximum degree, say d_n, to be such that $\sum_{k\geq d_n} p_k \approx 1/n$. The following theorem makes this heuristic statement precise and states the first-order growth rate of the maximum degree for three different classes of vertex-weight distributions. In all classes, we find, up to the leading order in the asymptotic expression of p_k, that $\sum_{k\geq d_n} p_k \approx 1/n$ is satisfied when considering the asymptotic expressions in Theorem 2.7.

Theorem 2.9 (Maximum degree in WRGs)**.** *Consider the WRG model as in Definition 2.1 and let* $I_n := \inf\{i \in [n] : \mathcal{Z}_n(i) \geq \mathcal{Z}_n(j) \text{ for all } j \in [n]\}$. *We consider the different cases of Assumption 2.3.*

(**Bounded**) *Let* $\theta_m := 1 + \mathbb{E}[W]/m$. *Then,*

$$\frac{\max_{i\in[n]} \mathcal{Z}_n(i)}{\log n} \xrightarrow{a.s.} \frac{1}{\log \theta_m}.$$

(**Gumbel**) *For sub-case* (**SV**),

$$\Big(\max_{i\in[n]} \frac{\mathcal{Z}_n(i)}{mb_n \log n}, \frac{\log I_n}{\log n}\Big) \xrightarrow{\mathbb{P}} (1,0). \tag{2.15}$$

For sub-case (**RV**), *let* $\gamma := 1/(\tau+1)$. *Then,*

$$\Big(\max_{i\in[n]} \frac{\mathcal{Z}_n(i)}{m(1-\gamma)b_{n^\gamma} \log n}, \frac{\log I_n}{\log n}\Big) \xrightarrow{a.s.} (1,\gamma). \tag{2.16}$$

Finally, for sub-case (**RaV**), *let* $t_n := \exp\{-\tau \log n/ \log(b_n)\}$. *Then,*

$$\Big(\max_{i\in[n]} \frac{\mathcal{Z}_n(i)}{mb_{t_n n} \log(1/t_n)}, \frac{\log I_n}{\log n}\Big) \xrightarrow{\mathbb{P}} (1,1). \tag{2.17}$$

(**Fréchet**) *Let* Π *be a Poisson point process (PPP) on* $(0,1)\times(0,\infty)$ *with intensity measure* $\nu(\mathrm{d}t, \mathrm{d}x) := \mathrm{d}t \times (\alpha-1)x^{-\alpha}\mathrm{d}x$. *Then, when* $\alpha > 2$,

$$(\max_{i\in[n]} \mathcal{Z}_n(i)/u_n, I_n/n) \xrightarrow{d} (m \max_{(t,f)\in\Pi} f \log(1/t), I_\alpha), \tag{2.18}$$

where $m \max_{(t,f)\in\Pi} f\log(1/t)$ *and* I_α *are independent, with* $I_\alpha \overset{d}{=} \mathrm{e}^{-W_\alpha}$ *and* W_α *a* $\Gamma(\alpha,1)$ *random variable, and where* $m \max_{(t,f)\in\Pi} f \log(1/t)$ *has a Fréchet distribution with shape parameter* $\alpha - 1$ *and scale parameter* $m\Gamma(\alpha)^{1/(\alpha-1)}$. *Finally, when* $\alpha \in (1,2)$,

$$(\max_{i\in[n]} \mathcal{Z}_n(i)/n, I_n/n) \xrightarrow{d} (Z, I), \tag{2.19}$$

for some random variable I *with values in* $(0,1)$ *and where*

$$Z = m \max_{(t,f)\in\Pi} f \int_t^1 \Big(\int_{(0,1)\times(0,\infty)} g\mathbb{1}_{\{u\leq s\}}\,\mathrm{d}\Pi(u,g)\Big)^{-1} \mathrm{d}s.$$

Remark 2.10. (i) Note that the asymptotics of the maximal degrees are the result of a non-trivial competition, where older vertices can achieve a higher degree because they have been in the system for longer, while younger vertices have the chance to have a big vertex-weight corresponding to a local maximum.

(ii) The result in (2.19) is equivalent to the behaviour of the PAF models with infinite mean power-law fitness random variables, as presented in [19]. Here, the influence of the fitness (vertex-weights) overpowers the preferential attachment mechanism so that the preferential attachment graph behaves like a weighted recursive graph.

(iii) The result in (2.18) can actually be extended to hold jointly for the K largest degrees and their locations as well, for any $K \in \mathbb{N}$. The limits $(Z^{(K)}, I_\alpha^{(K)})$ of the K^{th} largest degree and its location are independent, $I_\alpha^{(K)} \overset{d}{=} \mathrm{e}^{-W_\alpha^{(K)}}$, where the $(W_\alpha^{(K)})_{K\in\mathbb{N}}$ are i.i.d. $\Gamma(\alpha, 1)$ random variables, and

$$\mathbb{P}\Big(Z^{(K)} \leq x\Big) = \sum_{i=0}^{K-1} \frac{1}{i!}(\Gamma(\alpha)(x/m)^{-(\alpha-1)})^i \exp\{-\Gamma(\alpha)(x/m)^{-(\alpha-1)}\}.$$

(iv) We conjecture that the convergence in (2.15) can be strengthened to almost sure convergence. This is definitely the case for particular vertex-weight distributions that satisfy the **(Gumbel)**-**(SV)** sub-case, e.g. $W := \log W'$, where W' satisfies the **(Gumbel)**-**(RV)** sub-case.

(v) In (2.16) and (2.17) the rescaling of the maximum degree can be interpreted as $m(1-\gamma)c_1\gamma^{1/\tau}(\log n)^{1+1/\tau}$ and $me^{-1}c_2^{-1}(\log n)^{1-1/\tau}\exp\big\{c_1(\log n)^{1/\tau}\big\}$, respectively, since the lower order terms of b_n as in (2.1) and (2.2) can be ignored when considering only the first-order behaviour of the maximum degree. As a result, it should be possible to weaken the assumptions on the tail-distribution in the **(Gumbel)**-**(RV)** and **(Gumbel)**-**(RaV)** sub-cases to $\mathbb{P}(W \geq x) = \exp\big\{-(x/c_1)^\tau(1+o(1))\big\}$ and $\mathbb{P}(W \geq x) = \exp\big\{-(\log(x)/c_1)^\tau(1+o(1))\big\}$, respectively, such that the results in (2.16) and (2.17) still hold.

A result we have been unable to prove, but which we conjecture to be true, is the convergence of the location of the maximum degree in the WRG when the vertex-weights are almost surely bounded, which would improve the result proved for the random recursive tree by Banerjee and Bhamidi in [3] from convergence in probability to almost sure convergence and would extend this result to the m-DAG model and the WRG model.

Conjecture 2.11 (Location of the maximum degree in WRGs with bounded weights)**.** *Consider the WRG model as in Definition 2.1, let $I_n := \inf\{i \in [n] : \mathcal{Z}_n(i) \geq \mathcal{Z}_n(j)$ for all $j \in [n]\}$ and set $\theta_m := 1 + \mathbb{E}[W]/m$. When the vertex-weights satisfy the* **(Bounded)** *case,*

$$\frac{\log I_n}{\log n} \xrightarrow{a.s.} 1 - \frac{\theta_m - 1}{\theta_m \log(\theta_m)}.$$

We now concentrate on the case of unbounded weights. In Theorem 2.9, we note that in the **(Fréchet)** case the first order of the growth of the largest degree is random, while in the **(Gumbel)** case, the first order is deterministic. This is a general feature which also appears when considering the growth of the maximum of n i.i.d. weights (although the particular normalizations and limits are different). For the i.i.d. case it is also known that in the **(Gumbel)** case the second order is random. A natural conjecture would be that the same is true in for this model. Our next result shows that if we consider the indices in a compact window around the maximal one (as identified in Theorem 2.9), then this is indeed true.

To formulate our results, we introduce the following notation: For $0 < s < t < \infty$, $\gamma \in (0,1)$ and a strictly positive function f, define

$$\begin{aligned} C_n(\gamma, s, t, f) &:= \{i \in [n] : sf(n)n^\gamma \leq i \leq tf(n)n^\gamma\}, \\ I_n(\gamma, s, t, f) &:= \inf\{i \in C_n(\gamma, s, t, f) : \mathcal{Z}_n(i) \geq \mathcal{Z}_n(j) \text{ for all } j \in C_n(\gamma, s, t, f)\}. \end{aligned} \tag{2.20}$$

We abuse notation to also write $C_n(1,s,t,t_n)$ and $I_n(1,s,t,t_n)$ when we deal with vertices i such that $st_n n \le i \le tt_n n$ for some sequence $(t_n)_{n\in\mathbb{N}}$. We then present the following theorem:

Theorem 2.12 (Random second order asymptotics in the Gumbel case)**.** *In the same setting as in Theorem 2.9, we further assume that the vertex-weights fall into the sub-case* (**RV**)*. Let $\gamma := 1/(\tau+1)$ and let ℓ be a strictly positive function such that $\lim_{n\to\infty}\log(\ell(n))^2/\log n = \zeta_0$ for some $\zeta_0 \in [0,\infty)$. Furthermore, let Π be a Poisson point process (PPP) on $(0,\infty)\times\mathbb{R}$ with intensity measure $\nu(\mathrm{d}t,\mathrm{d}x) := \mathrm{d}t \times \mathrm{e}^{-x}\mathrm{d}x$. Then, when $\tau \in (0,1)$,*

$$\Big(\max_{i\in C_n(\gamma,s,t,\ell)} \frac{\mathcal{Z}_n(i) - m(1-\gamma)b_{n^\gamma}\log n}{m(1-\gamma)a_{n^\gamma}\log n}, \frac{I_n(\gamma,s,t,\ell)}{\ell(n)n^\gamma}\Big) \xrightarrow{d} \Big(\max_{\substack{(v,w)\in\Pi \\ v\in[s,t]}} w - \log v - \frac{\zeta_0(\tau+1)^2}{2\tau}, \mathrm{e}^U\Big), \tag{2.21}$$

where $U \sim \mathrm{Unif}(\log s, \log t)$ and the maximum over the PPP follows a Gumbel distribution with location parameter $\log\log(t/s) - \zeta_0(\tau+1)^2/2\tau$.

Finally, let us assume that the vertex-weights fall into the sub-class (**RaV**) *and let $t_n := \exp\{-\tau\log n/\log(b_n)\}$. Then, for any $0 < s < t < \infty$ and with Π and U as above,*

$$\Big(\max_{i\in C_n(1,s,t,t_n)} \frac{\mathcal{Z}_n(i) - mb_{t_n n}\log(1/t_n)}{ma_{t_n n}\log(1/t_n)}, \frac{I_n(1,s,t,t_n)}{t_n n}\Big) \xrightarrow{d} \Big(\max_{\substack{(v,w)\in\Pi \\ v\in[s,t]}} w - \log v, \mathrm{e}^U\Big), \tag{2.22}$$

where now the maximum follows a Gumbel distribution with location parameter $\log\log(t/s)$.

Remark 2.13 (The vertex with largest degree for (**Gumbel**) weights)**.** The restriction to $\tau \in (0,1)$ comes from the fact our result only looks at the fluctuations coming from the random weights and indeed the same statement is true for all $\tau > 0$ when looking at the conditional expected degrees (conditioned on the random weights), see Proposition 5.4 later on. By a central limit theorem-type argument we would expect that the fluctuations of the degree around its conditional mean are of the order square root of its variance (which is comparable to its mean and so of order $(\log n)^{(1/\tau+1)/2}$), therefore if $\tau > 1$ this term would be larger than the fluctuations coming from the random weights (which are of order $(\log n)^{1/\tau}$) and so we would expect a different scaling limit.

A standard Poisson process calculation (for more details see Section 3) shows that the limit random variables describing the second order growth of the near-maximal degree in Theorem 2.12 become infinite if we let $s \downarrow 0$ and $t \to \infty$. This phenomenon indicates that we need to consider a much larger window of indices to capture the true second order asymptotics of the maximal degree over the full set of indices. This fact also shows that the competition between the advantages of older vertices compared to vertices with high weight is very finely balanced. The following result captures the resulting effect on the second order asymptotics.

Theorem 2.14 (Precise second order asymptotics in the Gumbel case)**.** *In the same setting as in Theorem 2.9, we first assume that the vertex-weights fall into the sub-case* (**RV**) *and let $\gamma := 1/(\tau+1)$. For $\tau \in (0,1]$,*

$$\max_{i\in[n]} \frac{\mathcal{Z}_n(i) - m(1-\gamma)b_{n^\gamma}\log n}{m(1-\gamma)a_{n^\gamma}\log n\log\log n} \xrightarrow{\mathbb{P}} \frac{1}{2}. \tag{2.23}$$

Now assume that the vertex-weights fall into the sub-class (**RaV**) *and let $t_n := \exp\{-\tau\log n/\log(b_n)\}$. If $\tau \in (1,3]$,*

$$\max_{i\in[n]} \frac{\mathcal{Z}_n(i) - mb_{t_n n}\log(1/t_n)}{ma_{t_n n}\log(1/t_n)\log\log n} \xrightarrow{\mathbb{P}} \frac{1}{2}\Big(1 - \frac{1}{\tau}\Big), \tag{2.24}$$

whilst for $\tau > 3$,

$$\max_{i\in[n]} \frac{\mathcal{Z}_n(i) - mb_{t_n n}\log(1/t_n)}{ma_{t_n n}\log(1/t_n)(\log n)^{1-3/\tau}} \xrightarrow{\mathbb{P}} -\frac{\tau(\tau-1)^2}{2c_1^3}.$$

Remark 2.15. Though the result in (2.21) only holds for $\tau \in (0,1)$, the result in (2.23) turns out to hold for $\tau = 1$ as well. This slight deviation is due to the fact that the additional $\log\log n$ term allows us to prove concentration of the maximum degree around the maximum conditional mean degree, which cannot be done with the second order rescaling in (2.21) when $\tau = 1$.

As mentioned above, the problem of capturing the second order fluctuations in the (**Gumbel**)-(**RV**) case when $\tau > 1$ and for lighter tailed weights (including bounded weights) requires different techniques and is currently on-going research.

Remark 2.16 (More general model formulation)**.** As in [19], it is possible to prove some of the results for a more general class of models. More specifically, the results in Theorem 2.5 and the (**Fréchet**) case in Theorem 2.9 hold for a growing network that satisfies the following conditions as well: let $\Delta\mathcal{Z}_n(i) := \mathcal{Z}_{n+1}(i) - \mathcal{Z}_n(i)$. For all $n \in \mathbb{N}$:

(A1) $\mathbb{E}_W[\Delta\mathcal{Z}_n(i)] = W_i/S_n\mathbb{1}_{\{i\in[n]\}}$.
(A2) For all $k \in \mathbb{N}$, $\exists\, C_k > 0$ such that $\mathbb{E}_W\big[\prod_{j=0}^{k-1}(\Delta\mathcal{Z}_n(i) - j)\big] \le C_k\mathbb{E}_W[\Delta\mathcal{Z}_n(i)]$.
(A3) $\sup_{i=1,\dots,n} n\big|\mathbb{P}_W(\Delta\mathcal{Z}_n(i) = 1) - \mathbb{E}_W[\Delta\mathcal{Z}_n(i)]\big| \xrightarrow{a.s.} 0$.
(A4) Conditionally on $(W_i)_{i\in\mathbb{N}}$, $\{\Delta\mathcal{Z}_n(i)\}_{i\in[n]}$ is negatively quadrant dependent in the sense that for any $i \neq j$ and $k, l \in \mathbb{Z}^+$,

$$\mathbb{P}_W(\Delta\mathcal{Z}_n(i) \le k, \Delta\mathcal{Z}_n(j) \le l) \le \mathbb{P}_W(\Delta\mathcal{Z}_n(i) \le k)\,\mathbb{P}_W(\Delta\mathcal{Z}_n(j) \le l)\,.$$

If we further assume that $\Delta\mathcal{Z}_n(i) \in \{0,1\}$ then all the results presented in this paper hold as well.

Outline of the paper
In Section 3 we provide a short overview and explain the intuitive idea of the proofs of Theorem 2.5, 2.7, 2.9, 2.12 and 2.14. We then prove Theorem 2.5 and 2.7 in Section 4. In Section 5, we state and prove several propositions regarding the maximum conditional mean degree. We then discuss under which scaling the maximum degree concentrates around the maximum conditional expected degree in Section 6. Finally, we use these results in Section 7 to prove the main theorems, Theorem 2.9, 2.12 and 2.14. For clarity, we split the proof of Theorem 2.9 into three separate parts that deal with each of the cases outlined in the theorem separately.

3. Overview of the proofs

First, since the proof of Theorem 2.5 heavily relies on the proof of Theorem 2.4 in [19], we refer to [19, Section 3] for an overview of its proof. The same holds for Theorem 2.9, the (**Bounded**) case, which follows the same strategy as the proof of Theorem 2 in [8] but where we need to take extra care because of the random weights. Finally, the proof of Theorem 2.7 is mainly computational in nature and we therefore do not include an overview in this section.

Here, we provide an intuitive idea of the proof of Theorem 2.9, for the (**Gumbel**) and (**Fréchet**) cases, as well as Theorem 2.12 and 2.14. In the (**Gumbel**) and (**Fréchet**) cases of Theorem 2.9, the main idea consists of two ingredients: We first consider the asymptotics of the conditional expected degree $\mathbb{E}_W[\mathcal{Z}_n(i)]$ of a vertex $i \in [n]$, where we condition on the weights $(W_i)_{i\in[n]}$. Then in a second step, we show that the degrees concentrate around their conditional expected values.

More precisely for the concentration argument, we show that

$$|\max_{i\in[n]} \mathcal{Z}_n(i) - \max_{i\in[n]} \mathbb{E}_W[\mathcal{Z}_n(i)]|/g_n \xrightarrow{\mathbb{P}} 0, \tag{3.1}$$

for some suitable sequence $(g_n)_{n\in\mathbb{N}}$ such that g_n diverges with n. What sequence g_n is sufficient depends on the different vertex-weight distribution cases, as outlined in Assumption 2.3. For completeness, $g_n = mb_n \log n, g_n = m(1-\gamma)b_{n^\gamma} \log n, g_n = mb_{t_n n}\log(1/t_n)$, $g_n = u_n$ and $g_n = n$ for the **(Gumbel)**-**(SV)**, **(Gumbel)**-**(RV)**, **(Gumbel)**-**(RaV)** sub-cases and the **(Fréchet)** case with $\alpha > 2$ and $\alpha \in (1,2)$, respectively. (3.1) follows by applying standard large deviation bounds to $|\mathcal{Z}_n(i) - \mathbb{E}_W[\mathcal{Z}_n(i)]|$ for all $i \in [n]$, as in Proposition 6.1. We can also construct a concentration argument when g_n is equivalent to the second order growth rate of the maximum, however in this case a more careful analysis of the different terms in the large deviation bounds is required.

The bulk of the argument for our results is to show that the conditional expected degree behave as we claim above. For the first order asymptotics as in Theorem 2.9, we have in the **(Gumbel)** case,

$$\max_{i\in[n]} \frac{\mathbb{E}_W[\mathcal{Z}_n(i)]}{g_n} \xrightarrow{\mathbb{P}} 1,$$

with g_n as described above for the **(SV)**, **(RV)** and **(RaV)** sub-cases, whereas for the **(Fréchet)** case,

$$\max_{i\in[n]} \frac{\mathbb{E}_W[\mathcal{Z}_n(i)]}{u_n} \xrightarrow{d} m \max_{(t,f)\in\Pi} f\log(1/t),$$

$$\max_{i\in[n]} \frac{\mathbb{E}_W[\mathcal{Z}_n(i)]}{n} \xrightarrow{d} m \max_{(t,f)\in\Pi} f \int_t^1 \bigg(\int_{(0,1)\times(0,\infty)} g\mathbb{1}_{\{u\leq s\}} \mathrm{d}\,\Pi(u,g)\bigg)^{-1} \mathrm{d}s,$$

depending on whether $\alpha > 2$ or $\alpha \in (0,1)$. Together with (3.1), the results in Theorem 2.9 for the **(Gumbel)** and **(Fréchet)** cases follow. For the **(Gumbel)**-**(RV)** sub-case (and probably the **(Gumbel)**-**(SV)** sub-case, too), the result can be strengthened to almost sure convergence.

Let us delve a bit more into why the maximum conditional mean in-degree has the limits as claimed above. We stress that, as stated in Remark 2.2, $\mathbb{E}[W] = 1$ for the cases we discuss here. It is clear from the definition of the model that

$$\mathbb{E}_W[\mathcal{Z}_n(i)] = mW_i \sum_{j=i}^{n-1} 1/S_j \approx mW_i \log(n/i),$$

for any $i \in [n]$. Let us start with the **(Gumbel)**-**(RV)** sub-case, where we recall that $\gamma := 1/(1+\tau)$. If we set

$$\Pi_n := \sum_{i\geq 1} \delta_{(i/n,(W_i-b_n)/a_n)},$$

where δ is a Dirac measure, then classical extreme value theory tells us that (see e.g. [25])

$$\Pi_n \Rightarrow \Pi, \tag{3.2}$$

where Π is a Poisson point process on $(0,\infty)\times\mathbb{R}$ with intensity measure $\nu(\mathrm{d}t, \mathrm{d}x) := \mathrm{d}t \times \mathrm{e}^{-x}\mathrm{d}x$. Then, if we consider $i = t\ell(n)n^\gamma$ and $(W_i - b_{\ell(n)n^\gamma})/a_{\ell(n)n^\gamma} = f$ where ℓ is a strictly positive function such that $\log(\ell(n))^2/\log n$ converges, it follows that

$$\begin{aligned}\mathbb{E}_W[\mathcal{Z}_n(i)] &\approx mW_i\log(n/i) = m(b_{\ell(n)n^\gamma} + a_{\ell(n)n^\gamma} f)\log(1/(tn^{1-\gamma}\ell(n)))\\ &\approx mc_1\gamma^{1/\tau}(1-\gamma)(\log n)^{1/\tau+1} + mc_2\gamma^{1/\tau-1}(f(1-\gamma) - \tau\gamma\log t)(\log n)^{1/\tau}\\ &\approx m(1-\gamma)b_{n^\gamma}\log n + m(1-\gamma)a_{n^\gamma}\log n(f - \log t),\end{aligned} \tag{3.3}$$

when using that $a_n = c_2(\log n)^{1/\tau-1}$ and $b_n \sim c_1(\log n)^{1/\tau}$, using a Taylor approximation and leaving out all lower order terms. This yields the first order behaviour of the maximum conditional mean in-degree as well as its location $n^{\gamma(1+o(1))}$. This is proved rigorously in Proposition 5.4.

For the **(SV)** sub-case, as in Proposition 5.2, a similar approach as in (3.3) can be applied, though we set $i = tn^\beta$, $(W_i - b_{n^\beta})/a_{n^\beta} = f$ for any $\beta \in (0,1)$. We divide the right-hand side of (3.3) by $b_n \log n$ and observe that $b_{n^\beta} = \ell(\log(n^\beta)) = \ell(\beta \log n)$, so that it follows that b_{n^β}/b_n converges to 1 with n for any fixed $\beta \in (0,1)$ since ℓ is slowly varying at infinity. Then, the constant in front of the leading term is increasing in β, so that taking the limit $\beta \downarrow 0$ yields the required result.

Then, for the **(RaV)** sub-case, as in Proposition 5.5, we realise that the location of the maximum should grow faster than n^γ for any $\gamma \in (0,1)$, as the tails of these distribution are heavier than those of any distribution in the **(RV)** sub-case. By a similar argument as for the **(SV)** sub-case and using that now b_{n^β}/b_n converges to 0 with n, one might want to set $\beta = 1$, that is, the location of the maximum degree is of order n. However, this would imply that the growth rate of the maximum expected degree should be b_n. This is not the case, however, since for any $t \in (0,1)$, approximately,

$$\max_{i\in[n]} \mathbb{E}_W[\mathcal{Z}_n(i)]/b_n \geq (\max_{i\in[tn]} W_i/b_{tn}) \log(1/t) b_{tn}/b_n.$$

Since b_{tn}/b_n converges to 1 for t fixed (b_n is slowly varying) and the maximum converges to 1 in probability, letting t tend to 0 shows scaling by just b_n is insufficient. To find the correct behaviour, we need to let t tend to zero with n, i.e. $t = t_n$, such that $b_{t_n n}/b_n$ has a non-trivial limit (not 0 or 1). The sequence that satisfies this requirement is $t_n = \exp\{-\tau \log n/\log(b_n)\}$. This suggests that the location of the maximum degree is $t_n n$ and that the maximum degree grows as $b_{t_n n} \log(1/t_n)$.

The second order growth rate for the **(RaV)** sub-case, as in Theorem 2.12, is obtained in a similar way as in (3.3), where we now consider $i = st_n n$ and $(W_i - b_{t_n n})/a_{t_n n} = f$ for some $(s,f) \in (0,\infty)\times\mathbb{R}$. This yields

$$\begin{aligned}\mathbb{E}_W[\mathcal{Z}_n(i)] &\approx mW_i \log(n/i) = m(b_{t_n n} + a_{t_n n} f)\log(1/st_n)\\ &= mb_{t_n n}\log(1/t_n) + (ma_{t_n n}\log(1/t_n)f - mb_{t_n n}\log s) + ma_{t_n n} f \log(1/s).\end{aligned}$$

Here, the first order again appears in the first term on the right-hand side, and the second order can be obtained by realising that $b_{t_n n}/(a_{t_n n}\log(1/t_n)) \to 1$ as n tends to infinity. A similar approach using the weak convergence of Π_n to Π as in (3.2) allows us to obtain the required limits. For the results of the **(RV)** sub-case in Theorem 2.12, we take a more in-depth look at the approximation in (3.3). First, when subtracting the first term on the right-hand side and dividing by $m(1-\gamma)a_{n^\gamma}\log n$, we are left with exactly the functional which is used in the maximum over the Poisson point process in (2.21). When combining this with the weak convergence of Π_n to Π the desired result follows. To understand how the additional term in (2.21) arises, we include an extra lower order term in the approximation in (3.3). That is,

$$\begin{aligned}\mathbb{E}_W[\mathcal{Z}_n(i)] \approx{}& mc_1\gamma^{1/\tau}(1-\gamma)(\log n)^{1/\tau+1} + mc_2\gamma^{1/\tau-1}(f(1-\gamma) - \tau\gamma\log t)(\log n)^{1/\tau}\\ &- mc_1\frac{\gamma^{1/\tau}}{1-\gamma}\frac{1+\tau}{2}(\log n)^{1/\tau-1}\log(\ell(n))^2.\end{aligned}$$

Hence, the requirement that $\lim_{n\to\infty}\log(\ell(n))^2/\log n = \zeta_0$ ensures that the last term on the right-hand side does not grow faster than $(\log n)^{1/\tau}$. Also, when divided by the second order growth rate $m(1-\gamma)a_{n^\gamma}\log n$, the last term converges to $-(1+\tau)^2\zeta_0/(2\tau)$, exactly the additional term found in (2.21).

Somewhat surprising is that for both the **(RV)** and **(RaV)** sub-cases, when we consider not a compact window around the optimal index but instead all $i \in [n]$, we find that the second order correction as suggested above is insufficient, as can be observed in (2.23) and (2.24) in Theorem 2.14. The reason this behaviour is observed, loosely speaking, is that we can move away even further from what one would expect to be the optimal window, i.e. $\ell(n)n^\gamma$

for ℓ that do not grow or decay 'too quickly' in the **(RV)** sub-case and $t_n n$ in the **(RaV)** sub-case, and find higher degrees. As we have observed above, when setting $s = 0, t = \infty$ in the limit in (2.21), which would mimic considering all $i \in [n]$ rather than the indices in a compact window around n^γ, we obtain $\sup_{(v,w)\in\Pi} w - \log v$ (when we would set $\zeta_0 = 0$). It is readily checked that integrating the intensity measure ν over $\mathbb{R}_+ \times \mathbb{R}$ yields an infinitely large rate. Hence, $\sup_{(v,w)\in\Pi} w - \log v = \infty$ almost surely, so that the second order scaling $(1-\gamma)a_{n^\gamma} \log n$ is insufficient when considering all $i \in [n]$.

A more insightful argument is the following: consider the PPP limit as in (2.21) (with $\zeta_0 = 0$) and (2.22). Its distribution depends only on the ratio t/s. Thus, for any integer $j \in \mathbb{N}$ such that $j = o(\sqrt{\log n})$,

$$\max_{\mathrm{e}^{j-1}n^\gamma \le i \le \mathrm{e}^j n^\gamma} \frac{\mathcal{Z}_n(i) - m(1-\gamma)b_{n^\gamma}\log n}{m(1-\gamma)a_{n^\gamma}\log n} \xrightarrow{d} \Lambda_j,$$

where the $(\Lambda_j)_{j\in\mathbb{N}}$ are i.i.d. standard Gumbel random variables (where the location parameter equals 0). Their independence follows from the independence property of the PPP. Now, as the Gumbel distribution satisfies the **(Gumbel)**-**(RV)** sub-case with $\tau = 1, c_1 = 1, b = 0$, we can argue that, for any $\eta > 0$ and $x \in \mathbb{R}$,

$$\begin{aligned}
&\mathbb{P}\bigg(\max_{n^\gamma \le i \le \mathrm{e}^{(\log n)^{1/2-\eta}} n^\gamma} \frac{\mathcal{Z}_n(i) - m(1-\gamma)b_{n^\gamma}\log n}{m(1-\gamma)a_{n^\gamma}\log n} \le x(1/2-\eta)\log\log n\bigg) \\
&\quad = \mathbb{P}\bigg(\max_{1\le j\le (\log n)^{1/2-\eta}} \max_{\mathrm{e}^{j-1}n^\gamma \le i \le \mathrm{e}^j n^\gamma} \frac{\mathcal{Z}_n(i) - m(1-\gamma)b_{n^\gamma}\log n}{m(1-\gamma)a_{n^\gamma}\log n} \le x(1/2-\eta)\log\log n\bigg) \\
&\quad \approx \mathbb{P}\bigg(\max_{1\le j\le (\log n)^{1/2-\eta}} \Lambda_j \le x(1/2-\eta)\log\log n\bigg) \\
&\quad = \mathbb{P}(\Lambda_1 \le x(1/2-\eta)\log\log n)^{(\log n)^{1/2-\eta}} \\
&\quad = \exp\{-(\log n)^{1/2-\eta}\exp\{-x(1/2-\eta)\log\log n\}\},
\end{aligned}$$

which has limit 1 (resp. 0) if $x > 1$ (resp. $x < 1$). Then, as we can choose η arbitrarily close to 0, the result in (2.23) follows. Here it is essential that $j = o(\sqrt{\log n})$ to obtain the correct limit. Note that making the approximation $\approx$ in the above argument rigorous is the highly non-trivial part of the argument. The reason is that we cannot rely on the elegant theory of convergence to a Poisson point processes, but have to explicitly control the errors made in this approximation.

A similar reasoning can be applied for the **(Gumbel)**-**(RaV)** when $\tau \in (1,3]$, where now $j = o((\log n)^{(1-1/\tau)/2})$ needs to be satisfied. When $\tau > 3$, more care needs to be taken of lower-order terms that appear in the double exponent, yielding a different scaling.

In the **(Fréchet)** case, we now set

$$\Pi_n := \sum_{i=1}^n \delta_{(i/n, W_i/u_n)},$$

where again $\Pi_n \Rightarrow \Pi$, with Π a Poisson point process on $(0,1)\times(0,\infty)$ with intensity measure $\nu(\mathrm{d}t,\mathrm{d}x) := \mathrm{d}t \times (\alpha-1)x^{-\alpha}\mathrm{d}x$. Now considering $i = tn, W_i = fu_n$, yields

$$\mathbb{E}_W[\mathcal{Z}_n(i)/u_n] \approx m\frac{W_i}{u_n}\log(n/i) = mf\log(1/t),$$

which is the functional in the maximum in (2.18). Again, combining this with the weak convergence of Π_n yields the result. Finally, the heuristic idea for (2.19) is contained in [19, Section 3] as well, since the rescaled maximum degree in the preferential attachment model with additive fitness studied there has the same distributional limit when $\alpha \in (1,2)$.

4. The limiting degree sequence of weighted recursive graphs

In this section we prove Theorem 2.5 and 2.7. The proof of Theorem 2.5 follows the same steps as the proof of [19, Theorem 2.4] and we simply give an overview of the steps that need to be adjusted.

Proof of Theorem 2.5. First, we can, without loss of generality, assume that $\mathbb{E}[W] = 1$ when $\mathbb{E}[W] < \infty$ is satisfied. Let us start by discussing why the results hold for the model with a random out-degree, as in Remark 2.2(*ii*), as well. In [19, Theorem 2.4], the results also follow for a model with a random out-degree, the PAFRO model, and the adjustments required for the proof to work for this model are all made in [19, Lemma 4.2]. These adjustments account for the fact that the expected total in-degree in $\mathcal{G}_n$, i.e. $\sum_{i=1}^n \mathbb{E}[\mathcal{Z}_n(i)]$, can be larger than $m_0 + m(n-n_0)$ due to the random out-degree of each vertex, so that

$$\sum_{i=1}^n \mathbb{E}\left[\frac{\mathcal{Z}_n(i)+\mathcal{F}_i}{m_0+m(n-n_0)+S_n}\right] > 1$$

is possible, where $\mathcal{F}_i$ is the fitness of vertex i. Hence, adjustments are required to show that the sum of the expected connection probabilities does not exceed 1 'too much'. For the WRG model, however, even with a random out-degree, we still have that

$$\sum_{i=1}^n \mathbb{E}[W_i/S_n] = 1,$$

as the in-degree does not play a role in the connection probabilities. Hence, the equivalence of [19, Lemma 4.2] immediately holds for a model with such a random out-degree as well, from which the entire proof follows analogously.

To adapt the proof in [19] to the WRG model, we set $\bar{\mathcal{F}}_n = S_n/n$ in this model, which by the strong law of large numbers converges to 1 almost surely by our assumption that $\mathbb{E}[W] = 1$. As the vertex-weights are non-negative, we let $0 < f < f' < \infty$ and $\mathbb{F} = [0, \infty)$. Set $X_n := (1/n)\sum_{i\in\mathbb{I}_n} \mathcal{Z}_n(i)$ and $\mathbb{I}_n := \{i \in [n] | W_i \in (f, f']\}$. Now, following the same steps, we arrive at the upper and lower bound

$$\begin{aligned}X_{n+1} - X_n &\geq \frac{1}{n+1}\Big(-X_n + \frac{\mathbb{I}_n}{n}\frac{mf}{S_n/n}\Big) + \Delta R_n,\\ X_{n+1} - X_n &\leq \frac{1}{n+1}\Big(-X_n + \frac{\mathbb{I}_n}{n}\frac{mf'}{S_n/n}\Big) + \Delta R_n.\end{aligned}$$

Using the law of large numbers and [7, Lemma 3.1], this results in the upper and lower bound,

$$\liminf_{n\to\infty} X_n \geq mf\mu((f,f']), \qquad \limsup_{n\to\infty} X_n \leq mf'\mu((f,f']),$$

almost surely. The almost sure convergence of R_n follows from [19, Lemma 4.2], which proves the almost sure convergence of Γ_n in the weak* topology to Γ with a similar argument as in [19]. In the remainder of the proof, we let $X_n := \Gamma_n^{(k)}((f,f']) = (1/n)\sum_{i\in\mathbb{I}_n} \mathbb{1}_{\{\mathcal{Z}_n(i)=k\}}$. Again, following the same steps as in [19], replacing the terms $(k+\mathcal{F}_i)/(n\bar{\mathcal{F}}_n/m), (k+f')/(\bar{\mathcal{F}}_n/m)$ and $(f'-\mathcal{F}_i)/(n\bar{\mathcal{F}}_n/m)$ in (4.12) by $mW_i/S_n, mf'/(S_n/n), m(f'-W_i)/S_n$, respectively, it follows that we obtain the lower bound

$$X_{n+1} - X_n \geq \frac{1}{n+1}(A_n - B_n' X_n) + \Delta R_n,$$

where A_n, B_n' almost surely converge to

$$A := m\int_{(f,f']} x\,\Gamma^{(k-1)}(\mathrm{d}x), \qquad B' := \frac{1/m+f'}{1/m},$$

respectively, and where the almost sure convergence of R_n again follows from [19, Lemma 4.2]. For the proof of the convergence to these limits, in the arguments in the proof of

Theorem 2.4 in [19], (4.14) through (4.18), change $(k-1+x)$ to x and $(k+\mathcal{F}_i)/(n\bar{\mathcal{F}}_n/m)$ to mW_i/S_n. With a similar approach, an upper bound on the recursion $X_{n+1}-X_n$ can be obtained with sequences A_n, B_n that converge to A and B, respectively, with $B = 1+mf$. Now, applying [7, Lemma 3.1] yields

$$\begin{aligned}\liminf_{n\to\infty} X_n &\geq \frac{A}{B'} = \frac{1}{1/m+f'}\int_{(f,f']} x\,\Gamma^{(k-1)}(\mathrm{d}x),\\ \limsup_{n\to\infty} X_n &\leq \frac{A}{B} = \frac{1}{1/m+f}\int_{(f,f']} x\,\Gamma^{(k-1)}(\mathrm{d}x).\end{aligned}$$

Analogous to the proof in [19], we then obtain

$$\Gamma^{(k)}(\mathrm{d}x) = \Big(\frac{x}{x+1/m}\Big)^k \Gamma^{(0)}(\mathrm{d}x).$$

With similar adjustments, it follows that

$$\Gamma^{(0)}(\mathrm{d}x) = \frac{1/m}{x+1/m}\mu(\mathrm{d}x),$$

from which (2.3), (2.4) and (2.5) follow. Now, we prove (2.6) for $m = 1$ (the proof for $m > 1$ follows analogously). For the first steps, we can directly follow the proof of Theorem 2.6(iii) in [19]. Let $\beta \in (0, (2-\alpha)/(\alpha-1)\wedge 1)$. We obtain

$$\mathbb{P}\Big(E_n^c \cap \{W_{U_n} \leq n^\beta\}\Big) \leq \frac{1}{n}\sum_{j=1}^{n-1}\sum_{k=1}^{j} n^\beta \mathbb{E}\left[\frac{1}{S_j}\mathbb{1}_{\{W_k\leq n^\beta\}}\right] \leq Cn^{\beta-1}\sum_{j=1}^{n-1} j\mathbb{E}\left[\frac{1}{M_j}\right], \tag{4.1}$$

where we bound S_j from below by the maximum vertex-weight $M_j := \max_{i\in[j]} W_i$ and $C > 0$ is a constant. We can then bound the expected value of $1/M_j$ by

$$\mathbb{E}\left[1/M_j\right] \leq \mathbb{P}\Big(M_j \leq j^{1/(\alpha-1)-\varepsilon}\Big)/x_l + j^{-1/(\alpha-1)+\varepsilon}\mathbb{P}\Big(M_j \geq j^{1/(\alpha-1)-\varepsilon}\Big).$$

The second probability can be bounded by 1, and for j large, say $j > j_0 \in \mathbb{N}$, we can bound the first probability from above by

$$\mathbb{P}\Big(M_j \leq j^{1/(\alpha-1)-\varepsilon}\Big) \leq \exp\{-j^{(\alpha-1)\varepsilon/2}\},$$

which leads to the bound

$$\mathbb{E}\left[1/M_j\right] \leq \mathbb{1}_{\{j\leq j_0\}}/x_l + \mathbb{1}_{\{j>j_0\}}(1+1/x_l)j^{-1/(\alpha-1)+\varepsilon}.$$

We then use this in (4.1) to obtain

$$\mathbb{P}\Big(E_n^c \cap \{W_{U_n} \leq n^\beta\}\Big) \leq \widetilde{C}n^{\beta-((2-\alpha)/(\alpha-1)\wedge 1)+\varepsilon},$$

for some constant $\widetilde{C} > 0$. Combining this with

$$\mathbb{P}\Big(W_{U_n} \geq n^\beta\Big) = \ell(n^\beta)n^{-\beta/(\alpha-1)} \leq n^{-\beta/(\alpha-1)+\varepsilon},$$

for n sufficiently large, by [4, Proposition 1.3.6 (v)], yields

$$\mathbb{P}(E_n) \geq 1 - n^{-\beta(\alpha-1)+\varepsilon} - \widetilde{C}n^{\beta-((2-\alpha)/(\alpha-1)\wedge 1)+\varepsilon}.$$

Taking $C = 1+\widetilde{C}$ and choosing the optimal value of β, namely $\beta = ((2-\alpha)/(\alpha(\alpha-1)))\wedge(1/\alpha)$, yields the desired result and concludes the proof. □

Prior to proving Theorem 2.7, we prove Lemma 2.6 as stated in Section 2.

Proof of Lemma 2.6. Without loss of generality, we can set $x_0 = 1$. The claim relating the Weibull and Fréchet maximum domains of attractions follows directly from [25, Propositions 1.11 and 1.13] and the fact that $\mathbb{P}(W \geq 1-1/x) = \mathbb{P}\big((1-W)^{-1} \geq x\big)$.

By [25, Corollary 1.7], the random variable W belongs to the Gumbel MDA if and only if there exist a $z_0 < 1$ and measurable functions c, g, f such that

$$\lim_{x\to 1} c(x) = \widehat{c} > 0, \qquad \lim_{x\to 1} g(x) = 1,$$

and

$$\mathbb{P}(W \geq x) = c(x) \exp\Big\{ -\int_{z_0}^{x} g(t)/f(t)\,\mathrm{d}t \Big\}, \qquad z_0 < x < 1,$$

with f, known as the auxiliary function, being absolutely continuous, $f > 0$ on $(z_0, 1)$ and $\lim_{x\to 1} f'(x) = 0$. [25, Lemma 1.2] states that $\lim_{u\uparrow 1}(1-u)^{-1}f(u) = 0$ holds for the function f as well (though this is not a necessary condition for [25, Corollary 1.7]). We are required to find a $\widetilde{z}_0 < \widetilde{x}_0$ and measurable functions $\widetilde{c}, \widetilde{g}, \widetilde{f}$ with the same properties for the random variable $(1-W)^{-1}$ to prove one direction.

First, we readily have that $\widetilde{x}_0 = \sup\{x > 0 : \mathbb{P}\big((1-W)^{-1} \leq x\big) < 1\} = \infty$, so that $(1-W)^{-1}$ has unbounded support. Then, take $\widetilde{z}_0 = (1-z_0)^{-1}$ and note that $\widetilde{z}_0 < \infty$ as $z_0 < 1$. For any $x > \widetilde{z}_0$,

$$\begin{aligned}\mathbb{P}\big((1-W)^{-1} \geq x\big) &= \mathbb{P}(W \geq 1 - 1/x)\\ &= c(1-1/x)\exp\Big\{ -\int_{z_0}^{1-1/x} g(t)/f(t)\,\mathrm{d}t\Big\}\\ &= c(1-1/x)\exp\Big\{ -\int_{\widetilde{z}_0}^{x} g(1-1/u)/f(1-1/u)u^{-2}\,\mathrm{d}u\Big\}.\end{aligned}$$

We can thus set $\widetilde{c}(x) := c(1-1/x), \widetilde{g}(x) := g(1-1/x), \widetilde{f}(x) := f(1-1/x)x^2$. It directly follows that $\widetilde{f}$ is absolutely continuous and strictly positive on $(\widetilde{z}_0, \infty)$. Moreover,

$$\lim_{x\to\infty} \widetilde{c}(x) = \lim_{x\to\infty} c(1-1/x) = \lim_{u\uparrow 1} c(u) = \widehat{c}, \qquad \lim_{x\to\infty} \widetilde{g}(x) = \lim_{u\uparrow 1} g(u) = 1,$$

and

$$\widetilde{f}'(x) = -f'(1-1/x) + 2xf(1-1/x) = -f'(1-1/x) + 2(1-(1-1/x))^{-1}f(1-1/x).$$

Hence, we find that

$$\lim_{x\to\infty} \widetilde{f}'(x) = -\lim_{u\uparrow 1} f'(u) + 2\lim_{u\uparrow 1}(1-u)^{-1}f(u) = 0,$$

so that all the conditions of [25, Corollary 1.7] are satisfied for the tail distribution of $(1-W)^{-1}$.

For the other direction, we use the other result of [25, Lemma 1.2], which states that $\lim_{u\to\infty} u^{-1}f(u) = 0$ when f is an auxiliary function for an unbounded distribution belonging to the Gumbel MDA. With similar steps as above, the required result then follows as well. □

We now prove Theorem 2.7.

Proof of Theorem 2.7. We know from Theorem 2.5, with μ the distribution of W, that

$$p_k = \int_0^\infty \frac{\mathbb{E}[W]/m}{\mathbb{E}[W]/m + x}\Big(\frac{x}{\mathbb{E}[W]/m + x}\Big)^k \mu(\mathrm{d}x), \tag{4.2}$$

for any choice of vertex-weights W such that $\mathbb{E}[W] < \infty$.

In all cases of Theorem 2.7, we provide an upper and lower bound for p_k by using the assumptions on the tail distribution of the vertex-weights and the properties of the integrand of (4.2). We first discuss the **(Bounded)** case, for which $x_0 = \sup\{x \geq 0, \mu(0,x) < 1\} < \infty$, so that we can, without loss of generality, set $x_0 = 1$. In the final part, we then discuss the **(Gumbel)** and **(Fréchet)** cases for which $x_0 = \infty$.

For the bounded case, we recall that $\mathbb{E}[W]/m = \theta_m - 1$. It is straightforward to check that the integrand in (4.2) is unimodal and maximised at $x = \mathbb{E}[W]k/m$. Thus, for

$k > m/\mathbb{E}[W]$, the integrand is increasing and maximal at $x = 1$. This directly yields, for some non-negative sequence $(s_k)_{k\in\mathbb{N}}$ such that $s_k \uparrow 1$,

$$\begin{aligned} p_k &\leq \int_0^{s_k} \Big(\frac{s_k}{\theta_m - 1 + s_k}\Big)^k \mu(\mathrm{d}x) + \int_{s_k}^1 \frac{\theta_m - 1}{\theta_m}\theta_m^{-k}\mu(\mathrm{d}x) \\ &= \Big(\frac{\theta_m s_k}{\theta_m - 1 + s_k}\Big)^k \theta_m^{-k} + \frac{\theta_m - 1}{\theta_m}\theta_m^{-k}\mathbb{P}(W \geq s_k) \\ &\leq \exp\{-(1-\theta_m^{-1})(1-s_k)k\}\theta_m^{-k} + (1-\theta_m^{-1})\theta_m^{-k}\mathbb{P}(W \geq s_k)\,. \end{aligned} \tag{4.3}$$

For a lower bound, we again split the integral at s_k, but only keep the second integral. This yields

$$p_k \geq (1-\theta_m^{-1})\Big(1 - \frac{(\theta_m - 1)(1-s_k)}{\theta_m - 1 + s_k}\Big)^k \theta_m^{-k}\mathbb{P}(W \geq s_k)\,.$$

We now bound the second term from below by setting $s_k = 1 - 1/t_k$ and by writing

$$\Big(1 - \frac{(\theta_m - 1)/t_k}{\theta_m - 1/t_k}\Big)^k = \exp\Big\{k\log\Big(1 - \frac{(\theta_m-1)/t_k}{\theta_m - 1/t_k}\Big)\Big\} = \exp\{-(1-\theta_m^{-1})k/t_k\}(1+o(1)),$$

provided that $\sqrt{k} = o(t_k)$ (otherwise the $(1 + o(1))$ is to be included in the exponent) so that we arrive at

$$p_k \geq (1-\theta_m^{-1})\exp\{-(1-\theta_m^{-1})k/t_k\}\theta_m^{-k}\mathbb{P}(W \geq 1 - 1/t_k)\,(1+o(1)). \tag{4.4}$$

We now prove (2.7) through (2.11), starting with (2.7). Since W belongs to the Weibull MDA (with $x_0 = 1$), $(1-W)^{-1}$ belongs to the Fréchet MDA by Lemma 2.6, so that

$$\mathbb{P}(W \geq 1 - 1/x) = \mathbb{P}\big((1-W)^{-1} \geq x\big) = \ell(x)x^{-(\alpha-1)}, \tag{4.5}$$

for some slowly-varying function ℓ and $\alpha > 1$. Thus, with the upper bound in (4.3) we obtain

$$p_k \leq \Big[\exp\big\{-(1-\theta_m^{-1})k/t_k\big\} + (1-\theta_m^{-1})\ell(t_k)\exp\{-(\alpha-1)\log t_k\}\Big]\theta_m^{-k}.$$

We balance the two terms in the square brackets by setting $t_k = (1-\theta_m^{-1})k/((\alpha-1)\log k)$. This yields

$$p_k \leq \big(k^{-(\alpha-1)} + (1-\theta_m^{-1})^{2-\alpha}(\alpha-1)^{\alpha-1}\log(k)^{\alpha-1}\ell(t_k)k^{-(\alpha-1)}\big)\theta_m^{-k} = \overline{L}(k)k^{-(\alpha-1)}\theta_m^{-k},$$

where $\overline{L}(k) := 1 + (1-\theta_m^{-1})^{2-\alpha}(\alpha-1)^{\alpha-1}\log(k)^{\alpha-1}\ell(t_k)$, which is slowly varying as t_k is regularly varying and ℓ is slowly varying, so that $\ell(t_k)$ is slowly varying by [4, Proposition 1.5.7.(*ii*)].

For a lower bound we use (4.4) and (4.5) and set $t_k = k$ to obtain

$$p_k \geq \underline{L}(k)k^{-(\alpha-1)}\theta_m^{-k},$$

where $\underline{L}(k) := (1-\theta_m^{-1})\mathrm{e}^{-(1-\theta_m^{-1})}\ell(k)$, which proves the second part of (2.7).

To prove (2.8), we use an improved version of (4.3). Recall that $\gamma = 1/(\tau+1)$. Then we define sequences $t_{k,j} := d_j c_1^{1-\gamma}((1-\theta_m^{-1})k)^\gamma, k, j \in [J]$ for constants $d_1 < d_2 < \ldots < d_J$ and some $J \in \mathbb{N}$ that we choose at the end. We also write $f(x) := ((\theta_m-1)/(\theta_m - 1 + x))(x/(\theta_m - 1 + x))^k$ for simplicity. Then, we bound

$$p_k \leq \int_0^{1-1/t_{k,1}} f(x)\mu(\mathrm{d}x) + \sum_{j=1}^{J-1}\int_{1-1/t_{k,j}}^{1-1/t_{k,j+1}} f(x)\mu(\mathrm{d}x) + \int_{1-1/t_{k,J}}^1 f(x)\mu(\mathrm{d}x). \tag{4.6}$$

As $(1-W)^{-1}$ satisfies the **(RV)** sub-case, it follows that

$$\begin{aligned} \mathbb{P}(W \geq 1 - 1/t_{k,j}) &= \mathbb{P}\big((1-W)^{-1} \geq t_{k,j}\big) = (1+o(1))at_{k,j}^b\mathrm{e}^{-(t_{k,j}/c_1)^\tau} \\ &= (1+o(1))at_{k,j}^b\exp\Big\{-d_j^\tau\Big(\frac{(1-\theta_m^{-1})k}{c_1}\Big)^{1-\gamma}\Big\}. \end{aligned}$$

Also using that $f(x)$ is increasing on $[0,1]$ when $k > 1/(\theta_m - 1)$ allow us to bound p_k from above even further by

$$
\begin{aligned}
&f(1-1/t_{k,1}) + \sum_{j=1}^{J-1} f(1-1/t_{k,j+1})\mathbb{P}(W \geq 1-1/t_{k,j}) + f(1)\mathbb{P}(W \geq 1-1/t_{k,J})\\
&\leq \frac{\theta_m - 1}{\theta_m^{k+1}}\bigg[\Big(1-\frac{1}{t_{k,1}}\Big)^k\Big(1-\frac{1}{\theta_m t_{k,1}}\Big)^{-(k+1)} + \sum_{j=1}^{J-1}\Big(1-\frac{1}{t_{k,j+1}}\Big)^k\Big(1-\frac{1}{\theta_m t_{k,j+1}}\Big)^{-(k+1)}\\
&\quad\times a t_{k,j}^b \exp\Big\{-d_j^\tau\Big(\frac{(1-\theta_m^{-1})k}{c_1}\Big)^{1-\gamma}\Big\} + a t_{k,J}^b \exp\Big\{-d_J^\tau\Big(\frac{(1-\theta_m^{-1})k}{c_1}\Big)^{1-\gamma}\Big\}\bigg](1+o(1))\\
&\leq (1-\theta_m^{-1})\theta_m^{-k}\bigg[\exp\Big\{-\frac{1}{d_1}\Big(\frac{(1-\theta_m^{-1})k}{c_1}\Big)^{1-\gamma}\Big\}\\
&\quad+\sum_{j=1}^{J-1} a t_{k,j}^b \exp\Big\{-(d_{j+1}^{-1}+d_j^\tau)\Big(\frac{(1-\theta_m^{-1})k}{c_1}\Big)^{1-\gamma}\Big\}\\
&\quad+ a t_{k,J}^b \exp\Big\{-d_J^\tau\Big(\frac{(1-\theta_m^{-1})k}{c_1}\Big)^{1-\gamma}\Big\}\bigg](1+o(1)).
\end{aligned}
$$

Now, we note that the function $g(t) = t^{-1} + t^\tau$ has a minimum at $d^* = \tau^{-\gamma}$. Now, we choose $d_1 < d^*$ such that $1/d_1 > g(d^*) = \tau^\gamma/(1-\gamma)$ and similarly d_∞ such that $d_\infty^\tau > g(d^*)$. Given any $\varepsilon > 0$, we can now choose J sufficiently large such that for all $d, d' \in [d_1, d^*]$ with $|d-d'| \leq (d_\infty - d_1)/J$, we have that $|d^\tau - (d')^\tau| < \varepsilon$. Finally, we define $d_j = d_1 + \frac{j}{J}(d_\infty - d_1)$ for $j = 2, \dots, J$. In particular, it follows that for any $j = 1, \dots, J-1$,

$$
d_{j+1}^{-1} + d_j^\tau > d_{j+1}^{-1} + d_{j+1}^\tau - \varepsilon \geq g(d^*) - \varepsilon = \tau^\gamma/(1-\gamma) - \varepsilon.
$$

Substituting into the bound for p_k and using that $t_{k,j}^b = d_j^b c_1^{(1-\gamma)b}(1-\theta_m)^{\gamma b}k^{\gamma b} \leq Ck^{\gamma b}$ uniformly in $j \in [J]$ for some constant $C > 0$, we arrive at

$$
\begin{aligned}
p_k &\leq (1-\theta_m^{-1})\theta_m^{-k}\exp\Big\{-\Big(\frac{\tau^\gamma}{1-\gamma}-\varepsilon\Big)\Big(\frac{(1-\theta_m^{-1})k}{c_1}\Big)^{1-\gamma}\Big\}\Big(1+\sum_{j=1}^{J} aCk^{\gamma b}\Big)\\
&\leq \exp\Big\{-\Big(\frac{\tau^\gamma}{1-\gamma}-2\varepsilon\Big)\Big(\frac{(1-\theta_m^{-1})k}{c_1}\Big)^{1-\gamma}\Big\}\theta_m^{-k},
\end{aligned}
$$

where the last inequality holds for k large enough.

For a lower bound, we use (4.4), but now with $t_k = c_1^{1-\gamma}((1-\theta_m^{-1})k/\tau)^\gamma$. We thus obtain

$$
p_k \geq (1-\theta_m^{b\gamma-1})\tau^{-\gamma b}c_1^{b(1-\gamma)}k^{\gamma b}\exp\Big\{-(\tau^\gamma+\tau^{\gamma-1})\Big(\frac{(1-\theta_m^{-1})k}{c_1}\Big)^{1-\gamma}(1+o(1))\Big\}\theta_m^{-k}. \quad (4.7)
$$

As $\sqrt{k} = o(t_k)$ is not guaranteed for all values of $\tau > 0$, we include the $1+o(1)$ in the exponent. We thus obtain

$$
p_k \geq \exp\Big\{-\frac{\tau^\gamma}{1-\gamma}\Big(\frac{(1-\theta_m^{-1})k}{c_1}\Big)^{1-\gamma}(1+o(1))\Big\}\theta_m^{-k},
$$

which together with (4.7) yields (2.8).

We now prove (2.9), for which a similar approach is applied. Again, we choose a sequence $t_{k,1} < \ldots < t_{k,J}$ with $J = J(k)$ to be determined later. This time however, we take $t_{k,j} = (d_{J-j+1})^{-1}(1-\theta_m^{-1})k(\log k)^{-(\tau-1)}$ for a sequence $d_1 < d_2 < \ldots < d_J$ to be fixed later on, but such d_1 is bounded in k and $\log(d_J) = o(\log k)$. Then, by the same estimate as above but now using that $(1-W)^{-1}$ satisfies the **(RaV)** sub-case, we obtain

$$\begin{aligned} p_k \le (1-\theta_m^{-1})\theta_m^{-k}\Big[&\exp\Big\{-d_J(\log k)^{\tau-1}\Big\} \\ &+\sum_{j=1}^{J-1} a(\log t_{k,j})^b \exp\Big\{-d_{J-j}(\log k)^{\tau-1}-(\log(t_{k,j})/c_1)^\tau\Big\} \\ &+a(\log t_{k,J})^b\exp\{-(\log(t_{k,J})/c_1)^\tau\}\Big](1+o(1)). \end{aligned} \tag{4.8}$$

Now, we use that by a Taylor expansion

$$(\log t_{k,j}/c_1)^\tau \ge \Big(\frac{\log k}{c_1}\Big)^\tau\Big(1-\tau(\tau-1)\frac{\log\log k}{\log k}\Big)-\frac{\tau}{c_1^\tau}\log\Big(\frac{d_{J-j+1}}{1-\theta_m^{-1}}\Big)(\log k)^{\tau-1}.$$

Hence, we obtain that there exists a constant $C>0$ such that

$$\begin{aligned} p_k \le C\theta_m^{-k}(\log k)^{b\vee 0}\Big[&\exp\Big\{-d_J(\log k)^{\tau-1}\Big\} \\ &+\exp\Big\{-\Big(\frac{\log k}{c_1}\Big)^\tau\Big(1-\tau(\tau-1)\frac{\log\log k}{\log k}\Big)\Big\} \\ &\times\Big\{\sum_{j=1}^{J-1}\exp\Big\{-\Big(d_{J-j}-\frac{\tau}{c_1^\tau}\log\Big(\frac{d_{J-j+1}}{1-\theta_m^{-1}}\Big)\Big)(\log k)^{\tau-1}\Big\} \\ &+\exp\Big\{\frac{\tau}{c_1^\tau}\log\Big(\frac{d_1}{1-\theta_m^{-1}}\Big)(\log k)^{\tau-1}\Big\}\Big\}\Big]. \end{aligned} \tag{4.9}$$

We eventually choose d_J such that $d_J \ge c_1^{-\tau}\log k$, so that we can neglect the first term. Secondly, we notice that the function $f(x)=x-\tau c_1^{-\tau}\log(x/(1-\theta_m^{-1}))$ is minimised at $x^*=\tau c_1^{-\tau}$, so we choose d_1 small enough such that $\tau c_1^{-\tau}\log(d_1/(1-\theta_m^{-1})) < f(x^*) = \tau c_1^{-\tau}\log(\mathrm{e}c_1^\tau(1-\theta_m^{-1})/\tau)$. Therefore, we can neglect the first term and the term outside the sum and can concentrate on the sum itself and so need to estimate

$$\sum_{j=1}^{J-1}\exp\Big\{-\Big(d_j-\frac{\tau}{c_1^\tau}\log\Big(\frac{d_{j+1}}{1-\theta_m^{-1}}\Big)\Big)(\log k)^{\tau-1}\Big\} \tag{4.10}$$

Let d_∞ be big enough such that $\tau c_1^{-\tau}\log(d/(1-\theta_m^{-1}))\le d/2$ for all $d\ge d_\infty$ and also big enough such that $d_\infty \ge 2(f(x^*)+1)$. Given $\varepsilon>0$, let J' be such that $J'\ge\varepsilon^{-1}(d_\infty-d_1)$ (note that J' does not depend on k). Then define $d_j=d_1+(j/J')(d_\infty-d_1)$ for $j=1,\dots,J'$. Moreover, choose $d_j=d_\infty+(j-J')$ for $j\ge J'+1$. Finally, choose J such that $d_{J-1}\le(\log k)/(c_1)^\tau\le d_J$. We split the sum in (4.10) into summands smaller and bigger than J' and first consider

$$\begin{aligned} &\sum_{j=1}^{J'-1}\exp\Big\{-\Big(d_j-\frac{\tau}{c_1^\tau}\log\Big(\frac{d_{j+1}}{1-\theta_m^{-1}}\Big)\Big)(\log k)^{\tau-1}\Big\} \\ &\quad\le\sum_{j=1}^{J'-1}\exp\Big\{-\Big(d_{j+1}-\frac{\tau}{c_1^\tau}\log\Big(\frac{d_{j+1}}{1-\theta_m^{-1}}\Big)-\varepsilon\Big)(\log k)^{\tau-1}\Big\} \\ &\quad\le J'\exp\Big\{-(f(x^*)-\varepsilon)(\log k)^{\tau-1}\Big\}. \end{aligned}$$

Now, for the second sum we obtain by the assumptions on d_j,

$$
\begin{aligned}
\sum_{j=J'}^{J-1} \exp\Big\{ -\Big(d_j - \frac{\tau}{c_1^\tau}\log\Big(\frac{d_{j+1}}{1-\theta_m^{-1}}\Big)\Big)(\log k)^{\tau-1}\Big\}
&\le \sum_{j=J'}^{J-1} \exp\Big\{ -(d_j - d_{j+1}/2)(\log k)^{\tau-1}\Big\}\\
&\le \sum_{j=J'}^{J-1} \exp\Big\{ -\Big(\frac{1}{2}(d_\infty + (j-J')) - \frac{1}{2}\Big)(\log k)^{\tau-1}\Big\}\\
&\le \exp\Big\{ -\frac{1}{2}(d_\infty - 1)(\log k)^{\tau-1}\Big\}\sum_{j=0}^{\infty}\exp\Big\{-\frac{j}{2}(\log k)^{\tau-1}\Big\}
\end{aligned}
$$

By the assumption, we have that $(d_\infty - 1)/2 \ge f(x^*)$, so that combining the two last estimates with (4.9), we obtain the upper bound

$$
C_1\theta_m^{-k}(\log k)^{b\vee 0}\exp\Big\{ -\Big(\frac{\log k}{c_1}\Big)^{\tau}\Big(1-\tau(\tau-1)\frac{\log\log k}{\log k}\Big) - (\log k)^{\tau-1}(f(x^*)-\varepsilon)\Big\}, \tag{4.11}
$$

where C_1 is some positive constant. This produces the required bound as we recall that $f(x^*) = \tau c_1^{-\tau}\log(\mathrm{e}c_1^\tau(1-\theta_m^{-1})/\tau)$.

For a lower bound, we set $t_k = (1-\theta_m^{-1})k/(x^*(\log k)^{\tau-1})$, where $x^* = \tau c_1^{-\tau}$ as before. Then, we use (4.4) to find

$$
\begin{aligned}
p_k &\ge (1-\theta_m^{-1})a\log(t_k)^b\exp\{-(1-\theta_m^{-1})k/t_k - (\log(t_k)/c_1)^\tau\}\theta_m^{-k}(1+o(1))\\
&\ge C_2\log(k)^b\exp\Big\{ -\Big(\frac{\log k}{c_1}\Big)^{\tau}\Big(1-\tau(\tau-1)\frac{\log\log k}{\log k}\Big)\\
&\qquad - (x^* - \tau c_1^{-\tau}\log(x^*/(1-\theta_m^{-1})))(\log k)^{\tau-1}(1+o(1))\Big\}\theta_m^{-k},
\end{aligned}
$$

for some constant $C_2 > 0$, which proves the lower bound in (2.9) since we recall that $f(x^*) = \tau c_1^{-\tau}\log(\mathrm{e}c_1^\tau(1-\theta_m^{-1})/\tau)$.

Finally, we prove (2.11). As $q_0 = \mathbb{P}(W = x_0) = \mathbb{P}(W = 1) > 0$, we immediately obtain the lower bound

$$
p_k \ge q_0(1-\theta_m^{-1})\theta_m^{-k}.
$$

For an upper bound, recall s_k and r_k from (2.10). Then, (4.3) yields

$$
\begin{aligned}
p_k &\le \exp\{-(1-\theta_m^{-1})(1-s_k)k\}\theta_m^{-k} + (1-\theta_m^{-1})\theta_m^{-k}\mu((s_k,1)) + q_0(1-\theta_m^{-1})\theta_m^{-k}\\
&= q_0(1-\theta_m^{-1})\theta_m^{-k}\big(1+\mathcal{O}\big(\exp\{-(1-\theta_m^{-1})(1-s_k)k\}\vee\mu((s_k,1))\big)\big).
\end{aligned}
$$

By the right-continuity of the tail distribution, we obtain that the maximum in the big $\mathcal{O}$ notation equals $\exp\{-(1-\theta_m^{-1})(1-s_k)k\}$ by the definition of s_k, which proves (2.11).

In the final part of the proof, we prove (2.12) through (2.14) for unbounded weights, i.e. for which $x_0 = \infty$. Without loss of generality, we can now assume that $\mathbb{E}[W] = 1$, so that the expression in (4.2) simplifies to

$$
p_k = \mathbb{E}\left[\frac{1}{1+mW}\Big(\frac{mW}{1+mW}\Big)^k\right] = \int_0^\infty \frac{1}{1+mx}\Big(\frac{mx}{1+mx}\Big)^k\mu(\mathrm{d}x).
$$

Recall that the integrand on the right-hand side is a unimodal function which obtains its maximum at $x = k/m$. It thus is increasing (resp. decreasing) for $x < k/m$ (resp. $x > k/m$). The aim is (again) to identify an increasing, now diverging sequence s_k such that $s_k \le k/m$ and write p_k as

$$
p_k = \int_0^{s_k} \frac{1}{1+mx}\Big(\frac{mx}{1+mx}\Big)^k\mu(\mathrm{d}x) + \int_{s_k}^\infty \frac{1}{1+mx}\Big(\frac{mx}{1+mx}\Big)^k\mu(\mathrm{d}x), \tag{4.12}
$$

so that we can bound p_k from above and below by

$$\begin{aligned}
p_k &\leq \frac{1}{s_k}\Big(\Big(\frac{ms_k}{1+ms_k}\Big)^k + \mathbb{P}(W\geq s_k)\Big) \leq \frac{1}{s_k}\Big(\exp\Big\{-\frac{k}{ms_k}+\frac{k}{(ms_k)^2}\Big\}+\mathbb{P}(W\geq s_k)\Big),\\
p_k &\geq \Big(\frac{ms_k}{1+ms_k}\Big)^k\int_{s_k}^\infty \frac{1}{1+mx}\,\mu(\mathrm{d}x) \geq \Big(1-\frac{1}{ms_k}\Big)^k\int_{s_k}^\infty\frac{1}{1+mx}\,\mu(\mathrm{d}x).
\end{aligned} \tag{4.13}$$

Here, we use that the expression in the expected values is increasing when $W\leq s_k\leq k/m$ in the first term of upper bound, and for the lower bound we only consider the second expected value in (4.12) and use that $(x/(1+x))^k$ is increasing in x for any $k\in\mathbb{N}_0$. The goal is then to choose s_k such that the exponent and the tail probability in the upper bound are of the same order, and to choose (a possibly different) s_k for the lower bound such that the product of the exponent and integral behaves similar to the upper bound.

For (2.12), however, we use an improved upper bound. As in (4.6), writing $f(x)=(mx)^k(1+mx)^{-(k+1)}$ and taking sequences $s_{k,j}, k,j\in\mathbb{N}$ such that $s_{k,j}\leq s_{k,j+1}$ and $s_{k,j}\leq k/m$ (so that $f(s_{k,j})$ is increasing in j) for all $j\in\mathbb{N}$ and k large,

$$\begin{aligned}
p_k &\leq \int_0^{s_{k,1}} f(x)\mu(\mathrm{d}x)+\sum_{j=1}^{J-1}\int_{s_{k,j}}^{s_{k,j+1}} f(x)\mu(\mathrm{d}x)+\int_{s_{k,J}}^\infty f(x)\mu(\mathrm{d}x)\\
&\leq f(s_{k,1})+\sum_{j=1}^{J-1} f(s_{k,j+1})\mathbb{P}(W\geq s_{k,j})+\mathbb{P}(W\geq s_{k,J})\,.
\end{aligned}$$

Then using similar bounds as in (4.13), we obtain the upper bound

$$\begin{aligned}
&\frac{1}{s_{k,1}}\exp\Big\{-\frac{k}{ms_{k,1}}+\frac{k}{(ms_{k,1})^2}\Big\}\\
&+\sum_{j=1}^{J-1}\frac{1}{s_{k,j+1}}\exp\Big\{-\frac{k}{ms_{k,j+1}}+\frac{k}{(ms_{k,j+1})^2}\Big\}as_{k,j}^b\exp\Big\{-(s_{k,j}/c_1)^\tau\Big\}(1+o(1))\\
&+as_{k,J}^{b-1}\exp\Big\{-(s_{k,J}/c_1)^\tau\Big\}(1+o(1)),
\end{aligned} \tag{4.14}$$

where $J\geq 2$ is some large integer. We then set $s_{k,j}=d_jc_1^{1-\gamma}(k/m)^\gamma$ for some constants $d_1<d_2<\ldots<d_J$ (so that $s_{k,j}\leq s_{k,j+1}$ and $s_{k,j}\leq k/m$ holds for all $j\in\mathbb{N}$ and all k large) and note that this bound is similar to the one developed in the proof of the upper bound in (2.8), but with $1-\theta_m^{-1}$ replaced by $1/m$, some additional lower order terms in the exponents and different constants. We can thus use the same approach to conclude that for any fixed $\varepsilon>0$, we can take J large enough such that we obtain the upper bound

$$\begin{aligned}
p_k &\leq \exp\Big\{-(1-\varepsilon)\frac{\tau^\gamma}{1-\gamma}\Big(\frac{k}{c_1m}\Big)^{1-\gamma}\Big\}\Big(\frac{m^\gamma}{d_1c_1^{1-\gamma}k^\gamma}\\
&\qquad + ac_1^{(b-1)(1-\gamma)}m^{-(b-1)\gamma}k^{(b-1)\gamma}\sum_{j=1}^{J-1}d_j^bd_{j+1}^{-1}+ad_J^{b-1}c_1^{(1-\gamma)(b-1)}m^{-\gamma(b-1)}k^{\gamma(b-1)}\Big)\\
&\leq \exp\{-(1-2\varepsilon)\frac{\tau^\gamma}{1-\gamma}\Big(\frac{k}{c_1m}\Big)^{1-\gamma}\Big\}.
\end{aligned} \tag{4.15}$$

For a lower bound we set $s_k=c_1^{1-\gamma}(k/(\tau m))^\gamma$ and use (4.13) to obtain

$$p_k\geq\Big(1-(c_1m)^{-(1-\gamma)}(k/\tau)^{-\gamma}\Big)^k\int_{s_k}^\infty\frac{1}{1+mx}\mu(\mathrm{d}x). \tag{4.16}$$

The first term can be written as

$$\exp\{k\log(1-(c_1m)^{-(1-\gamma)}(k/\tau)^{-\gamma})\}=\exp\{-(c_1m)^{-(1-\gamma)}\tau^\gamma k^{1-\gamma}(1+o(1))\}. \tag{4.17}$$

For the second term, we use that for any $\varepsilon \in (0,\tau)$ and all k sufficiently large,

$$(s_k + s_k^{1-\tau+\varepsilon})^b = s_k^b(1+o(1)), \qquad \Big(\frac{s_k + s_k^{1-\tau+\varepsilon}}{c_1}\Big)^\tau \le (s_k/c_1)^\tau + 2\tau s_k^\varepsilon/c_1^\tau,$$

so that $\mathbb{P}\big(W \ge s_k + s_k^{1-\tau+\varepsilon}\big) = o(\mathbb{P}(W \ge s_k))$. As a result we can bound the integral in (4.16) from below by

$$\begin{aligned}\frac{1}{m+1}\int_{s_k}^{s_k+s_k^{1-\tau+\varepsilon}} x^{-1}\mu(\mathrm{d}x) &\ge \frac{1+o(1)}{(m+1)s_k}\big(\mathbb{P}(W\ge s_k) - \mathbb{P}\big(W \ge s_k + s_k^{1-\tau+\varepsilon}\big)\big)\\ &\ge \frac{a(1+o(1))}{m+1} s_k^{b-1}\exp\big\{-(s_k/c_1)^\tau\big\}\\ &\ge Kk^{\gamma(b-1)}\exp\big\{-\tau^{\gamma-1}(k/(c_1m))^{1-\gamma}\big\}.\end{aligned}$$

for some small constant $K>0$. Combining this with (4.17) in (4.16), we arrive at

$$\begin{aligned}p_k &\ge \exp\{-(c_1m)^{-(1-\gamma)}\tau^\gamma k^{1-\gamma}(1+o(1))\}Kk^{\gamma(b-1)}\exp\big\{-\tau^{\gamma-1}(k/(c_1m))^{1-\gamma}\big\}\\ &= \exp\Big\{-\frac{\tau^\gamma}{1-\gamma}(k/(c_1m))^{1-\gamma}(1+o(1))\Big\}.\end{aligned}$$

Combined with (4.15) this proves (2.12).

For the **(RaV)** sub-case, we use a similar approach as for the **(RV)** sub-case. For $j = 1,\dots,J$, we now set $s_{k,j} = d_{J-j+1}^{-1}(k/m)(\log(k/m))^{-(\tau-1)}$ for a sequence $d_1 < d_2 < \ldots < d_J$ and J to be determined later on, such that d_1 is bounded in k and $\log d_J = o(\log k)$. Additionally, we use that the weights satisfy the **(RaV)** sub-case and use (4.14) to obtain

$$\begin{aligned}p_k \le{}& \Big[\frac{1}{s_{k,1}}\exp\Big\{-\frac{k}{ms_{k,1}} + \frac{k}{(ms_{k,1})^2}\Big\}\\ &+ \sum_{j=1}^{J-1}\frac{1}{s_{k,j+1}}\exp\Big\{-\frac{k}{ms_{k,j+1}}+\frac{k}{(ms_{k,j+1})^2}\Big\}a(\log s_{k,j})^b\exp\Big\{-(\log(s_{k,j})/c_1)^\tau\Big\}\\ &+ as_{k,J}^{-1}\log s_{k,J}^b\exp\Big\{-(\log(s_{k,J})/c_1)^\tau\Big\}\Big](1+o(1))\\ ={}& \Big[d_J\frac{m\log(k)^{\tau-1}}{k}\exp\Big\{-d_J(\log(k/m))^{\tau-1}+o(1)\Big\}\\ &+\sum_{j=1}^{J-1} ad_{J-j}\frac{m\log(k)^{b+\tau-1}}{k}\exp\Big\{-d_{J-j}(\log(k/m))^{\tau-1} - (\log(s_{k,j})/c_1)^\tau + o(1)\Big\}\\ &+ad_1\frac{m\log(k)^{b+\tau-1}}{k}\exp\Big\{-(\log(s_{k,J}/c_1)^\tau\Big\}\Big](1+o(1)).\end{aligned}$$

We find that determining the optimal value of the $d_1,\dots,d_J$ follows a similar approach as in the case when $(1-W)^{-1}$ satisfies the **(RaV)** sub-case in (4.8) through (4.11) (but with k replaced with k/m in the exponent and $1-\theta_m^{-1}$ omitted). As a result, we obtain for any $\varepsilon > 0$,

$$p_k \le k^{-1}\exp\Big\{-\Big(\frac{\log(k/m)}{c_1}\Big)^\tau\Big(1-\tau(\tau-1)\frac{\log\log(k/m)}{\log(k/m)} + \frac{\tau\log(\mathrm{e}c_1^\tau/\tau)-\varepsilon}{\log(k/m)}\Big)\Big\}. \tag{4.18}$$

For a lower bound on p_k we set $s_k = c_1^\tau\tau^{-1}(k/m)(\log(k/m))^{-(\tau-1)}$. As $s_k/\sqrt{k}$ diverges, it follows that we can improve the bound in (4.13) to find for some small constant $C>0$,

$$p_k \ge C\exp\{-k/(ms_k)\}\int_{s_k}^\infty \frac{1}{1+mx}\,\mu(\mathrm{d}x) \ge C_m\exp\{-k/(ms_k)\}\int_{s_k}^{2s_k} x^{-1}\,\mu(\mathrm{d}x), \tag{4.19}$$

for some constant $C_m > 0$. Now, since $\tau > 1$, when k is large,

$$(\log(2s_k)/c_1)^\tau \le (\log(s_k)/c_1)^\tau + 2\tau c_1^{-\tau}\log 2(\log s_k)^{\tau-1},$$

so that $\mathbb{P}(W \geq 2s_k) = o(\mathbb{P}(W \geq s_k))$. We can thus bound (4.19) from below by

$$\begin{aligned} &C_m \exp\{-k/(ms_k)\}(2s_k)^{-1}\mathbb{P}(W \geq s_k)\,(1+o(1)) \\ &\quad \geq C_2 s_k^{-1}(\log s_k)^b \exp\{-(\log(s_k)/c_1)^\tau - k/(ms_k)\}, \end{aligned}$$

for some constant $C_2 > 0$. Using the precise value of s_k and a Taylor expansion of $(\log s_k)^\tau$ yields

$$p_k \geq C_2 k^{-1} \exp\Big\{ -\Big(\frac{\log k}{c_1}\Big)^\tau \Big(1 - \tau(\tau-1)\frac{\log\log k}{\log k} + \frac{\tau \log(\mathrm{e}c_1^\tau/\tau)}{\log k}(1+o(1))\Big)\Big\}.$$

Combined with (4.18) this proves (2.13).

Finally, we prove (2.14). We first set $s_k = k/(m(\alpha-1+\varepsilon)\log k)$ for some $\varepsilon > 0$ (note $s_k \leq k/m$). Then, using (4.13) we bound p_k from above by

$$\begin{aligned} p_k &\leq m(\alpha-1+\varepsilon)k^{-1}\log k\big(k^{-(\alpha-1+\varepsilon)}(1+o(1)) + \ell(s_k)(m(\alpha-1+\varepsilon)\log k)^{\alpha-1}k^{-(\alpha-1)}\big) \\ &= o(k^{-\alpha}) + L(k)k^{-\alpha}, \end{aligned}$$

where $L(k) := (m(\alpha-1+\varepsilon)\log k)^\alpha \ell(k/(m(\alpha-1+\varepsilon)\log k))$ is slowly varying by [4, Proposition 1.5.7 (ii)]. The required upper bound is obtained by taking $\overline{\ell}(k) := (1+\varepsilon)L(k)$.

To conclude the proof, we construct a lower bound for p_k. We set $s_k = k/m$ and use the improved lower bound for p_k as in the first line of (4.19) to obtain

$$\begin{aligned} p_k &\geq \frac{C}{\mathrm{e}}\int_{k/m}^\infty \frac{1}{1+mx}\mu(\mathrm{d}x) \geq \frac{Cm}{\mathrm{e}(1+2k)}(\mathbb{P}(W \geq k/m) - \mathbb{P}(W \geq 2k/m)) \\ &= \frac{Cm}{3\mathrm{e}}k^{-1}\ell(k/m)(k/m)^{-(\alpha-1)}\Big(1 - \frac{\ell(2k/m)}{\ell(k/m)}2^{-(\alpha-1)}\Big). \end{aligned}$$

As ℓ is slowly-varying at infinity, is follows that the last term can be bounded from below by a constant, as the fraction converges to one, and that $\ell(k/m) \geq \ell(k)/2$, when k is large. As a result,

$$p_k \geq C_2 \ell(k)k^{-\alpha} =: \underline{\ell}(k)k^{-\alpha},$$

where $C_2 > 0$ is a suitable constant. □

5. The maximum conditional mean degree in WRGs

It turns out that the analysis of the maximum degree of WRGs can be carried out via the maximum of the conditional mean degrees under certain assumptions on the vertex-weight distribution. For WRGs it is necessary for the vertex-weights to have unbounded support. To this end, we formulate four propositions to describe the behaviour of the maximum conditional mean degree when the vertex-weights satisfy the different conditions in Assumption 2.3. Let us first introduce an important quantity, namely the location of the maximum conditional mean degree,

$$\widetilde{I}_n := \inf\{i \in [n] : \mathbb{E}_W[\mathcal{Z}_n(i)] \geq \mathbb{E}_W[\mathcal{Z}_n(j)] \text{ for all } j \in [n]\}.$$

Furthermore, it is important to note that, as $\mathcal{Z}_n(i)$ is a sum of indicator random variables for any $i \in [n]$, its conditional mean equals

$$\mathbb{E}_W[\mathcal{Z}_n(i)] = mW_i \sum_{j=i}^{n-1} \frac{1}{S_j}, \tag{5.1}$$

where we recall that $S_j = \sum_{\ell=1}^j W_\ell$. This is also true when we work with the model with a *random out-degree*, as discussed in Remark 2.2(*ii*), so that all the results in the upcoming propositions also hold for this model when setting $m = 1$.

Another important result that we use throughout the proofs of the propositions is the following lemma. We note that the conditions in the lemma are satisfied for all cases in

Assumption 2.3 such that $\mathbb{E}[W] < \infty$. A similar result (under a different condition) can be found in [2, Theorem 1].

Lemma 5.1. *Let $W_1, \dots,$ be i.i.d. non-negative random variables such that $W_i > 0$ a.s. and $\mathbb{E}\left[W^{1+\varepsilon}\right] < \infty$ for some $\varepsilon > 0$. Moreover, we assume that $\mathbb{E}[W_i] = 1$ and write $S_n = \sum_{i=1}^n W_i$. Then, there exists an almost surely finite random variable Y such that*

$$\sum_{j=1}^{n-1} \frac{1}{S_j} - \log n \xrightarrow{a.s.} Y. \tag{5.2}$$

Proof. We first write

$$\sum_{j=1}^{n-1} \frac{1}{S_j} - \log n = \sum_{j=1}^{n-1} \frac{j - S_j}{jS_j} + \sum_{j=1}^{n-1} \frac{1}{j} - \log n =: \sum_{j=1}^{n-1} \frac{j - S_j}{jS_j} + E_n, \tag{5.3}$$

where E_n is deterministic and converges to the Euler-Mascheroni constant. Therefore, it suffices to show that the first sum on the right hand side is almost surely absolutely convergent.

By the strong law of large numbers and since $\mathbb{E}[W_i] = 1$, almost surely there exists J such that that $S_j > \frac{1}{2}j$ for all $j \geq J$ almost surely. So, we can bound almost surely,

$$\sum_{j=1}^{n-1} \frac{|j - S_j|}{jS_j} \leq \sum_{j=1}^{J-1} \frac{|j - S_j|}{jS_j} + 2\sum_{j=J}^{n-1} \frac{|j - S_j|}{j^2}.$$

The first term is finite almost surely as each $W_i > 0$ as. We now claim that the second term has a finite mean. Namely, using the ε from the assumption,

$$\mathbb{E}\left[\sum_{j=J}^{n-1} \frac{|j - S_j|}{j^2}\right] \leq \sum_{j=1}^{\infty} \frac{1}{j^2} \mathbb{E}\left[|j - S_j|^{1+\varepsilon}\right]^{1/(1+\varepsilon)} \leq \sum_{j=1}^{\infty} \frac{c_\varepsilon}{j^2} j^{1/(1+\varepsilon)},$$

which is summable, where $c_\varepsilon > 0$ is a constant and where we use a Zygmund-Marcinkiewicz bound, see [14, Corollary 8.2] in the last step (note this bound can only be used for $\varepsilon \in (0, 1]$, but when $\varepsilon > 1$ we can always take a smaller value of ε such that the assumptions of the lemma are still satisfied). Therefore, the sum on the right hand side in (5.3) is almost surely (absolutely) convergent, which completes the proof. □

Finally, we recall that it suffices to state the proofs of the results below for $m = 1$ only, as the expected degrees scale linearly with m, see (5.1).

Proposition 5.2. *Consider the WRG model as in Definition 2.1 and suppose the vertex-weights satisfy the* **(Gumbel)-(SV)** *sub-case in Assumption 2.3. Then,*

$$\left(\max_{i \in [n]} \frac{\mathbb{E}_W[\mathcal{Z}_n(i)]}{mb_n \log n}, \frac{\log \widetilde{I}_n}{\log n}\right) \xrightarrow{\mathbb{P}} (1, 0). \tag{5.4}$$

Proof. Let $\beta \in (0, 1)$. Using (5.1) with $m = 1$ and restricting the maximum to only vertices $i \in [n^{1-\beta}]$, we can bound

$$\max_{i \in [n]} \frac{W_i \sum_{j=i}^{n-1} 1/S_j}{b_n \log n} \geq \max_{i \in [n^{1-\beta}]} \frac{W_i \sum_{j=\lceil n^{1-\beta} \rceil}^{n-1} 1/S_j}{b_n \log n} = \Big(\max_{i \in [n^{1-\beta}]} \frac{W_i}{b_{n^{1-\beta}}}\Big) \frac{\sum_{j=\lceil n^{1-\beta} \rceil}^{n-1} 1/S_j}{\log n} \frac{b_{n^{1-\beta}}}{b_n}.$$

We then note that $b_{n^{1-\beta}}/b_n = \ell((1-\beta)\log n)/\ell(\log n) \to 1$ as n tends to infinity, since ℓ is slowly varying at infinity. Furthermore, the maximum on the right-hand side tends to 1 in probability and the fraction in the middle converges to β almost surely by (5.2). Thus, with high probability,

$$\max_{i \in [n]} \frac{W_i \sum_{j=i}^{n-1} 1/S_j}{b_n \log n} \geq \beta, \tag{5.5}$$

where we note that we can choose β arbitrarily close to 1. Furthermore, we can immediately obtain an upper bound of the form

$$\max_{i\in[n]} \frac{W_i \sum_{j=i}^{n-1} 1/S_j}{b_n \log n} \le \Big(\max_{i\in[n]} \frac{W_i}{b_n}\Big) \frac{\sum_{j=1}^{n-1} 1/S_j}{\log n}.$$

Here, both the maximum and the second fraction tend to one, the former in probability and the latter almost surely. Hence, with high probability,

$$\max_{i\in[n]} \frac{W_i \sum_{j=i}^{n-1} 1/S_j}{b_n \log n} \le 1+\eta,$$

for any $\eta > 0$. Together with (5.5) this yields the first part of (5.4). Now, for the second part, let $\varepsilon > 0$ and let us define the event, for $\eta < \varepsilon$,

$$E_n := \Big\{ \max_{i\in[n]} \frac{W_i \sum_{j=i}^{n-1} 1/S_j}{b_n \log n} \ge 1-\eta \Big\},$$

which holds with high probability by (5.5) as we can choose $\beta > 1-\eta$. With this definition,

$$\begin{aligned}\mathbb{P}\bigg(\frac{\log \widetilde{I}_n}{\log n} > \varepsilon\bigg) &= \mathbb{P}\bigg(\Big\{\frac{\log \widetilde{I}_n}{\log n} > \varepsilon\Big\} \cap E_n\bigg) + \mathbb{P}(E_n^c)\\ &\le \mathbb{P}\bigg(\max_{i>n^\varepsilon} \frac{W_i \sum_{j=i}^{n-1} 1/S_j}{b_n \log n} \ge 1-\eta\bigg) + \mathbb{P}(E_n^c)\,.\end{aligned}$$

By (5.5) the second probability on the right-hand side tends to zero with n. What remains to show is that the same holds for the first probability. Via an upper bound where we substitute $j = \lfloor n^\varepsilon \rfloor$ for $j = i$ in the summation, we immediately obtain

$$\mathbb{P}\bigg(\max_{i>n^\varepsilon} \frac{W_i \sum_{j=i}^{n-1} 1/S_j}{b_n \log n} \ge 1-\eta\bigg) \le \mathbb{P}\bigg(\Big(\max_{i>n^\varepsilon} \frac{W_i}{b_n}\Big) \frac{\sum_{j=\lfloor n^\varepsilon \rfloor}^{n-1} 1/S_j}{\log n} \ge 1-\eta\bigg) \to 0.$$

Indeed, the maximum over the fitness values scaled by b_n tends to one in probability, and the sum scaled by $\log n$ converges to $1-\varepsilon$ almost surely, so that the product of the two converges to $1-\varepsilon < 1-\eta$ in probability. This concludes the proof. ☐

Before we turn our attention to the maximum conditional mean in-degree in the WRG model for the **(Gumbel)**-**(RV)** sub-case, we first inspect the behaviour of maxima of i.i.d. vertex-weights in this class:

Lemma 5.3 (Almost sure convergence of rescaled maximum vertex-weight)**.** *Let $(W_i)_{i\in\mathbb{N}}$ be i.i.d. random variables that satisfy the* **(Gumbel)**-**(RV)** *sub-case in Assumption 2.3. Then,*

$$\max_{i\in[n]} \frac{W_i}{b_n} \xrightarrow{a.s.} 1.$$

Proof. The almost sure convergence holds for a particular case of a distribution in the **(Gumbel)**-**(RV)** sub-case, as follows from [17, Lemma 4.1]. That is, when the vertex-weights are i.i.d. copies of a random variable W with distribution

$$\mathbb{P}(W \ge x) = \exp\{-x^\tau\}, \tag{5.6}$$

with $\tau \in (0,1]$. We observe that this is indeed a particular example of the **(Gumbel)**-**(RV)** sub-case, with $c_1 = a = 1, b = 0, \tau \in (0,1]$ and where the asymptotic equivalence of the tail distribution is replaced with an equality. Lemma 4.1 in [17] provides an almost sure lower and upper bound for the maximum of n i.i.d. random variables with a distribution as in (5.6). The leading order term in these bounds is asymptotically equal to b_n (with $c_1 = a = 1, b = 0, \tau \in (0,1]$), from which the statement of the lemma follows.

We observe that Lemma 4.1 in [17] can be easily extended to hold for any $\tau > 1$ as well, in which case only lower order terms may need to be adjusted slightly, so that the leading order terms are still asymptotically equivalent to b_n. Thus, it remains to show that for any $\tau>0$, we can extend the case $c_1=a=1, b=0$ to any $c_1, a>0, b\in\mathbb{R}$.

To that end, let $(W_i)_{i\in\mathbb{N}}$ be i.i.d. copies of a random variable W with a tail distribution as in the **(Gumbel)**-**(RV)** sub-case. This implies that there exists a function ℓ such that $\ell(x)\to 1$ as $x\to\infty$, and

$$\mathbb{P}(W\geq x)=\ell(x)ax^b\mathrm{e}^{-(x/c_1)^\tau}.$$

Let $(X_i)_{i\in\mathbb{N}}$ be i.i.d. copies of a random variable X with a tail distribution as in (5.6), which are also independent of the W_i. As follows from [17, Lemma 4.1] and the first steps of the proof,

$$\max_{i\in[n]}\frac{X_i}{b_n}\xrightarrow{a.s.}1. \tag{5.7}$$

Let us write $b_n(X), b_n(W)$ to distinguish between the respective first order growth-rate sequences of X and W, respectively. Define the functions $f,h:\mathbb{R}\to\mathbb{R}$ as $f(x):=x(c_1^{-\tau}-(b\log x+\log a)/x^\tau)^{1/\tau}$ and $h(x):=(f(x)^\tau-\log(\ell(x)))^{1/\tau}$, so that for all $x>0$,

$$\mathbb{P}(W\geq x)=\ell(x)ax^b\exp\{-(x/c_1)^\tau\}=\ell(x)\exp\{-f(x)^\tau\}=\exp\{-h(x)^\tau\}=\mathbb{P}(X\geq h(x)).$$

Hence, $W\overset{d}{=}h^{\leftarrow}(X)$, where $h^{\leftarrow}$ is the generalised inverse of h, defined as $h^{\leftarrow}(x):=\inf\{y\in\mathbb{R}:h(y)\geq x\}, x\in\mathbb{R}$. We can write h as

$$\begin{aligned}h(x)&=f(x)\Big(1-\frac{\log(\ell(x))}{f(x)^\tau}\Big)^{1/\tau}\\&=x\Big(c_1^{-\tau}-\frac{b\log x+\log a}{x^\tau}\Big)^{1/\tau}\Big(1-\frac{\log(\ell(x))}{(x/c_1)^\tau-(b\log x+\log a)}\Big)^{1/\tau}\\&=:xL(x).\end{aligned}$$

Note that $L(x)\to 1/c_1$ as x tends to infinity, so that h is regularly varying at infinity with exponent 1. [4, Theorem 1.5.12] then tells us that there exists a slowly-varying function $\widetilde{L}$ such that

$$\lim_{x\to\infty}\widetilde{L}(x)L(x\widetilde{L}(x))=1, \tag{5.8}$$

which implies that $h^{\leftarrow}(x)\sim\widetilde{L}(x)x$ and that $\widetilde{L}(x)$ converges to c_1. Since $h^{\leftarrow}$ is increasing, we obtain

$$\max_{i\in[n]}\frac{W_i}{b_n(W)}=\frac{h^{\leftarrow}(\max_{i\in[n]}X_i)}{\widetilde{L}(\max_{i\in[n]}X_i)\max_{i\in[n]}X_i}\frac{\max_{i\in[n]}X_i}{b_n(X)}\frac{b_n(X)}{b_n(W)}\widetilde{L}(\max_{i\in[n]}X_i)\xrightarrow{a.s.}1,$$

since the maximum over X_i tends to infinity with n almost surely, $b_n(X)/b_n(W)\sim 1/c_1$, by (5.7) and (5.8) and the continuous mapping theorem. □

With this lemma at hand, we now investigate the maximum conditional mean in-degree of the WRG when the vertex-weights satisfy the **(Gumbel)**-**(RV)** sub-case.

Proposition 5.4. *Consider the WRG model as in Definition 2.1, suppose the vertex-weights satisfy the* **(Gumbel)**-**(RV)** *sub-case in Assumption 2.3 and recall the sets C_n from (2.20). Let $\gamma:=1/(\tau+1)$, let ℓ be a strictly positive function such that for some $\zeta_0\in[0,\infty)$,*

$$\lim_{n\to\infty}\frac{\log(\ell(n))^2}{\log n}=\zeta_0, \tag{5.9}$$

and let Π be a PPP on $(0,\infty)\times\mathbb{R}$ with intensity measure $\nu(\mathrm{d}t,\mathrm{d}x):=\mathrm{d}t\times\mathrm{e}^{-x}\mathrm{d}x$. Then, for any $0<s<t<\infty$, as $n\to\infty$,

$$\begin{aligned}&\Big(\max_{i\in[n]}\frac{\mathbb{E}_W[\mathcal{Z}_n(i)]}{m(1-\gamma)b_{n^\gamma}\log n},\frac{\log\widetilde{I}_n}{\log n}\Big)\xrightarrow{a.s.}(1,\gamma),\\&\max_{i\in C_n(\gamma,s,t,\ell)}\frac{\mathbb{E}_W[\mathcal{Z}_n(i)]-m(1-\gamma)b_{n^\gamma}\log n}{m(1-\gamma)a_{n^\gamma}\log n}\xrightarrow{d}\max_{\substack{(v,w)\in\Pi\\ v\in(s,t)}}w-\log v-\frac{\zeta_0(\tau+1)^2}{2\tau}.\end{aligned} \tag{5.10}$$

Furthermore,

$$\max_{i\in[n]}\frac{\mathbb{E}_W[\mathcal{Z}_n(i)]-m(1-\gamma)b_{n^\gamma}\log n}{m(1-\gamma)a_{n^\gamma}\log n\log\log n}\xrightarrow{\mathbb{P}}\frac{1}{2}. \tag{5.11}$$

Proof. We start by proving the first order growth rate of the maximum, as in the first line of (5.10). We can immediately construct the lower bound

$$\frac{\max_{i\in[n]}W_i\sum_{j=i}^{n-1}1/S_j}{(1-\gamma)b_{n^\gamma}\log n}\geq\frac{\max_{i\in[n^\gamma]}W_i\sum_{j=n^\gamma}^{n-1}1/S_j}{(1-\gamma)b_{n^\gamma}\log n}, \tag{5.12}$$

and the right-hand side converges almost surely to 1 by (5.2) and Lemma 5.3. For an upper bound, we first define the sequence $(\widetilde{\varepsilon}_k)_{k\in\mathbb{Z}_+}$ as

$$\widetilde{\varepsilon}_k=\frac{\gamma}{2}\Big(1-\Big(\frac{1-\gamma}{1-(\gamma-\widetilde{\varepsilon}_{k-1})}\Big)^\tau\Big)+\frac{1}{2}\widetilde{\varepsilon}_{k-1},\quad k\geq1,\qquad\widetilde{\varepsilon}_0=\gamma. \tag{5.13}$$

This sequence is defined in such a way that it is decreasing and tends to zero with k, and the maximum over indices i such that $n^{\gamma-\widetilde{\varepsilon}_{k-1}}\leq i\leq n^{\gamma-\widetilde{\varepsilon}_k}$ is almost surely bounded away from 1: For any $k\geq1$, we obtain the upper bound

$$\begin{aligned}\max_{i\in[n^{\gamma-\widetilde{\varepsilon}_k}]}\frac{W_i\sum_{j=i}^{n-1}1/S_j}{(1-\gamma)b_{n^\gamma}\log n}&=\max_{1\leq j\leq k}\max_{n^{\gamma-\widetilde{\varepsilon}_{j-1}}\leq i\leq n^{\gamma-\widetilde{\varepsilon}_j}}\frac{W_i\sum_{j=i}^{n-1}1/S_j}{(1-\gamma)b_{n^\gamma}\log n}\\&\leq\max_{1\leq j\leq k}\Big[\Big(\max_{i\in[n^{\gamma-\widetilde{\varepsilon}_j}]}\frac{W_i}{b_{n^{\gamma-\widetilde{\varepsilon}_j}}}\Big)\frac{\sum_{j=n^{\gamma-\widetilde{\varepsilon}_{j-1}}}^{n-1}1/S_j}{(1-\gamma)\log n}\frac{b_{n^{\gamma-\widetilde{\varepsilon}_j}}}{b_{n^\gamma}}\Big],\end{aligned} \tag{5.14}$$

which, using the asymptotics of b_n, (5.2) and Lemma 5.3 converges almost surely to

$$c_k:=\max_{1\leq j\leq k}\frac{1-(\gamma-\widetilde{\varepsilon}_{j-1})}{1-\gamma}\Big(\frac{\gamma-\widetilde{\varepsilon}_j}{\gamma}\Big)^{1/\tau}, \tag{5.15}$$

which is strictly smaller than one by the choice of the sequence $(\widetilde{\varepsilon}_k)_{k\geq0}$. Now, by writing, for some $\eta>0$ to be specified later,

$$E_n:=\Big\{\max_{i\in[n]}\frac{W_i\sum_{j=i}^{n-1}1/S_j}{(1-\gamma)b_{n^\gamma}\log n}\geq1-\eta\Big\},$$

which holds almost surely for all n large by (5.12), we obtain, for any $\varepsilon>0$,

$$\begin{aligned}\Big\{\frac{\log\widetilde{I}_n}{\log n}<\gamma-\varepsilon\Big\}&\subseteq\Big(\Big\{\frac{\log\widetilde{I}_n}{\log n}<\gamma-\varepsilon\Big\}\cap E_n\Big)\cup E_n^c\\&\subseteq\Big\{\max_{i<n^{\gamma-\varepsilon}}\frac{W_i\sum_{j=i}^{n-1}1/S_j}{(1-\gamma)b_{n^\gamma}\log n}\geq1-\eta\Big\}\cup E_n^c.\end{aligned}$$

The second event in the union on the right-hand side holds for finitely many n only, almost surely. For the first event in the union, we use (5.14) for a fixed k large enough such that $\widetilde{\varepsilon}_k<\varepsilon$ to obtain

$$\Big\{\max_{i<n^{\gamma-\varepsilon}}\frac{W_i\sum_{j=i}^{n-1}1/S_j}{(1-\gamma)b_{n^\gamma}\log n}\geq1-\eta\Big\}\subseteq\Big\{\max_{i<n^{\gamma-\widetilde{\varepsilon}_k}}\frac{W_i\sum_{j=i}^{n-1}1/S_j}{(1-\gamma)b_{n^\gamma}\log n}\geq1-\eta\Big\} \tag{5.16}$$

If we then choose η small enough such that

$$\begin{aligned}c_k&=\max_{1\leq j\leq k}2^{-1/\tau}\Big(\frac{(1-(\gamma-\widetilde{\varepsilon}_{j-1}))^\tau(\gamma-\widetilde{\varepsilon}_{j-1})}{(1-\gamma)^\tau\gamma}+1\Big)^{1/\tau}\\&=2^{-1/\tau}\Big(\frac{(1-(\gamma-\widetilde{\varepsilon}_{k-1}))^\tau(\gamma-\widetilde{\varepsilon}_{k-1})}{(1-\gamma)^\tau\gamma}+1\Big)^{1/\tau}<1-\eta,\end{aligned}$$

which is possible due to the fact that the expression on the left of the second line is increasing to 1 in k, we find that the event on the right-hand side of (5.16) holds for finitely many n only. Thus, almost surely, the event $\{\log(\widetilde{I}_n)/\log n<\gamma-\varepsilon\}$ holds for finitely many n only, irrespective of the value of ε. With a similar argument, and using a sequence $(\varepsilon_k)_{k\in\mathbb{Z}_+}$, defined as

$$\varepsilon_k = \frac{1-\gamma}{2}\Big(1-\Big(\frac{\gamma+\varepsilon_{k-1}}{\gamma}\Big)^{-1/\tau}\Big)+\frac{1}{2}\varepsilon_{k-1}, \quad k\geq 1, \qquad \varepsilon_0 = 1-\gamma,$$

we find that the maximum is not obtained at $n^{\gamma+\varepsilon}\leq i\leq n$ for any $\varepsilon>0$ almost surely as well, which proves the second part of the first line of (5.10). This also allows for a tighter upper bound of the maximum. On the event that the maximum is obtained at an index i such that $n^{\gamma-\varepsilon}\leq i\leq n^{\gamma+\varepsilon}$,

$$\frac{\max_{i\in[n]} W_i\sum_{j=i}^{n-1}1/S_j}{(1-\gamma)b_{n^\gamma}\log n} = \max_{n^{\gamma-\varepsilon}\leq i\leq n^{\gamma+\varepsilon}} \frac{W_i\sum_{j=i}^{n-1}1/S_j}{(1-\gamma)b_{n^\gamma}\log n} \leq \max_{i\in[n^{\gamma+\varepsilon}]}\frac{W_i}{b_{n^{\gamma+\varepsilon}}}\frac{\sum_{j=n^{\gamma-\varepsilon}}^{n-1}1/S_j}{(1-\gamma)\log n}\frac{b_{n^{\gamma+\varepsilon}}}{b_{n^\gamma}},$$

which, again using the asymptotics of b_n, (5.2) and Lemma 5.3 converges almost surely to $(1+\varepsilon/(1-\gamma))(1+\varepsilon/\gamma)^{1/\tau}$. This upper bound decreases to 1 as ε tends to zero, so that the upper bound can be chosen arbitrarily close to 1 by choosing ε sufficiently small. Hence, the left-hand side exceeds $1+\delta$, for any $\delta>0$, only finitely many times. As the event on which this upper bound is constructed holds almost surely eventually for all n, for any fixed $\varepsilon>0$, the first part of the first line of (5.10) follows.

Restricted second-order fluctuations. We now turn to the second line of (5.10), which deals with the second order growth rate of the maximum conditional mean with indices in a specific compact range. For ease of writing, we omit the arguments and write $C_n := C_n(\beta,s,t,\ell)$.

We use results from extreme value theory regarding the convergence of particular point processes to obtain the results. Let us define the point process

$$\Pi_n := \sum_{i=1}^n \delta_{(i/n,(W_i-b_n)/a_n)}.$$

By [25, Corollary 4.19], when the W_i are i.i.d. random variables in the Gumbel maximum domain of attraction (which is the case for the **(Gumbel)-(RV)** sub-case), then Π_n has a weak limit Π in the vague topology, a PPP on $(0,\infty)\times(-\infty,\infty]$ with intensity measure $\nu(\mathrm{d}t,\mathrm{d}x)=\mathrm{d}t\times \mathrm{e}^{-x}\mathrm{d}x$. Here, we understand the topology on $(-\infty,\infty]$ such that sets of the form $[a,\infty]$ for $a\in\mathbb{R}$ are compact and we are crucially using that the measure $\mathrm{e}^{-x}\mathrm{d}x$ is finite on these compact sets.

Rather than considering the time-scale n and all $i\in[n]$, we consider the time-scale $\ell(n)n^\gamma$ and $i\in C_n$, and show that the rescaled conditional mean in-degrees can be written as a continuous functional of the point process $\Pi_{\ell(n)n^\gamma}$ with vanishing error terms. Thus, we write

$$\begin{aligned}\frac{W_i\sum_{j=i}^{n-1}1/S_j - b_{\ell(n)n^\gamma}\log(n^{1-\gamma}/\ell(n))}{a_{\ell(n)n^\gamma}\log(n^{1-\gamma}/\ell(n))} &= \frac{W_i-b_{\ell(n)n^\gamma}}{a_{\ell(n)n^\gamma}}\frac{\sum_{j=i}^{n-1}1/S_j}{\log(n^{1-\gamma}/\ell(n))}-\log\Big(\frac{i}{\ell(n)n^\gamma}\Big)\\ &\quad+\frac{b_{\ell(n)n^\gamma}}{a_{\ell(n)n^\gamma}\log(n^{1-\gamma}/\ell(n))}\Big(\sum_{j=i}^{n-1}1/S_j-\log(n/i)\Big)\\ &\quad-\Big(\frac{b_{\ell(n)n^\gamma}}{a_{\ell(n)n^\gamma}\log(n^{1-\gamma}/\ell(n))}-\Big)\log(i/\ell(n)n^\gamma).\end{aligned}$$

We then let, for $0<s<t<\infty, f\in\mathbb{R}$,

$$\widetilde{C}_n(f) := \{i\in C_n : (W_i-b_{\ell(n)n^\gamma})/a_{\ell(n)n^\gamma}\geq f\}. \tag{5.17}$$

Then, for C_n (as well as $\widetilde{C}_n(f)$), we can bound

$$\Big|\max_{i\in C_n}\frac{W_i\sum_{j=i}^{n-1}1/S_j-b_{\ell(n)n^\gamma}\log(n^{1-\gamma}/\ell(n))}{a_{\ell(n)n^\gamma}\log(n^{1-\gamma}/\ell(n))} \\ -\max_{i\in C_n}\Big(\frac{(W_i-b_{\ell(n)n^\gamma})\sum_{j=i}^{n-1}1/S_j}{a_{\ell(n)n^\gamma}\log(n^{1-\gamma}/\ell(n))}-\log\Big(\frac{i}{\ell(n)n^\gamma}\Big)\Big)\Big|$$

from above by

$$\frac{b_{\ell(n)n^\gamma}}{a_{\ell(n)n^\gamma}\log(n^{1-\gamma}/\ell(n))}\max_{i\in C_n}\Big|\sum_{j=i}^{n-1}1/S_j-\log(n/i)\Big| \\ +\Big|\frac{b_{\ell(n)n^\gamma}}{a_{\ell(n)n^\gamma}\log(n^{1-\gamma}/\ell(n))}-1\Big|\max_{i\in C_n}|\log(i/\ell(n)n^\gamma)|. \tag{5.18}$$

Since $\lim_{n\to\infty}\log(\ell(n))/\log n=0$ by (5.9), it immediately follows that $\ell(n)n^\gamma=n^{\gamma+o(1)}$. Since $b_n=g(\log n), a_n=\widetilde{g}(\log n)$ with $g,\widetilde{g}$ regularly-varying (see Remark 2.4), it follows that $b_{\ell(n)n^\gamma}\sim b_{n^\gamma}$, $a_{\ell(n)n^\gamma}\sim a_{n^\gamma}$, $\log(n^{1-\gamma}/\ell(n))\sim(1-\gamma)\log n$, so that the first fraction on the first line and the term on the second line in absolute value tend to one and zero, respectively. It also follows from (5.2) that $\sum_{j=i}^{n-1}1/S_j-\log(n/i)$ converges almost surely for any fixed $i\in\mathbb{N}$, so the maximum on the second line tends to zero almost surely, as the sequence in the absolute value is a Cauchy sequence almost surely (and all $i\in C_n$ tend to infinity with n). Finally, we can bound the maximum on the last line by $\max\{|\log t|,|\log s|\}$, so that the entire expression in (5.18) converges to zero almost surely. As a result, we obtain

$$\Big|\max_{i\in C_n}\frac{W_i\sum_{j=i}^{n-1}1/S_j-b_{\ell(n)n^\gamma}\log(n^{1-\gamma}/\ell(n))}{a_{\ell(n)n^\gamma}\log(n^{1-\gamma}/\ell(n))} \\ -\max_{i\in C_n}\Big[\frac{(W_i-b_{\ell(n)n^\gamma})}{a_{\ell(n)n^\gamma}}\frac{\sum_{j=i}^{n-1}1/S_j}{\log(n^{1-\gamma}/\ell(n))}-\log\Big(\frac{i}{\ell(n)n^\gamma}\Big)\Big]\Big|\xrightarrow{a.s.}0. \tag{5.19}$$

From the weak convergence of $\Pi_n\xrightarrow{d}\Pi$ in the vague topology, it follows that

$$\sum_{i\in\widetilde{C}_n}\delta\Big(i/(\ell(n)n^\gamma),\frac{W_i-b_{\ell(n)n^\gamma}}{a_{\ell(n)n^\gamma}}\Big)\Rightarrow\sum_{\substack{(v,w)\in\Pi\\ v\in[s,t],w\geq f}}\delta((v,w)),$$

on the space of point measures equipped with the vague topology, where $\delta(\cdot)$ is a Dirac measure. It is straight-forward to extend this convergence (using that $[s,t]\times[f,\infty]$ is a compact set) to show that

$$\sum_{i\in\widetilde{C}_n}\delta\Big(i/(\ell(n)n^\gamma),\frac{W_i-b_{\ell(n)n^\gamma}}{a_{\ell(n)n^\gamma}},\frac{\sum_{j=i}^{n-1}1/S_j}{\log(n^{1-\gamma}/\ell(n))}\Big)\Rightarrow\sum_{\substack{(v,w)\in\Pi\\ v\in[s,t],w\geq f}}\delta((v,w,1)).$$

Hence, combining this with the continuous mapping theorem and (5.19) yields

$$\max_{i\in\widetilde{C}_n}\frac{W_i-b_{\ell(n)n^\gamma}}{a_{\ell(n)n^\gamma}}\frac{\sum_{j=i}^{n-1}1/S_j}{\log(n^{1-\gamma}/\ell(n))}-\log\Big(\frac{i}{\ell(n)n^\gamma}\Big)\xrightarrow{d}\max_{\substack{(v,w)\in\Pi\\ v\in[s,t],w\geq f}}\Big(w-\log v\Big), \tag{5.20}$$

as element-wise multiplication and taking the maximum of a finite number of elements is a continuous operation (which uses that by [25, Proposition 3.13] vague convergence on a compact set is the same as pointwise convergence). Now, we intend to show that the same result holds when considering $i\in C_n$, that is, the distributional convergence still holds when omitting the constraint on the size of the W_i (see (5.17)). Let $\eta>0$ be fixed, and for any closed $D\subset\mathbb{R}$, let let $D_\eta:=\{x\in\mathbb{R}\,|\,\inf_{y\in D}|x-y|\leq\eta\}$ be its η-enlargement. We define the random variables and events

$$\begin{aligned}X_{n,i}&:=\frac{W_i-b_{\ell(n)n^\gamma}}{a_{\ell(n)n^\gamma}}\frac{\sum_{j=i}^{n-1}1/S_j}{\log(n^{1-\gamma}/\ell(n))}-\log(i/(\ell(n)n^\gamma)),\quad i\in[n],\\ E_n(\eta)&:=\{|\max_{i\in C_n}X_{n,i}-\max_{i\in\widetilde{C}_n(f)}X_{n,i}|<\eta\},\qquad A_n(\eta):=\{\max_{i\in C_n}X_{n,i}\in D_\eta\},\end{aligned}$$

and note that $D_0=D$, by definition of D_η. The aim is to show that

$$\limsup_{n\to\infty}\mathbb{P}(A_n(0))=\limsup_{n\to\infty}\mathbb{P}\Big(\max_{i\in C_n}X_{n,i}\in D\Big)\leq\mathbb{P}\Big(\max_{\substack{(v,w)\in\Pi\\ v\in[s,t]}}(w-\log v)\in D\Big),\tag{5.21}$$

as the Portmanteau lemma then yields

$$\max_{i\in C_n(\gamma,s,t,\ell(n))}\frac{W_i-b_{\ell(n)n^\gamma}}{a_{\ell(n)n^\gamma}}\frac{\sum_{j=i}^{n-1}1/S_j}{\log(n^{1-\gamma}/\ell(n))}-\log(i/(\ell(n)n^\gamma))\xrightarrow{d}\max_{\substack{(v,w)\in\Pi\\ v\in[s,t]}}w-\log v.\tag{5.22}$$

To achieve (5.21), we first use the events $E_n(\eta)$ to bound

$$\mathbb{P}(A_n(0))\leq\mathbb{P}(A_n(0)\cap E_n(\eta))+\mathbb{P}(E_n(\eta)^c)\,,\tag{5.23}$$

and then show that the second probability on the right-hand side converges to zero and the first probability on the right-hand side converges to the right-hand side of (5.21) when $n\to\infty$ and $\eta\downarrow 0$.

We first deal with the first probability on the right-hand side. Note that this probability can be bounded from above by

$$\mathbb{P}(A_n(0)\cap E_n(\eta))\leq\mathbb{P}\Big(\max_{i\in\widetilde{C}_n(f)}X_{n,i}\in D_\eta\Big),$$

due to the definition of the event $E_n(\eta)$. From (5.20) we then obtain that

$$\limsup_{n\to\infty}\mathbb{P}\Big(\max_{i\in\widetilde{C}_n}X_{n,i}\in D_\eta\Big)=\mathbb{P}\Big(\max_{\substack{(v,w)\in\Pi\\ v\in[s,t],w\geq f}}w-\log v\in D_\eta\Big).\tag{5.24}$$

What remains is to remove the restriction that $w\geq f$ of the maximum. Note that

$$\begin{aligned}\Big|\max_{\substack{(v,w)\in\Pi\\ v\in[s,t]}}\Big(w-\log v\Big)-\max_{\substack{(v,w)\in\Pi\\ v\in[s,t],w\geq f}}\Big(w-\log v\Big)\Big|&\leq\max\Big\{0,\max_{\substack{(v,w)\in\Pi\\ v\in[s,t],w\leq f}}w-\log v\Big\}\\ &\leq\max\Big\{0,f-\log s\Big\},\end{aligned}\tag{5.25}$$

which tends to zero almost surely when $f\to-\infty$. Hence, using (5.24) and (5.25), we obtain

$$\lim_{f\to-\infty}\limsup_{n\to\infty}\mathbb{P}\Big(\max_{i\in\widetilde{C}_n}X_{n,i}\in D_\eta\Big)\leq\mathbb{P}\Big(\max_{\substack{(v,w)\in\Pi\\ v\in[s,t]}}(w-\log v)\in D_\eta\Big).\tag{5.26}$$

We then show that the second probability on the right-hand side of (5.23) tends to zero. We bound

$$\Big|\max_{i\in C_n}X_{n,i}-\max_{i\in\widetilde{C}_n(f)}X_{n,i}\Big|\leq\max\Big\{0,\max_{i\in C_n\backslash\widetilde{C}_n(f)}X_{n,i}\Big\}.\tag{5.27}$$

As we intend to let f go to $-\infty$, we can assume $f<0$. Then, as $(W_i-b_{\ell(n)n^\gamma})/a_{\ell(n)n^\gamma}<f<0$ for all $i\in C_n\backslash\widetilde{C}_n(f)$, we obtain the upper bound

$$\Big|\max_{i\in C_n}X_{n,i}-\max_{i\in\widetilde{C}_n(f)}X_{n,i}\Big|\leq\max\Big\{0,f\frac{\sum_{j=t\ell(n)n^\gamma}^{n-1}1/S_j}{\log(n^{1-\gamma}/\ell(n))}-\log s\Big\},$$

and the right-hand side converges almost surely to $\max\{0,f-\log s\}$ as n tends to infinity. Then, as f tends to $-\infty$, this maximum tends to zero. So, the absolute value on the left-hand side of (5.27) tends to zero in probability, and therefore the second probability on the right-hand side of (5.23) converges to zero as $n\to\infty$, then $f\to-\infty$ for any $\eta>0$. Combining this with (5.26) and (5.23), we obtain

$$\limsup_{n\to\infty}\mathbb{P}(A_n(0))\leq \mathbb{P}\Big(\max_{\substack{(v,w)\in\Pi\\ v\in[s,t]}}(w-\log v)\in D_\eta\Big).$$

We can then take the limit $\eta\downarrow 0$ on the right-hand side, since the properties of the Poisson point process imply that the maximum does not hit the boundary of D almost surely and by the continuity of the probability measure $\mathbb{P}$. We thus arrive at (5.21) which then implies (5.22) via the Portmanteau lemma. Together with (5.19) and Slutsky's theorem, it follows that

$$\max_{i\in C_n(\gamma,s,t,\widetilde{\ell}(n))}\frac{W_i\sum_{j=i}^{n-1}1/S_j-b_{\ell(n)n^\gamma}\log(n^{1-\gamma}/\ell(n))}{a_{\ell(n)n^\gamma}\log(n^{1-\gamma}/\ell(n))}\xrightarrow{d}\max_{\substack{(v,w)\in\Pi\\ v\in(s,t)}}w-\log v,$$

so that the same results hold for the re-scaled maximum conditional mean degree by (5.1). What remains is to show that the same result is obtained when $\ell(n)n^\gamma$ is replaced with n^γ in the first and second order rescaling, from which the second line of (5.10) follows. To obtain this, we show that

$$\begin{aligned}\lim_{n\to\infty}\frac{(1-\gamma)a_{n^\gamma}\log n}{a_{\ell(n)n^\gamma}\log(n^{1-\gamma}/\ell(n))}&=1,\\ \lim_{n\to\infty}\frac{b_{\ell(n)n^\gamma}\log(n^{1-\gamma}/\ell(n))-(1-\gamma)b_{n^\gamma}\log n}{(1-\gamma)a_{n^\gamma}\log n}&=-\frac{\zeta_0(\tau+1)^2}{2\tau},\end{aligned}\tag{5.28}$$

after which the Convergence-to-Types theorem [25, Proposition 0.2] yields the required result.

we now prove (5.28). First, it immediately follows from Remark 2.4 that

$$\frac{a_{\ell(n)n^\gamma}\log(n^{1-\gamma}/\ell(n))}{(1-\gamma)a_{n^\gamma}\log n}=\Big(1+\frac{\log(\ell(n))}{\gamma\log n}\Big)^{1/\tau-1}\Big(1-\frac{\log(\ell(n))}{(1-\gamma)\log n}\Big)\to 1,$$

since we assume that $\log(\ell(n))^2/\log n\to\zeta_0$, so that the first condition in (5.28) is satisfied. Then,

$$\begin{aligned}&b_{\ell(n)n^\gamma}-b_{n^\gamma}\\ &\quad=c_1(\gamma\log n)^{1/\tau}\Big[\Big(1+\frac{\log(\ell(n))}{\gamma\log n}\Big)^{1/\tau}-1\Big]\\ &\qquad+\frac{c_1}{\tau}(\gamma\log n)^{1/\tau-1}\Big[\Big(1+\frac{\log(\ell(n))}{\gamma\log n}\Big)^{1/\tau-1}-1\Big]\Big(\frac{b}{\tau}\log(\gamma\log n)+b\log c_1+\log a\Big)\\ &\qquad+\frac{bc_1}{\tau^2}(\gamma\log n)^{1/\tau-1}\Big(1+\frac{\log(\ell(n))}{\gamma\log n}\Big)^{1/\tau-1}\log\Big(1+\frac{\log(\ell(n))}{\gamma\log n}\Big).\end{aligned}$$

Also,

$$\begin{aligned}b_{\ell(n)n^\gamma}\log(\ell(n))&=c_1(\gamma\log n)^{1/\tau}\log(\ell(n))\Big(1+\frac{\log(\ell(n))}{\gamma\log n}\Big)^{1/\tau}\\ &\quad+\frac{c_1}{\tau}(\gamma\log n)^{1/\tau-1}\log(\ell(n))\Big(1+\frac{\log(\ell(n))}{\gamma\log n}\Big)^{1/\tau-1}\Big[\frac{b}{\tau}\log(\gamma\log n)\\ &\quad+b\log c_1+\frac{b}{\tau}\log\Big(1+\frac{\log(\ell(n))}{\gamma\log n}\Big)+\log a\Big].\end{aligned}$$

Using Taylor expansions for the terms containing $1 + \log(\ell(n))/(\gamma \log n)$ in both these expressions and combining them, yields

$$\begin{aligned}
&b_{\ell(n)n^\gamma} \log(n^{1-\gamma}/\ell(n)) - b_{n^\gamma}(1-\gamma)\log n \\
&\quad = (b_{\ell(n)n^\gamma} - b_{n^\gamma})(1-\gamma)\log n - b_{\ell(n)n^\gamma}\log(\ell(n)) \\
&\quad = -\frac{c_1(\tau+1)}{2\tau}(\gamma\log n)^{1/\tau-1}\log(\ell(n))^2 + \frac{c_1 b}{\tau}(\gamma \log n)^{1/\tau-1}\log(\ell(n)) \\
&\qquad - c_1\Big(\frac{b}{\tau}\log(\gamma\log n) + b\log c_1 + \log a\Big)(\gamma\log n)^{1/\tau-1}\log(\ell(n)) + x_n,
\end{aligned}$$

where x_n consists of lower order terms such that $x_n = o((\log n)^{1/\tau-1}\log(\ell(n)))$. Thus, we obtain

$$\begin{aligned}
\frac{b_{\ell(n)n^\gamma}\log(n^{1-\gamma}/\ell(n)) - (1-\gamma)b_{n^\gamma}\log n}{(1-\gamma)a_{n^\gamma}\log n} &\sim -\frac{((\tau+1)\log(\ell(n)))^2}{2\tau\log n} + \frac{b(\tau+1)}{\tau}\frac{\log(\ell(n))}{\log n} \\
&\quad - (\tau+1)\Big[\frac{b}{\tau}\log(\gamma\log n) + \log\big(ac_1^b\big)\Big]\frac{\log(\ell(n))}{\log n}.
\end{aligned}$$

Since $(\log \ell(n))^2/\log n$ converges to $\zeta_0 \in [0,\infty)$, it follows that the second condition in (5.28) is indeed satisfied.

Unrestricted second order fluctuations. We finally prove (5.11), which describes the second order fluctuations when the index set is allowed to range the full set $[n]$. We first remark that it suffices to study $\max_{i\in[n]} W_i \log(n/i)$ rather than $\max_{i\in[n]} \mathbb{E}_W[\mathcal{Z}_n(i)]$. This is due to the following:

$$\begin{aligned}
\frac{|\max_{i\in[n]}\mathbb{E}_W[\mathcal{Z}_n(i)] - \max_{i\in[n]} W_i\log(n/i)|}{a_n\log n\log\log n} &\le \frac{\max_{i\in[n]} W_i}{a_n\log n\log\log n}\Big|\sum_{j=i}^{n-1} 1/S_j - \log(n/i)\Big| \\
&= \frac{1}{a_n\log n\log\log n}\max_{i\in[n]} W_i|Y_n - Y_i|,
\end{aligned}$$

where $Y_n := \sum_{j=1}^{n-1} 1/S_j - \log n$. By (5.2), Y_n converges almost surely to Y, which is almost surely finite as well. Hence, $\sup_{i\in\mathbb{N}} Y_i$ is almost surely finite, too. This yields the upper bound

$$\frac{\max_{i\in[n]} W_i}{a_n\log n}\frac{|Y_n| + \sup_{i\in\mathbb{N}}|Y_i|}{\log\log n}.$$

The first fraction converges almost surely by Lemma 5.3 and as $a_n \log n \sim b_n/\tau$, and the second fraction converges to zero almost surely. We thus find that

$$\frac{1}{a_n\log n\log\log n}|\max_{i\in[n]}\mathbb{E}_W[\mathcal{Z}_n(i)] - \max_{i\in[n]} W_i\log(n/i)| \xrightarrow{a.s.} 0, \tag{5.29}$$

and thus it suffices to prove that

$$\max_{i\in[n]}\frac{W_i\log(n/i) - (1-\gamma)b_{n^\gamma}\log n}{(1-\gamma)a_{n^\gamma}\log n\log\log n} \xrightarrow{\mathbb{P}} \frac{1}{2}. \tag{5.30}$$

Therefore, we set for $i \in [n]$.

$$X_{n,i} := \frac{W_i\log(n/i) - (1-\gamma)b_{n^\gamma}\log n}{(1-\gamma)a_{n^\gamma}\log n\log\log n}.$$

For an upper bound on the maximum of the $X_{n,i}$, we consider different ranges of indices i separately. We concentrate on the range $i \ge n^\gamma$, the case $i \le n^\gamma$ follows by completely analogous arguments. For a lower bound on the maximum, we choose a convenient range of indices.

Let $\varepsilon \in (0, 1-\gamma)$. First of all, we notice that by the same argument as in the proof of the first line of (5.10), there exists a constant $C < 1$ (which is similar to the constant in (5.15)) such that almost surely

$$\max_{n^{\gamma+\varepsilon}<i\leq n} W_i \log(n/i) \leq C(1-\gamma)b_{n^\gamma}\log n.$$

It then follows that the rescaled maximum diverges to $-\infty$ almost surely.

As the next step, we consider the range of $n^\gamma \leq i \leq \mathrm{e}^{k_n}n^\gamma$, where $k_n = \sqrt{\log n}\log\log n$. This turns out to give the main contribution to the maximum of the $X_{n,i}$. We now fix $x > 1/2$ and let $\delta > 0$. Then,

$$\begin{aligned}
&\mathbb{P}\Big(\max_{n^\gamma\leq i\leq \mathrm{e}^{k_n}n^\gamma} \frac{W_i\log(n/i)-(1-\gamma)b_{n^\gamma}\log n}{(1-\gamma)a_{n^\gamma}\log n\log\log n}\leq x\Big)\\
&=\prod_{i=n^\gamma}^{\mathrm{e}^{k_n}n^\gamma}\mathbb{P}\Big(W_i\leq \frac{1-\gamma}{1-\log i/\log n}(b_{n^\gamma}+a_{n^\gamma}x\log\log n)\Big)\\
&\geq \exp\Big\{-(1+\delta)\sum_{i=n^\gamma}^{\mathrm{e}^{k_n}n^\gamma}\mathbb{P}\Big(W\geq \frac{1-\gamma}{1-\log i/\log n}(b_{n^\gamma}+a_{n^\gamma}x\log\log n)\Big)\Big\}\\
&\geq \exp\Big\{-(1+\delta)\sum_{j=1}^{k_n}\sum_{i=\mathrm{e}^{j-1}n^\gamma}^{\mathrm{e}^{j}n^\gamma}\mathbb{P}\Big(W\geq \frac{1-\gamma}{1-\gamma-(j-1)/\log n}(b_{n^\gamma}+a_{n^\gamma}x\log\log n)\Big)\Big\},
\end{aligned}\tag{5.31}$$

where we use that $1-y\geq \mathrm{e}^{-(1+\delta)y}$ for all y sufficiently small and that the tail probability is decreasing to zero, uniformly in i, in the last two steps. Since the probability is no longer dependent on i, we can also omit the inner sum and replace it by $\lfloor \mathrm{e}^j n^\gamma\rfloor - \lceil \mathrm{e}^{j-1}n^\gamma\rceil \leq (\mathrm{e}-1)\mathrm{e}^{j-1}n^\gamma$. Also using that $\mathbb{P}(W\geq y)\leq (1+\delta)ay^b\exp\{-(y/c_1)^\tau\}$ for all y sufficiently large, it follows that for any $x\in\mathbb{R}$ and n sufficiently large we obtain the lower bound

$$\begin{aligned}
\exp\Big\{&-(1+\delta)^2a(\mathrm{e}-1)\sum_{j=1}^{k_n}\mathrm{e}^{j-1}n^\gamma\Big(\frac{1-\gamma}{1-\gamma-(j-1)/\log n}(b_{n^\gamma}+a_{n^\gamma}x\log\log n)\Big)^b\\
&\times\exp\Big\{-\Big(\frac{1}{c_1}\frac{1-\gamma}{1-\gamma-(j-1)/\log n}(b_{n^\gamma}+a_{n^\gamma}x\log\log n)\Big)^\tau\Big\}\Big\}.
\end{aligned}\tag{5.32}$$

We first bound the fraction $(1-\gamma)/(1-\gamma)-(j-1)/\log n)$ from above by $1+\delta$ if $b\geq 0$ and from below by 1 if $b<0$, which holds uniformly in j for n large, in the outer exponent. Then, when we combine all other terms that contain j, we find

$$\exp\Big\{(j-1)-\Big(\frac{(1-\gamma)\log n}{(1-\gamma)\log n-(j-1)}\Big)^\tau\Big(\frac{b_{n^\gamma}+a_{n^\gamma}x\log\log n}{c_1}\Big)^\tau\Big\}.\tag{5.33}$$

Define $\mathrm{e}_n := (1-\gamma)\log n$, then we have by a Taylor expansion that there exists a constant $C_\tau>0$ such that uniformly for $|y|\leq \mathrm{e}_n/2$, we have that

$$1+\tau\frac{y}{\mathrm{e}_n}+\frac{\tau(1+\tau)}{2}\Big(\frac{y}{\mathrm{e}_n}\Big)^2\leq\Big(\frac{\mathrm{e}_n}{\mathrm{e}_n-y}\Big)^\tau\leq 1+\tau\frac{y}{\mathrm{e}_n}+C_\tau\Big(\frac{y}{\mathrm{e}_n}\Big)^2.\tag{5.34}$$

We also need that again by a Taylor expansion and the explicit form of a_n, b_n as stated in Remark 2.4, we have that

$$\begin{aligned}
\Big(\frac{b_{n^\gamma}+a_{n^\gamma}x\log\log n}{c_1}\Big)^\tau &= (b_{n^\gamma}/c_1)^\tau\Big(1+\tau\frac{a_{n^\gamma}}{b_{n^\gamma}}x\log\log n(1+o(1))\Big)\\
&= (b_{n^\gamma}/c_1)^\tau + x\log\log n(1+o(1)),
\end{aligned}$$

and similarly since $\gamma = 1/(1+\tau)$,

$$\frac{(b_{n^\gamma}/c_1)^\tau}{\mathrm{e}_n}=\frac{1}{(1-\gamma)}\frac{\log n^\gamma}{\log n}+\mathcal{O}\Big(\frac{\log\log n}{\log n}\Big)=\frac{1}{\tau}+\mathcal{O}\Big(\frac{\log\log n}{\log n}\Big).$$

Combining all these estimates, we obtain the following upper bound on (5.33):

$$\begin{aligned}
&\exp\Big\{j-1-\Big(1+\tau\frac{j-1}{\mathrm{e}_n}+\frac{\tau(1+\tau)}{2}\Big(\frac{j-1}{\mathrm{e}_n}\Big)^2\Big)((b_{n^\gamma}/c_1)^\tau+x\log\log n(1+o(1)))\Big\}\\
&\quad=\exp\Big\{-(b_{n^\gamma}/c_1)^\tau-x(\log\log n)(1+o(1))\\
&\qquad\qquad-(j-1)\,\mathcal{O}\Big(\frac{\log\log n}{\log n}\Big)-\frac{(1+\tau)(j-1)^2}{2(1-\gamma)\log n}(1+o(1))\Big\}\\
&\quad\leq\exp\big\{-(b_{n^\gamma}/c_1)^\tau-x(\log\log n)(1+o(1))+o(1)\big\},
\end{aligned}\tag{5.35}$$

where we used in the last step that $j\leq k_n=o(\log n/\log\log n)$ and the last term in the exponent is negative. Hence, combining this with (5.31) and (5.32), we obtain

$$\begin{aligned}
&\mathbb{P}\Big(\max_{n^\gamma\leq i\leq \mathrm{e}^{k_n}n^\gamma}\frac{W_i\log(n/i)-(1-\gamma)b_{n^\gamma}\log n}{(1-\gamma)a_{n^\gamma}\log n\log\log n}\leq x\Big)\\
&\quad\geq\exp\big\{-(1+\delta)^{2+b\vee 0}a(\mathrm{e}-1)k_nn^\gamma(b_{n^\gamma})^b\exp\{-(b_{n^\gamma}/c_1)^\tau-x(\log\log n)(1+o(1))\}\\
&\quad=\exp\big\{-(1+\delta)^{2+b\vee 0}(\mathrm{e}-1)k_n(\log n)^{-x(1+o(1))}\big\},
\end{aligned}$$

where we used that $n\mathbb{P}(W>b_n)\sim anb_n^b\mathrm{e}^{-(b_n/c_1)^\tau}\sim 1$ (see e.g. [25, Equation (1.1')] with $x=0$). Hence, by our choice of $k_n=\sqrt{\log n}\log\log n$ and $x>1/2$, the latter expression converges to 1 and we have shown that for any $\eta>0$, with high probability,

$$\max_{n^\gamma\leq i\leq \mathrm{e}^{k_n}n^\gamma}X_{n,i}\leq\frac{1}{2}+\eta.$$

Next, we consider an upper bound on the maximum for the range $\mathrm{e}^{k_n}n^\gamma\leq i\leq n^{\gamma+\varepsilon}$, where $\varepsilon\in(0,1-\gamma)$ and $k_n=\sqrt{\log n}\log\log n$ are as above. Again, we take $x\in\mathbb{R}$ and use the same idea as in the first step. This time, however, we need to be more careful in the intermediate step (5.35). For $k_n\leq j\leq\varepsilon\log n$, we obtain an upper bound on the expression in (5.33)

$$\begin{aligned}
&\exp\Big\{(j-1)-\Big(\frac{(1-\gamma)\log n}{(1-\gamma)\log n-(j-1)}\Big)^\tau\Big(\frac{b_{n^\gamma}+a_{n^\gamma}x\log\log n}{c_1}\Big)^\tau\Big\}\\
&\quad\leq\exp\Big\{-(b_{n^\gamma}/c_1)^\tau-x(\log\log n)(1+o(1))\\
&\qquad\qquad-(j-1)\mathcal{O}\Big(\frac{\log\log n}{\log n}\Big)-\frac{(1+\tau)(j-1)^2}{2(1-\gamma)\log n}(1+o(1))\Big\}\\
&\quad\leq\exp\Big\{-(b_{n^\gamma}/c_1)^\tau-x(\log\log n)(1+o(1))+C_1(j-1)\frac{\log\log n}{\log n}-C_2(j-1)^2\frac{1}{\log n}\Big\},
\end{aligned}$$

for suitable constants $C_1,C_2>0$. We now note that the right-hand side is decreasing in j as long as $j>(C_1)/(2C_2)\log\log n$. However, $k_n\gg\log\log n$, so that we obtain the following upper bound on the previous display that holds uniformly for $k_n\leq j\leq\varepsilon\log n$,

$$\exp\Big\{-(b_{n^\gamma}/c_1)^\tau-x(\log\log n)(1+o(1))+C_1k_n\frac{\log\log n}{\log n}-C_2k_n^2\frac{1}{\log n}\Big\}.$$

Using this bound in the same way as before (following the analogous steps as in (5.31) and (5.32)) we obtain for any $x\in\mathbb{R}$,

$$\begin{aligned}
&\mathbb{P}\Big(\max_{\mathrm{e}^{k_n}n^\gamma\leq i\leq n^{\gamma+\varepsilon}}\frac{W_i\log(n/i)-(1-\gamma)b_{n^\gamma}\log n}{(1-\gamma)a_{n^\gamma}\log n\log\log n}\leq x\Big)\\
&\quad\geq\exp\Big\{-(1+\delta)^{2+b\vee 0}(\mathrm{e}-1)\varepsilon(\log n)^{1-x(1+o(1))}\exp\Big\{C_1k_n\frac{\log\log n}{\log n}-C_2k_n^2\frac{1}{\log n}\Big\}\Big\}
\end{aligned}$$

Finally, since $k_n=\sqrt{\log n}\log\log n$, the right-hand side converges to 1. Therefore, we have shown that for any $x\in\mathbb{R}$, with high probability

$$\max_{\mathrm{e}^{k_n}n^\gamma\leq i\leq n^{\gamma+\varepsilon}}X_{n,i}\leq x.$$

In a similar way to the upper bound, we can construct a lower bound on the maximum by restricting to the indices $1 \leq i \leq k_n'$, where $k_n' = \sqrt{\log n}/(\log\log n)$. In this case, we consider $x < 1/2$ and use k_n' instead of k_n in the argument above. We omit the $(1+\delta)$ term in (5.31), bound the probability from below using $(1-\delta)$ rather than $(1+\delta)$, use the upper bound in (5.34) and obtain thus

$$\begin{aligned}\mathbb{P}\Big(&\max_{n^\gamma \leq i \leq \mathrm{e}^{k_n'} n^\gamma} \frac{W_i \log(n/i) - (1-\gamma) b_{n^\gamma} \log n}{(1-\gamma) a_{n^\gamma} \log n \log\log n} \leq x\Big)\\ &\leq \exp\Big\{-(1-\delta)^{1+b\wedge 0}(\mathrm{e}-1)k_n'(\log n)^{-x(1+o(1))}\exp\Big\{-\frac{C_\tau}{2\tau(1-\gamma)}\frac{(k_n')^2}{\log n}(1+o(1))\Big\}\Big\}.\end{aligned}$$

The latter term converges to zero as $x < 1/2$ and therefore, we have shown that for any $\eta > 0$, with high probability

$$\max_{i\in[n]} X_{n,i} \geq \max_{1\leq i \leq \mathrm{e}^{k_n'} n^\gamma} X_{n,i} \geq \frac{1}{2} - \eta.$$

This completes the argument for all $n^\gamma \leq i \leq n$. The argument for $1 \leq i \leq n^\gamma$ works completely analogously, so that we have shown (5.30), which completes the proof. $\square$

Proposition 5.5. *Consider the WRG model as in Definition 2.1 and suppose the vertex-weights satisfy the* **(Gumbel)**-**(RaV)** *sub-case in Assumption 2.3 and let* $t_n := \exp\{-\tau \log n/\log(b_n)\}$. *Then,*

$$\Big(\max_{i\in[n]} \frac{\mathbb{E}_W[\mathcal{Z}_n(i)]}{m b_{t_n n}\log(1/t_n)}, \frac{\log \widetilde{I}_n}{\log n}\Big) \xrightarrow{\mathbb{P}} (1,1). \tag{5.36}$$

Moreover, recall C_n *from* (2.20) *and let* Π *be a PPP on* $(0,\infty)\times\mathbb{R}$ *with intensity measure* $\nu(\mathrm{d}t,\mathrm{d}x) := \mathrm{d}t \times \mathrm{e}^{-x}\mathrm{d}x$. *Then, for any* $0 < s < t < \infty$,

$$\max_{i\in C_n(1,s,t,t_n)} \frac{\mathbb{E}_W[\mathcal{Z}_n(i)] - m b_{t_n n}\log(1/t_n)}{m a_{t_n n}\log(1/t_n)} \xrightarrow{d} \max_{\substack{(v,w)\in\Pi\\ v\in(s,t)}} w - \log v, \tag{5.37}$$

and, when $\tau \in (1,3]$,

$$\max_{i\in[n]} \frac{\mathbb{E}_W[\mathcal{Z}_n(i)] - m b_{t_n n}\log(1/t_n)}{m a_{t_n n}\log(1/t_n)\log\log n} \xrightarrow{\mathbb{P}} \frac{1}{2}\Big(1-\frac{1}{\tau}\Big), \tag{5.38}$$

whilst for $\tau > 3$,

$$\max_{i\in[n]} \frac{\mathbb{E}_W[\mathcal{Z}_n(i)] - m b_{t_n n}\log(1/t_n)}{m a_{t_n n}\log(1/t_n)(\log n)^{1-3/\tau}} \xrightarrow{\mathbb{P}} -\frac{\tau(\tau-1)^2}{2c_1^3}. \tag{5.39}$$

Proof. First, we show that, similar to (5.29),

$$\Big|\max_{i\in[n]} \frac{W_i \sum_{j=i}^{n-1} 1/S_j}{b_n} - \max_{i\in[n]} \frac{W_i}{b_n}\log(n/i)\Big| \xrightarrow{\mathbb{P}} 0, \tag{5.40}$$

so that in what follows we can work with the rightmost expression in the absolute value rather than the leftmost. This directly follows from writing the absolute value as

$$\Big|\max_{i\in[n]} \frac{W_i \sum_{j=i}^{n-1} 1/S_j}{b_n} - \max_{i\in[n]} \frac{W_i}{b_n}\log\Big(\frac{n}{i}\Big)\Big| \leq \max_{i\in[n]} \frac{W_i}{b_n}\Big|\sum_{j=i}^{n-1}\frac{1}{S_j} - \log\Big(\frac{n}{i}\Big)\Big| = \max_{i\in[n]} \frac{W_i}{b_n}|Y_n - Y_i|,$$

where $Y_n := \sum_{j=1}^{n-1} 1/S_j - \log n$, which converges almost surely by (5.2). We then split the maximum into two parts to obtain the upper bound, for any $\gamma \in (0,1)$,

$$\max_{i\in[n^\gamma]} \frac{W_i}{b_{n^\gamma}}(|Y_n| + \sup_{j\geq 1}|Y_j|)\frac{b_{n^\gamma}}{b_n} + \max_{n^\gamma \leq i \leq n} \frac{W_i}{b_n} \max_{n^\gamma \leq i \leq n} |Y_n - Y_i|.$$

The first maximum converges to 1 in probability, the term in the brackets converges almost surely and the second fraction tends to zero, as we recall from Remark 2.4 that $b_n = g(\log n)$ with g a rapidly-varying function at infinity. This implies, for any $\gamma \in (0,1)$, by the definition of a rapidly-varying function, that $b_{n^\gamma}/b_n = g(\gamma\log n)/g(\log n)$ converges to zero with n. Similarly, the second maximum converges to 1 in probability and the third maximum tends to zero almost surely, as Y_n is a Cauchy sequence almost surely. In total, the entire expression tends to zero in probability.

For the next part, we define

$$\ell(x) := c_1 + c_2 x^{-1}\Big(\frac{b}{\tau}\log x + b\log c_1 + \log a\Big).$$

Then, as we are working in the **(Gumbel)**-**(RaV)** sub-case in Assumption 2.3, we can write $b_n = \exp\{(\log n)^{1/\tau}\ell(\log n)\}$.

Using t_n we can show that for any fixed $r\in\mathbb{R}$ or $r = r(n)$ that does not grow 'too quickly' with n, $b_{t_n^r n}/b_n \sim \mathrm{e}^{-r}$. Namely, uniformly in $r = r(n) \le C\log\log(b_n)$ (for any constant $C>0$),

$$\begin{aligned}\frac{b_{t_n^r n}}{b_n} &= \exp\Big\{(\log n)^{1/\tau}\Big(\Big(1+r\frac{\log t_n}{\log n}\Big)^{1/\tau}\ell\Big(\log n\Big(1+r\frac{\log t_n}{\log n}\Big)\Big)-\ell(\log n)\Big)\Big\}\\ &\sim \exp\Big\{(\log n)^{1/\tau}\Big(\ell\Big(\log n\Big(1+r\frac{\log t_n}{\log n}\Big)\Big)-\ell(\log n)\Big)\\ &\qquad + (1/\tau)r\log t_n(\log n)^{1/\tau-1}\ell\Big(\log n\Big(1+r\frac{\log t_n}{\log n}\Big)\Big)\Big\},\end{aligned}\tag{5.41}$$

where we applied a Taylor approximation to $(1+r\log t_n/\log n)^{1/\tau}$, which holds uniformly in r as long as $r = o(\log n/\log t_n) = o(\log b_n)$. It is elementary to show that for such r, the first term in the exponent on the last line of (5.41) tends to zero. Thus, uniformly in $r\le C\log\log(b_n)$,

$$\frac{b_{t_n^r n}}{b_n} \sim \exp\Big\{-r\frac{\ell(\log n(1+r\log t_n/\log n))}{\ell(\log n)}\Big\} \sim \mathrm{e}^{-r},\tag{5.42}$$

where the last step follows a similar argument to the one used to show that the first term on the right-hand side of (5.41) tends to zero.

We thus note that by (5.42) and (5.40) it suffices to show that

$$\max_{i\in[n]}\frac{W_i\log(n/i)}{b_n\log(1/t_n)} \xrightarrow{\mathbb{P}} 1/\mathrm{e},\tag{5.43}$$

to prove (5.36).

We start by providing a lower bound to the left-hand side of (5.43). For some fixed $r>0$, we write

$$\max_{i\in[n]}\frac{W_i\log(n/i)}{b_n\log(1/t_n)} \ge \max_{i\in[t_n^r n]}\frac{W_i}{b_{t_n^r n}}\frac{\log(n/(t_n^r n))}{\log(1/t_n)}\frac{b_{t_n^r n}}{b_n} = \max_{i\in[t_n^r n]}\frac{W_i}{b_{t_n^r n}}r\frac{b_{t_n^r n}}{b_n}.$$

By (5.42), it follows that this lower bound converges in probability to $r\mathrm{e}^{-r}$. To maximise this expression, we choose $r=1$ giving the value $1/\mathrm{e}$ as claimed.

For an upper bound, we split the maximum into multiple parts which cover different ranges of the indices i. First, for ease of writing, let us denote

$$X_{n,i} := \frac{W_i\log(n/i)}{b_n\log(1/t_n)}.$$

Fix $\varepsilon>0$, then set $N = \lceil 2\log\log(b_n)/\varepsilon\rceil$, and define

$$r_0 = \mathrm{e}^{-1}, \text{ and } \quad r_i = r_0 + \varepsilon i \quad \text{for } i = 1,\dots,N.$$

Then,

$$\max_{i\in[n]}X_{n,i} \le \max\Big\{\max_{i\in[t_n^{r_N}n]}X_{n,i}, \max_{k=1,\dots,N}\max_{t_n^{r_k}n<i\le t_n^{r_{k-1}}n}X_{n,i}, \max_{t_n^{r_0}n<i\le n}X_{n,i}\Big\}.\tag{5.44}$$

We now bound each of these three parts separately. We start with the middle term and note that for $k \in \{1, \dots, N\}$,

$$\max_{t_n^{r_k} n < i \le t_n^{r_{k-1}} n} X_{n,i} = \max_{t_n^{r_k} n < i \le t_n^{r_{k-1}} n} \frac{W_i \log(n/i)}{b_n \log(1/t_n)} \le r_k \frac{b_{t_n^{r_{k-1}} n}}{b_n} \max_{t_n^{r_k} n < i \le t_n^{r_{k-1}} n} \frac{W_i}{b_{t_n^{r_{k-1}} n}}.$$

If we now define for $k = 0, \dots, N$,

$$A_n(k) := \max_{t_n^{r_{k+1}} n < i \le t_n^{r_k} n} \frac{W_i}{b_{t_n^{r_k} n}},$$

then, by (5.42), we have that

$$\begin{aligned} \max_{k=1,\dots,N} \max_{t_n^{r_k} n < i \le t_n^{r_{k-1}} n} X_{n,i} &\le (1+\varepsilon) \max_{k=1,\dots,N} r_k \mathrm{e}^{-r_{k-1}} A_n(k-1) \\ &\le (1+\varepsilon) \sup_{x \ge 1/\mathrm{e}} x \mathrm{e}^{-x+\varepsilon} \max_{k=0,\dots,N-1} A_n(k). \\ &\le (1+\varepsilon) \mathrm{e}^{-1+\varepsilon} \max_{k=0,\dots,N-1} A_n(k), \end{aligned} \tag{5.45}$$

using as before that $x \mapsto x\mathrm{e}^{-x}$ is maximised at $x = 1$. Similarly, we can bound the the last term in (5.44) as

$$\max_{t_n^{r_0} n < i \le n} X_{n,i} \le r_0 \max_{t_n^{r_0} n < i \le n} \frac{W_i}{b_n} = \mathrm{e}^{-1} A_n,$$

where we recall that $r_0 = 1/\mathrm{e}$ and we set $A_n := \max_{t_n^{r_0} n < i \le n} W_i / b_n$. Finally, for the first term in (5.44), we get that

$$\max_{i \in [t_n^{r_N} n]} X_{n,i} \le \frac{b_{t_n^{r_N} n}}{b_n} \frac{\log n}{\log(1/t_n)} \max_{i \in [t_n^{r_N} n]} \frac{W_i}{b_{t_n^{r_N} n}} \le \frac{1+\varepsilon}{\tau} \mathrm{e}^{-r_N} \log(b_n) \max_{i \in [t_n^{r_N} n]} \frac{W_i}{b_{t_n^{r_N} n}}. \tag{5.46}$$

Now using that $r_N \ge 2 \log\log(b_n)$ by definition, we find that the right-hand side is $o_{\mathbb{P}}(1)$.

Combining (5.44) with the estimates in (5.45)-(5.46), we obtain

$$\max_{i \in [n]} X_{n,i} \le (1+\varepsilon) \mathrm{e}^{-1+\varepsilon} \max\Big\{ \max_{k=0\dots,N-1} A_n(k), A_n \Big\}. \tag{5.47}$$

Since $\varepsilon > 0$ is arbitrary, it suffices to show that the maximum on the right-hand side is bounded by $1+\varepsilon$ with high probability. Using that the random variables follow a distribution as in the **(Gumbel)**-**(RaV)** case in Assumption 2.3, we can write, using a union bound and a large $C > 0$,

$$\begin{aligned} \mathbb{P}\Big(\max_{i \in [n]} \frac{W_i}{b_n} \ge 1+\varepsilon \Big) &\le C n \log((1+\varepsilon) b_n)^b \exp\{-(\log((1+\varepsilon) b_n)/c_1)^\tau\} \\ &= C n \log((1+\varepsilon) b_n)^b \exp\Big\{ -(\log(b_n)/c_1)^\tau \Big(1 + \frac{\log(1+\varepsilon)}{\log(b_n)} \Big)^\tau \Big\}. \end{aligned}$$

We now use the expression of b_n as in the **(Gumbel)**-**(RaV)** case in Assumption 2.3 to obtain the upper bound

$$\widetilde{C} \log(b_n)^b \exp\Big\{ \log n \Big(1 - \Big(1 + \frac{(b/\tau) \log\log n + b \log c_1 + \log \tau}{\tau \log n} \Big)^\tau \Big(1 + \frac{\log(1+\varepsilon)}{\log(b_n)} \Big)^\tau \Big) \Big\},$$

where $\widetilde{C} > 0$ is a suitable constant. Using a Taylor approximation on the terms in the exponent and using the asymptotics of $\log(b_n)$, we find an upper bound

$$K_1 (\log n)^{b/\tau} \exp\{-K_2 (\log n)^{1-1/\tau}\}, \tag{5.48}$$

for some constants $K_1, K_2 > 0$ and n sufficiently large. Note that this expression tends to zero as $\tau > 1$. Now, we aim to apply this bound to the maximum in (5.47). First, we use a union bound to arrive at

$$\mathbb{P}\Big(\max\Big\{\max_{k=0,\dots,N-1} A_n(k), A_n\Big\}\geq 1+\varepsilon\Big)\leq \sum_{k=0}^{N-1}\mathbb{P}\Big(\max_{i\in[t_n^{r_k}n]} W_i/b_{t_n^{r_k}n}\geq 1+\varepsilon\Big) + \mathbb{P}\Big(\max_{i\in[n]} W_i/b_n\geq 1+\varepsilon\Big).$$

The last term tends to zero with n. For the sum we use (5.48) and note that this upper bound tends to zero slowest for $k=N-1$, so that we obtain the upper bound

$$\begin{aligned}\sum_{k=0}^{N-1}\mathbb{P}\Big(\max_{i\in[t_n^{r_k}n]} W_i/b_{t_n^{r_k}n}\geq 1+\varepsilon\Big)&\leq NK_1\log(t_n^{r_{N-1}}n)^{b/\tau}\exp\{-K_2\log(t_n^{r_{N-1}}n)^{1-1/\tau}\}\\ &\leq K_3\log\log(b_n)(\log n)^{b/\tau}\exp\{-K_4(\log n)^{1-1/\tau}\},\end{aligned}$$

for some constants K_3, K_4, since $r_{N-1}=\mathcal{O}(\log\log(b_n))$, which again tends to zero with n as $\tau>1$.

Finally, we prove the convergence of $\log(\widetilde{I}_n)/\log n$. Let $\eta\in(0,1)$. Then, the event

$$E_n:=\Big\{\max_{i\in[n]}\frac{\mathbb{E}_W[\mathcal{Z}_n(i)]}{b_{t_nn}\log(1/t_n)}\geq\eta\Big\}$$

holds with high probability by the above. Using this and (5.40) yields, for $\varepsilon>0$,

$$\begin{aligned}\mathbb{P}\bigg(\frac{\log\widetilde{I}_n}{\log n}<1-\varepsilon\bigg)&\leq\mathbb{P}\bigg(\Big\{\frac{\log\widetilde{I}_n}{\log n}<1-\varepsilon\Big\}\cap E_n\bigg)+\mathbb{P}(E_n^c)\\ &\leq\mathbb{P}\bigg(\max_{i<n^{1-\varepsilon}}\frac{W_i\log(n/i)}{b_{t_nn}\log(1/t_n)}\geq\eta\bigg)+\mathbb{P}(E_n^c)\,.\end{aligned}$$

The second probability tends to zero with n and the first can be bounded from above by

$$\mathbb{P}\bigg(\max_{i\leq n^{1-\varepsilon}}\frac{W_i}{b_{n^{1-\varepsilon}}}\frac{b_{n^{1-\varepsilon}}\log(b_n)}{b_{t_nn}}\geq\tau\eta\bigg)\,. \tag{5.49}$$

Now,

$$\frac{b_{n^{1-\varepsilon}}\log(b_n)}{b_{t_nn}}\sim\exp\Big\{1+(\log n)^{1/\tau}\ell(\log n)\Big((1-\varepsilon)^{1/\tau}\frac{\ell((1-\varepsilon)\log n)}{\ell(\log n)}-1\Big)+\log\log(b_n)\Big\},$$

which, since ℓ is a slowly-varying function at infinity and $(1-\varepsilon)^{1/\tau}<1$, tends to zero with n. As the maximum in (5.49) tends to 1 in probability, we obtain that the probability in (5.49) tends to zero with n.

To prove (5.37) we use a similar argument as in the proof of Proposition 5.4, as distributions satisfying the **(Gumbel)**-**(RaV)** sub-case also fall in the Gumbel MDA. Namely, we can write

$$\begin{aligned}\frac{W_i\sum_{j=i}^{n-1}1/S_j-b_{t_nn}\log(1/t_n)}{a_{t_nn}\log(1/t_n)}&=\frac{W_i-b_{t_nn}}{a_{t_nn}}\frac{\sum_{j=i}^{n-1}1/S_j}{\log(1/t_n)}-\log\Big(\frac{i}{t_nn}\Big)\\ &\quad+\frac{b_{t_nn}}{a_{t_nn}\log(1/t_n)}\Big(\sum_{j=i}^{n-1}1/S_j-\log(n/i)\Big)\\ &\quad-\Big(\frac{b_{t_nn}}{a_{t_nn}\log(1/t_n)}-1\Big)\log\Big(\frac{i}{t_nn}\Big),\end{aligned}$$

so that

$$\begin{aligned}&\bigg|\max_{i\in C_n}\frac{W_i\sum_{j=i}^{n-1}1/S_j-b_{t_nn}\log(1/t_n)}{a_{t_nn}\log(1/t_n)}-\max_{i\in C_n}\Big(\frac{W_i-b_{t_nn}}{a_{t_nn}}\frac{\sum_{j=i}^{n-1}1/S_j}{\log(1/t_n)}-\log\Big(\frac{i}{t_nn}\Big)\Big)\bigg|\\ &\leq\frac{b_{t_nn}}{a_{t_nn}\log(1/t_n)}\max_{i\in C_n}\bigg|\sum_{j=i}^{n-1}\frac{1}{S_j}-\log(n/i)\bigg|+\bigg|\frac{b_{t_nn}}{a_{t_nn}\log(1/t_n)}-1\bigg|\max_{i\in C_n}\bigg|\log\Big(\frac{i}{t_nn}\Big)\bigg|,\end{aligned} \tag{5.50}$$

where we omit the arguments of $C_n(1,s,t,t_n)$ for brevity. We now note that, in the (**RaV**) sub-case, $a_n := c_2(\log n)^{1/\tau-1}b_n$, which yields

$$\frac{b_{t_nn}}{a_{t_nn}\log(1/t_n)} = \Big(c_2\log(t_nn)^{1/\tau-1}\log(1/t_n)\Big)^{-1} \sim \Big(c_2(\log n)^{1/\tau-1}\frac{\tau\log n}{\log(b_n)}\Big)^{-1} \sim \frac{\ell(\log n)}{c_1},$$

which converges to 1 as n tends to infinity. Here, we use that $c_1 = c_2\tau$, and that $\log(b_n) = (\log n)^{1/\tau}\ell(\log n)$, with $\lim_{x\to\infty}\ell(x) = c_1$. It follows, with a similar argument as in the proof of Proposition 5.4, that the right-hand side of (5.50) converges to zero almost surely. Now, the rest of the proof of (5.37) follows the exact same approach as the proof of Proposition 5.4.

Finally, we prove (5.38) and (5.39). Again, we study $\max_{i\in[n]} W_i\log(n/i)$ rather than $\max_{i\in[n]}\mathbb{E}_W[\mathcal{Z}_n(i)]$. The general approach is similar to the proof of (5.11) in Proposition 5.4, though the details differ. We first consider the case $\tau\in(1,3]$ and then tend to the case $\tau>3$. In both cases, we prove a lower and upper bound. Moreover, for the lower bound we need only consider indices $t_nn\le i\le \mathrm{e}^{k_n}t_nn$ for a particular choice of k_n.

Fix $\tau\in(1,3], x\in\mathbb{R}$ and let $k_n := \sqrt{(\tau-1)/c_1}\sqrt{(\log n)^{1-1/\tau}\log\log n}$. We bound

$$\begin{aligned}
&\mathbb{P}\Big(\max_{t_nn\le i\le \mathrm{e}^{k_n}t_nn}\frac{W_i\log(n/i)-b_{t_nn}\log(1/t_n)}{a_{t_nn}\log(1/t_n)\log\log n}\le x\Big)\\
&\quad=\mathbb{P}\Big(\max_{t_nn\le i\le \mathrm{e}^{k_n}t_nn} W_i\log(n/i)\le \log(1/t_n)(b_{t_nn}+a_{t_nn}x\log\log n)\Big)\\
&\quad=\prod_{i=t_nn}^{\mathrm{e}^{k_n}t_nn}\Big(1-\mathbb{P}\Big(W\ge\frac{\log(1/t_n)}{\log(n/i)}(b_{t_nn}+a_{t_nn}x\log\log n)\Big)\Big)\\
&\quad\le\exp\Big\{-\sum_{i=t_nn}^{\mathrm{e}^{k_n}t_nn}\mathbb{P}\Big(W\ge\frac{\log(1/t_n)}{\log(n/i)}(b_{t_nn}+a_{t_nn}\log\log n)\Big)\Big\}\\
&\quad\le\exp\Big\{-\sum_{j=1}^{k_n}\sum_{i=\mathrm{e}^{j-1}t_nn}^{\mathrm{e}^{j}t_nn}\mathbb{P}\Big(W\ge\frac{\log(1/t_n)}{\log(1/t_n)-j}(b_{t_nn}+a_{t_nn}x\log\log n)\Big)\Big\}.
\end{aligned}\tag{5.51}$$

We then obtain an upper bound by bounding the probability from below. So, for $\delta>0$ small and n large, we arrive at the upper bound

$$\begin{aligned}
&\exp\Big\{-(1-\delta)a(\mathrm{e}-1)\sum_{j=1}^{k_n}\mathrm{e}^{j-1}t_nn\Big(\log\Big(\frac{\log(1/t_n)}{\log(1/t_n)-j}(b_{t_nn}+a_{t_nn}x\log\log n)\Big)\Big)^{b}\\
&\qquad\times\exp\Big\{-\Big(\log\Big(\frac{\log(1/t_n)}{\log(1/t_n)-j}(b_{t_nn}+a_{t_nn}x\log\log n)\Big)/c_1\Big)^{\tau}\Big\}\Big\}.
\end{aligned}\tag{5.52}$$

As $\log(1/t_n)/(\log(1/t_n)-j)=1+o(1)$ uniformly in j, we can write the inner exponent as

$$\begin{aligned}
&-\Big(\log\Big(\frac{\log(1/t_n)}{\log(1/t_n)-j}(b_{t_nn}+a_{t_nn}x\log\log n)\Big)/c_1\Big)^{\tau}\\
&\quad=-\Big(-\frac{1}{c_1}\log\Big(1-\frac{j}{\log(1/t_n)}\Big)+\frac{1}{c_1}\log(b_{t_nn}+a_{t_nn}x\log\log n)\Big)^{\tau}\\
&\quad=-\Big(\sum_{\ell=1}^{\infty}\frac{1}{\ell c_1}\Big(\frac{j}{\log(1/t_n)}\Big)^{\ell}+\log(t_nn)^{1/\tau}+\frac{1}{\tau}\log(t_nn)^{1/\tau-1}\log\big(a(c_1(\log t_nn)^{1/\tau})^b\big)\\
&\qquad+\sum_{\ell=1}^{\infty}(-1)^{\ell-1}\frac{1}{\ell c_1}\big(c_2\log(t_nn)^{1/\tau-1}x\log\log n\big)^{\ell}\Big)^{\tau}.
\end{aligned}$$

Taking out a factor $\log(t_nn)^{1/\tau}$ then yields

$$
\begin{aligned}
&-\log(t_nn)\Big(1+\sum_{\ell=1}^{\infty}\frac{1}{\ell c_1}\Big(\frac{j}{\log(1/t_n)}\Big)^{\ell}\log(t_nn)^{-1/\tau}+\frac{\log\big(a(c_1(\log t_nn)^{1/\tau})^b\big)}{\tau\log(t_nn)}\\
&+\sum_{\ell=1}^{\infty}(-1)^{\ell-1}\frac{1}{\ell c_1}\big(c_2\log(t_nn)^{1/\tau-1}x\log\log n\big)^{\ell}\log(t_nn)^{-1/\tau}\Big)^{\tau}\\
&\quad=-\log(t_nn)-\log\big(a(c_1(\log t_nn)^{1/\tau})^b\big)-x\log\log n-\frac{\log(t_nn)^{1-1/\tau}}{c_2\log(1/t_n)}j\\
&\quad-\frac{\log(t_nn)^{1-1/\tau}}{2c_2\log(1/t_n)^2}j^2(1+o(1))+o(1).
\end{aligned}
\tag{5.53}
$$

Now, for $\ell\in\{1,2\}$,

$$
\begin{aligned}
\frac{1}{\ell c_2}\frac{j^{\ell}\log(t_nn)^{1-1/\tau}}{\log(1/t_n)^{\ell}}&=\frac{j^{\ell}}{\ell c_2}c_2^{\ell}(\log n)^{-(\ell-1)(1-1/\tau)}\Big(1+\frac{\log t_n}{\log n}\Big)^{1-1/\tau}\Big(\frac{\log(b_n)}{c_1(\log n)^{1/\tau}}\Big)^{\ell}\\
&=\frac{j^{\ell}}{\ell}\big(c_2(\log n)^{-(1-1/\tau)}\big)^{\ell-1}\Big(1-\frac{\tau-1}{\log(b_n)}(1+\mathcal{O}((\log n)^{-1/\tau}))\Big)\\
&\quad\times\Big(1+\frac{1}{\tau}(\log n)^{-1}\log\big(a(c_1(\log n)^{1/\tau})^b\big)\Big)^{\ell},
\end{aligned}
$$

where the last two terms are both $(1+\mathcal{O}((\log n)^{-1/\tau}))$. So, we obtain for the right hand side of (5.53)

$$
\begin{aligned}
&-\log(t_nn)-\log\big(a(c_1(\log t_nn)^{1/\tau})^b\big)-x\log\log n-\frac{j^2}{2}c_2(\log n)^{-(1-1/\tau)}\\
&\qquad-j\Big(1+\frac{1}{\tau}(\log n)^{-1}\log\big(a(c_1(\log n)^{1/\tau})^b\big)-\frac{\tau-1}{\log(b_n)}\Big)+o(1),
\end{aligned}
\tag{5.54}
$$

uniformly in $j\in[k_n]$, where we note that the final $o(1)$ term vanishes uniformly in j since $\tau\in(1,3]$. Using this in (5.52), we arrive at

$$
\begin{aligned}
&\mathbb{P}\Big(\max_{t_nn\le i\le \mathrm{e}^{k_n}t_nn}\frac{W_i\log(n/i)-b_{t_nn}\log(1/t_n)}{a_{t_nn}\log(1/t_n)\log\log n}\le x\Big)\\
&\quad\le\exp\Big\{-(1-\delta)(1-1/\mathrm{e})\sum_{j=1}^{k_n}(\log n)^{-x}(1+o(1))\exp\big\{\mathrm{e}_nj-d_nj^2\big\}\Big\},
\end{aligned}
\tag{5.55}
$$

where

$$
\mathrm{e}_n:=\frac{\tau-1}{\log(b_n)}-\frac{1}{\tau}(\log n)^{-1}\log\big(a(c_1(\log n)^{1/\tau})^b\big),\qquad d_n:=\frac{c_2}{2(\log n)^{1-1/\tau}}.
$$

The expression $\mathrm{e}_nj-d_nj^2$ is increasing for $j\le \mathrm{e}_n/(2d_n)=o(k_n)$, so bound the sum from below by

$$
\begin{aligned}
\sum_{j=1}^{k_n}\mathrm{e}^{\mathrm{e}_nj-d_nj^2}&=\sum_{j=1}^{k_n}\exp\Big\{-d_n\Big(j-\frac{\mathrm{e}_n}{2d_n}\Big)^2+\frac{\mathrm{e}_n^2}{4d_n}\Big\}\\
&\ge\exp\big\{\mathrm{e}_n^2/(4d_n)\big\}\int_0^{k_n}\exp\Big\{-d_n\Big(y-\frac{\mathrm{e}_n}{2d_n}\Big)^2\Big\}\,\mathrm{d}y.
\end{aligned}
$$

Set $\mu_n:=\mathrm{e}_n/(2d_n),\sigma_n:=1/\sqrt{2d_n}$ and let $X_n\sim\mathcal{N}(\mu_n,\sigma_n^2)$ be a normal random variable. Then, we can write this as

$$
\begin{aligned}
&\exp\big\{\mathrm{e}_n^2/(4d_n)\big\}\sigma_n\sqrt{2\pi}\int_0^{k_n}\frac{1}{\sigma_n\sqrt{2\pi}}\exp\Big\{-\frac{1}{2}\Big(\frac{y-\mu_n}{\sigma_n}\Big)^2\Big\}\,\mathrm{d}y\\
&\quad=\exp\big\{\mathrm{e}_n^2/(4d_n)\big\}\sigma_n\sqrt{2\pi}\,\mathbb{P}(X_n\in(0,k_n))\,,
\end{aligned}
$$

Let $Z\sim\mathcal{N}(0,1)$ be a standard normal. We can then write the last line as

$$\exp\left\{\mathrm{e}_n^2/(4d_n)\right\}\sigma_n\sqrt{2\pi}\mathbb{P}\Big(Z\in\Big(-\frac{\mu_n}{\sigma_n},\frac{k_n-\mu_n}{\sigma_n}\Big)\Big). \tag{5.56}$$

It is clear that for $\tau\leq 3$,

$$\begin{aligned}\frac{\mathrm{e}_n^2}{4d_n}&=\frac{\tau(\tau-1)^2}{2c_1^3}(\log n)^{1-3/\tau}(1+o(1)),\\ \frac{k_n-\mu_n}{\sigma_n}&=\sqrt{(1-1/\tau)\log\log n}(1+o(1)),\\ \frac{\mu_n}{\sigma_n}&=\sqrt{\frac{\tau(\tau-1)^2}{c_1^3}}(\log n)^{(1-3/\tau)/2}(1+o(1)).\end{aligned} \tag{5.57}$$

It thus follows that, when $\tau\in(1,3]$, the probability as well as the exponential term in (5.56) converge to a strictly positive constant. So, for some $K>0$, we bound the expression in (5.56) from below by

$$K\sigma_n=\frac{K}{\sqrt{c_2}}(\log n)^{(1-1/\tau)/2}.$$

Using this in (5.55) finally yields the lower bound

$$\exp\Big\{-(1-\delta)(1-1/\mathrm{e})(\log n)^{(1-1/\tau)/2-x}(1+o(1))\Big\},$$

which converges to zero for any $x<(1-1/\tau)/2$. We thus arrive at, with high probability,

$$\max_{i\in[n]}\max_{t_nn\leq i\leq \mathrm{e}^{k_n}t_nn}\frac{W_i\log(n/i)-b_{t_nn}\log(1/t_n)}{a_{t_nn}\log(1/t_n)\log\log n}\geq\frac{1}{2}\Big(1-\frac{1}{\tau}\Big)+\eta,$$

for any $\eta>0$.

To prove a matching upper bound, we split the set $[n]$ into four parts: the indices $1\leq i\leq \mathrm{e}^{-k_n}t_nn$, $\mathrm{e}^{-k_n}t_nn\leq i\leq t_nn$, $t_nn\leq i\leq \mathrm{e}^{k_n}t_nn$ and $\mathrm{e}^{k_n}t_nn\leq i\leq n$. We prove an upper bound for all four ranges of indices, a union bound then concludes the proof. The proof for the first two ranges of indices is analogous to the proof for the latter two, so we focus on the latter two. Let us start with the range $t_nn\leq i\leq \mathrm{e}^{k_n}t_nn$. We can use the same approach as above, though with minor adaptations. First, using that $1-y\geq\exp\{-(1+\delta)y\}$ for y sufficiently small and $\delta>0$ fixed, we obtain

$$\begin{aligned}&\mathbb{P}\Big(\max_{t_nn\leq i\leq \mathrm{e}^{k_n}t_nn}\frac{W_i\log(n/i)-b_{t_nn}\log(1/t_n)}{a_{t_nn}\log(1/t_n)\log\log n}\leq x\Big)\\ &\geq\exp\Big\{-(1+\delta)\sum_{j=1}^{k_n}\sum_{i=\mathrm{e}^{j-1}t_nn}^{\mathrm{e}^jt_nn}\mathbb{P}\Big(W\geq\frac{\log(1/t_n)}{\log(1/t_n)-(j-1)}(b_{t_nn}+a_{t_nn}x\log\log n)\Big)\Big\}.\end{aligned}$$

We then bound the probability from above and use the same Taylor expansions to obtain a lower bound of the same form as (5.52). We bound the sum over j from above by

$$\sum_{j=1}^{k_n}\mathrm{e}^{\mathrm{e}_n(j-1)-d_n(j-1)^2}\leq\exp\left\{\mathrm{e}_n^2/(4d_n)\right\}\sigma_n\sqrt{2\pi}\mathbb{P}(X_n\in(0,k_n)),$$

where $d_n,\mathrm{e}_n,\mu_n,\sigma_n$ and X_n are as above. With a similar argument, we obtain an upper bound

$$\widetilde{K}(\log n)^{(1-1/\tau)/2},$$

for some appropriate constant $\widetilde{K}>0$. We thus obtain

$$\mathbb{P}\Big(\max_{t_nn\leq i\leq \mathrm{e}^{k_n}t_nn}\frac{W_i\log(n/i)-b_{t_nn}\log(1/t_n)}{a_{t_nn}\log(1/t_n)\log\log n}\leq x\Big)\geq\exp\left\{-(1+\delta)^2(\mathrm{e}-1)(\log n)^{(1-1/\tau)/2-x)}\right\},$$

which converges to one for any $x>(1-1/\tau)/2$.

To prove an upper bound for the range $\mathrm{e}^{k_n} t_n n \leq i \leq n$, a slight adaptation is required. We write,

$$
\begin{aligned}
\mathbb{P}\Big(& \max_{\mathrm{e}^{k_n} t_n n \leq i \leq n} \frac{W_i \log(n/i) - b_{t_n n}\log(1/t_n)}{a_{t_n n}\log(1/t_n)\log\log n} \leq x\Big) \\
& \geq \exp\Big\{ -(1+\delta) \sum_{j=1}^{\log(1/t_n)-k_n} \sum_{i=\mathrm{e}^{j-1+k_n}t_n n}^{\mathrm{e}^{j+k_n}t_n n} \mathbb{P}\Big(W \geq \frac{\log(1/t_n)(b_{t_n n}+a_{t_n n}x\log\log n)}{\log(1/t_n)-(j-1+k_n)}\Big)\Big\},
\end{aligned}
$$

and bound the probability from above by

$$
\begin{aligned}
(1+\delta)a\Big(& \log\Big(\frac{\log(1/t_n)}{\log(1/t_n)-(j-1+k_n)}(b_{t_n n}+a_{t_n n}x\log\log n)\Big)\Big)^b \\
& \times \exp\Big\{ -\Big(\log\Big(\frac{\log(1/t_n)}{\log(1/t_n)-(j-1+k_n)}(b_{t_n n}+a_{t_n n}x\log\log n)\Big)/c_1\Big)^\tau\Big\}.
\end{aligned}
$$

Since the fraction $\log(1/t_n)/(\log(1/t_n)-(j-1+k_n))$ is no longer $1+o(1)$ uniformly in j, we treat this term somewhat differently. We write the exponent as

$$
\begin{aligned}
& -(\log(b_{t_n n})/c_1)^\tau - x\log\log n(1+o(1)) \\
& \qquad + \frac{1}{c_2}\log(t_n n)^{1-1/\tau}\log\Big(1-\frac{j-1+k_n}{\log(1/t_n)}\Big)\Big(1+\mathcal{O}\Big(\frac{\log\log n}{(\log n)^{1/\tau}}\Big)\Big).
\end{aligned}
$$

Then using that as before $t_n n a(\log(b_{t_n n}))^b \exp\{-(b_{t_n n}/c_1)^\tau\} \sim 1$, this yields a lower bound

$$
\begin{aligned}
& \exp\Big\{ -(1+\delta)^2(\mathrm{e}-1) \sum_{j=1}^{\log(1/t_n)-k_n} (\log n)^{-x(1+o(1))}(1+o(1)) \\
& \quad \times \exp\Big\{ j-1+k_n + \frac{\log(t_n n)^{1-1/\tau}}{c_2}\log\Big(1-\frac{j-1+k_n}{\log(1/t_n)}\Big)\Big(1+\mathcal{O}\Big(\frac{\log\log n}{(\log n)^{1/\tau}}\Big)\Big)\Big\}\Big\}.
\end{aligned} \tag{5.58}
$$

The mapping $x \mapsto x + f_n\log(1-x/g_n)$, $x < g_n$, for some sequences f_n, g_n, is maximised at $x = g_n - f_n$. In this case, with $f_n := c_2^{-1}\log(t_n n)^{1-1/\tau}, g_n = \log(1/t_n)$, the mapping is maximised at

$$
g_n - f_n = \frac{\tau(\tau-1)}{c_1^2}(\log n)^{1-2/\tau} - \frac{\tau^3(\tau-1)}{2c_1^3}(\log n)^{1-3/\tau} + o(1).
$$

Since $(\log n)^{1-2/\tau} = o(k_n)$ when $\tau \in (1,3]$, as $(1-1/\tau)/2 \geq 1-2/\tau$ for $\tau\in(1,3]$, it follows that the inner exponent is largest when $j=1$. This yields the lower bound

$$
\begin{aligned}
& \exp\Big\{ -(1-\delta)^2(\mathrm{e}-1)\log(1/t_n)(\log n)^{-x(1+o(1))}(1+o(1)) \\
& \quad \times \exp\Big\{\Big(-K_1\frac{k_n}{\tau\log n}\log\log n + \frac{(\tau-1)k_n}{\log b_n} - \frac{1}{2}\frac{k_n^2}{\log(1/t_n)}\Big)\Big(1+\mathcal{O}\Big(\frac{\log\log n}{(\log n)^{1/\tau}}\Big)\Big)\Big\}\Big\},
\end{aligned}
$$

for some small constant $K_1 > 0$. The first two terms in the inner exponent are negligible (compared to the last term) by the choice of k_n. The last term equals $-((1-1/\tau)/2)\log\log n(1+o(1))$, so that we finally obtain the lower bound

$$
\exp\big\{ -K_2(\log n)^{(1-1/\tau)/2-x(1+o(1))}\big\},
$$

for some sufficiently large $K_2 > 0$. It thus follows that for any $x > (1-1/\tau)/2$ the lower bound converges to one. Together with the result for the range of indices $t_n n \leq i \leq \mathrm{e}^{k_n} t_n n$ (and a similar result for $1 \leq i \leq t_n n$, with analogous proofs), the upper bound then follows, and finishes the proof for the case $\tau \in (1,3]$.

When $\tau > 3$, we set $k_n := (\tau(\tau-1)/c_1^2)(\log n)^{1-2/\tau}$. For a lower bound on the maximum, we again consider the indices $t_n n \leq i \leq \mathrm{e}^{k_n} t_n n$. The steps in (5.51) through (5.56) are still valid when replacing $\log\log n$ with $(\log n)^{1-3/\tau}$. The only minor differences are that $\mathrm{e}_n/d_n \sim k_n$ rather than $o(k_n)$ and that the $o(1)$ term in (5.54) needs to be replaced by

$\mathcal{O}\big(\sqrt{\log\log n(\log n)^{1-5/\tau}}\big)$, though this changes nothing for the rest of the argument. The second quantity in (5.57) does change, however. We now find that

$$\frac{k_n-\mu_n}{\sigma_n}=\sqrt{\frac{\tau(\tau-1)^2}{c_1^3}}(\log n)^{(1-5/\tau)/2}\frac{\log\big(a\big(c_1(\log n)^{1/\tau}\big)^b\big)}{\tau}(1+o(1)).$$

As a result, the probability in (5.56) still converges to a constant, but now the exponential term diverges. We thus obtain the lower bound

$$\begin{aligned}&\mathbb{P}\Big(\max_{t_nn\le i\le \mathrm{e}^{k_n}t_nn}\frac{W_i\log(n/i)-b_{t_nn}\log(1/t_n)}{a_{t_nn}\log(1/t_n)(\log n)^{1-3/\tau}}\le x\Big)\\&\quad\le\exp\Big\{-(1-\delta)(1-1/\mathrm{e})(\log n)^{(1-1/\tau)/2}(1+o(1))\\&\quad\times\exp\Big\{\Big(\frac{\tau(\tau-1)^2}{2c_1^3}+x\Big)(\log n)^{1-3/\tau}(1+o(1))+\mathcal{O}\Big(\sqrt{\log\log n(\log n)^{1-5/\tau}}\Big)\Big\}\Big\},\end{aligned}$$

which converges to zero for any $x>-\tau(\tau-1)^2/(2c_1^3)$.

To prove an upper bound, we again adjust the arguments for the $\tau\in(1,3]$ case. Again, we substitute $(\log n)^{1-3/\tau}$ for $\log\log n$. The lower bound on the probability for indices $t_nn\le i\le \mathrm{e}^{k_n}t_nn$ remains valid, so that we obtain a lower bound that converges to zero for any $x<-\tau(\tau-1)^2/(2c_1^3)$.

For the range $\mathrm{e}^{k_n}t_nn\le i\le n$, we find that the expression in (5.58) still holds (again when switching $(\log n)^{1-3/\tau}$ for $\log\log n$). However, we improve on the accuracy of (5.58) by including more terms of the Taylor expansion. This yields, for an appropriate constant $K_3>0$,

$$\begin{aligned}&\exp\Big\{-K_3\sum_{j=1}^{\log(1/t_n)-k_n}\exp\Big\{j-1+k_n+\frac{1}{c_2}\log(t_nn)^{1-1/\tau}\log\Big(1-\frac{j-1+k_n}{\log(1/t_n)}\Big)\\&\qquad+\frac{\tau(\tau-1)}{2c_1^2}\log(t_nn)^{1-2/\tau}\Big(\log\Big(1-\frac{j-1+k_n}{\log(1/t_n)}\Big)\Big)^2(1+o(1))+x(\log n)^{1-3/\tau}\Big\}\Big\}.\end{aligned}$$

As is the case when $\tau\in(1,3]$, the inner exponent is largest when $j=1$, yielding the lower bound

$$\begin{aligned}&\exp\Big\{-K_3\log(1/t_n)\exp\Big\{k_n+\frac{1}{c_2}\log(t_nn)^{1-1/\tau}\log\Big(1-\frac{k_n}{\log(1/t_n)}\Big)(1+o(1))\\&\qquad+x(\log n)^{1-3/\tau}+O\Big(\sqrt{\log\log n(\log n)^{1-5/\tau}}\Big)\Big\}\Big\}.\end{aligned}$$

Now, applying a Taylor expansion to the logarithmic term, we can write the inner exponent as

$$\begin{aligned}&k_n-\frac{\log(t_nn)^{1-1/\tau}}{c_2\log(1/t_n)}k_n-\frac{\log(t_nn)^{1-1/\tau}}{2c_2\log(1/t_n)^2}k_n^2+x(\log n)^{1-3/\tau}+O\Big(\sqrt{\log\log n(\log n)^{1-5/\tau}}\Big)\\&\quad=\Big(k_n\frac{\tau-1}{c_1(\log n)^{1/\tau}}-\frac{c_2}{2}k_n^2(\log n)^{-(1-1/\tau)}\Big)(1+o(1))+x(\log n)^{1-3/\tau}+o\big((\log n)^{1-3/\tau}\big)\\&\quad=\frac{\tau(\tau-1)^2}{2c_1^3}(\log n)^{1-3/\tau}(1+o(1))+x(\log n)^{1-3/\tau}+o\big((\log n)^{1/3-\tau}\big).\end{aligned}$$

Concluding, we obtain the lower bound

$$\exp\Big\{-K_3\log(1/t_n)\exp\Big\{\Big(\frac{\tau(\tau-1)^2}{2c_1^3}(1+o(1))+x\Big)(\log n)^{1-3/\tau}+o\big((\log n)^{1/3-\tau}\big)\Big\}\Big\},$$

so that we obtain a limit of one when choosing any $x<-\tau(\tau-1)^2/(2c_1^3)$, which concludes the proof. □

Proposition 5.6. *Consider the WRG model as in Definition 2.1 and suppose the vertex-weights satisfy the* (**Fréchet**) *case in Assumption 2.3. Let* Π *be a PPP on* $(0,1)\times(0,\infty)$ *with intensity measure* $\nu(\mathrm{d}t,\mathrm{d}x) := \mathrm{d}t \times (\alpha-1)x^{-\alpha}\mathrm{d}x,\ x>0$*. When* $\alpha>2$*,*

$$\max_{i\in[n]}\mathbb{E}_W[\mathcal{Z}_n(i)/u_n] \xrightarrow{d} m \max_{(t,f)\in\Pi} f\log(1/t),$$

and when $\alpha\in(1,2)$*,*

$$\max_{i\in[n]}\mathbb{E}_W[\mathcal{Z}_n(i)/n] \xrightarrow{d} m \max_{(t,f)\in\Pi} f\int_t^1\bigg(\int_{(0,1)\times(0,\infty)} g\mathbb{1}_{\{u\leq s\}}\,\mathrm{d}\Pi(u,g)\bigg)^{-1}\mathrm{d}s.$$

Proof. First, let $\alpha>2$. We first claim that

$$\bigg|\max_{i\in[n]}\mathbb{E}_W[\mathcal{Z}_n(i)/u_n] - m\max_{i\in[n]}\frac{W_i\log(n/i)}{u_n}\bigg| \xrightarrow{\mathbb{P}} 0. \tag{5.59}$$

The claim's proof follows a similar structure as that of (5.40). Let us define the point process

$$\Pi_n := \sum_{i=1}^n \delta_{(i/n,W_i/u_n)}.$$

By [25], when the W_i are i.i.d. random variables in the Fréchet maximum domain of attraction with parameter $\alpha-1$, then Π is the weak limit of Π_n. Since $W_i\log(n/i)/u_n$ is a continuous mapping of $(i/n, W_i/u_n)$ and since taking the maximum is a continuous mapping too, it follows that

$$\max_{i\in[n]}\frac{W_i\log(n/i)}{u_n} \xrightarrow{d} \max_{(t,f)\in\Pi} f\log(1/t),$$

which, together with (5.59), yields the desired result. We now consider $\alpha\in(1,2)$. Note that

$$\max_{i\in[n]}\mathbb{E}_W[\mathcal{Z}_n(i)/n] = m\max_{i\in[n]}\frac{W_i}{n}\sum_{j=i}^{n-1}\frac{1}{S_j}.$$

The distributional convergence of the maximum on the right-hand side to the desired limit is proved in [19, Proposition 5.1], which concludes the proof. □

6. Concentration of the maximum degree

In this section we provide an important step to prove Theorems 2.9, 2.12 and 2.14: we discuss the concentration of the maximum degree around the maximum conditional mean degree, the behaviour of which is discussed in the previous section. To obtain this result, we combine union bounds with precise large deviation bounds for $|\mathcal{Z}_n(i)-\mathbb{E}_W[\mathcal{Z}_n(i)]|, i\in[n]$, using that $\mathcal{Z}_n(i)$ is a sum of independent indicator random variables. To this end, we present the following proposition:

Proposition 6.1. *Consider the WRG model as in Definition 2.1 and recall the vertex-weight conditions as in Assumption 2.3. When the vertex-weights satisfy the* (**Gumbel**)*-*(**SV**) *sub-case, for any* $\eta>0$*,*

$$\lim_{n\to\infty}\mathbb{P}\Big(\max_{i\in[n]}\big|\mathcal{Z}_n(i)-\mathbb{E}_W[\mathcal{Z}_n(i)]\big|\geq \eta b_n\log n\Big)=0. \tag{6.1}$$

When the vertex-weights satisfy the (**Gumbel**)*-*(**RV**) *sub-case,*

$$\max_{i\in[n]}\big|\mathcal{Z}_n(i)-\mathbb{E}_W[\mathcal{Z}_n(i)]\big|/b_n\log n \xrightarrow{\mathbb{P}-\text{a.s.}} 0. \tag{6.2}$$

Furthermore, when the vertex-weights satisfy the (**Gumbel**)*-*(**RaV**) *sub-case, let* $t_n := \exp\{-\tau\log n/\log(b_n)\}$*. Then, for any* $\eta>0$*,*

$$\lim_{n\to\infty}\mathbb{P}\Big(\max_{i\in[n]}\big|\mathcal{Z}_n(i)-\mathbb{E}_W[\mathcal{Z}_n(i)]\big|\geq \eta a_{t_n n}\log(1/t_n)\Big)=0. \tag{6.3}$$

Now suppose the vertex-weights satisfy the (**Fréchet**) *case. When $\alpha > 2$, for any $\eta > 0$,*

$$\lim_{n\to\infty} \mathbb{P}\Big(\max_{i\in[n]} \big|\mathcal{Z}_n(i) - \mathbb{E}_W[\mathcal{Z}_n(i)]\big| > \eta u_n\Big) = 0. \tag{6.4}$$

Similarly, when $\alpha \in (1,2)$, for any $\eta > 0$,

$$\lim_{n\to\infty} \mathbb{P}\Big(\max_{i\in[n]} \big|\mathcal{Z}_n(i) - \mathbb{E}_W[\mathcal{Z}_n(i)]\big| > \eta n\Big) = 0. \tag{6.5}$$

Finally, again assume the vertex-weights satisfy the (**Gumbel**)*-*(**RV**) *sub-case. Then, for any $\tau \in (0,1]$ and $\eta > 0$,*

$$\lim_{n\to\infty} \mathbb{P}\Big(\big|\max_{i\in[n]} \mathcal{Z}_n(i) - \max_{i\in[n]} \mathbb{E}_W[\mathcal{Z}_n(i)]\big| \geq \eta a_n \log n \log\log n\Big) = 0. \tag{6.6}$$

Also, let ℓ be a strictly positive function such that $\lim_{n\to\infty} \log(\ell(n))^2/\log n = \zeta_0$ for some $\zeta_0 \in [0,\infty)$. Recall $C_n(\gamma,s,t,\ell)$ from (2.20). Then, for any $0 < s < t < \infty, \tau \in (0,1)$ and $\eta > 0$,

$$\lim_{n\to\infty} \mathbb{P}\Big(\big|\max_{i\in C_n(\gamma,s,t,\ell)} \mathcal{Z}_n(i) - \max_{i\in C_n(\gamma,s,t,\ell)} \mathbb{E}_W[\mathcal{Z}_n(i)]\big| \geq \eta a_n \log n\Big) = 0. \tag{6.7}$$

Remark 6.2. The first five results of Proposition 6.1, as in (6.1)-(6.5) directly imply the concentration of the maximum degree due to the reversed triangle inequality for the supremum norm. That is, for any $I_n \subseteq [n]$,

$$\big|\max_{i\in I_n} \mathcal{Z}_n(i) - \max_{i\in I_n} \mathbb{E}_W[\mathcal{Z}_n(i)]\big| \leq \max_{i\in I_n} \big|\mathcal{Z}_n(i) - \mathbb{E}_W[\mathcal{Z}_n(i)]\big|.$$

Proof. We provide a proof for $m = 1$, as the proof for $m > 1$ follows in the same way. As mentioned above the statement of Proposition 6.1, the aim is to provide large deviation bounds for the quantities $|\mathcal{Z}_n(i) - \mathbb{E}_W[\mathcal{Z}_n(i)]|$ for $i \in [n]$, using that $\mathcal{Z}_n(i)$ is a sum of independent indicator random variables, combined with union bounds. For the results in (6.1) through (6.5), rather crude bounds suffice. To prove the more subtle results for the (**Gumbel**)-(**RV**) sub-case in (6.6) and (6.7), where the deviations around the maximum conditional mean are of smaller order compared to (6.2), more careful union bounds and large deviation bounds are provided.

Concentration under first-order scaling, convergence in probability. We start by proving the results in which the degrees are scaled by the first order growth-rate and the convergence holds in probability, as in (6.1), (6.3), (6.4) and (6.5). By using a large deviation bound for a sum of independent Bernoulli random variables (recall that $\mathcal{Z}_n(i)$ is such a sum), see e.g. [16, Theorem 2.21], we obtain

$$\begin{aligned}\mathbb{P}_W\big(\big|\mathcal{Z}_n(i) - \mathbb{E}_W[\mathcal{Z}_n(i)]\big| \geq g_n\big) &\leq 2\exp\Big\{-\frac{g_n^2}{2(\mathbb{E}_W[\mathcal{Z}_n(i)] + g_n)}\Big\}\\ &\leq 2\exp\Big\{-\frac{g_n^2}{2(\max_{i\in[n]} \mathbb{E}_W[\mathcal{Z}_n(i)] + g_n)}\Big\},\end{aligned} \tag{6.8}$$

for any non-negative sequence $(g_n)_{n\in\mathbb{N}}$. We start by considering (6.1), so that $g_n = \eta b_n \log n$. Hence, the fraction on the right-hand side is $g_n B_n$ for some random variable B_n that converges in probability to some positive constant (see Propositions 5.2). Using a union bound then yields

$$\begin{aligned}\mathbb{P}_W\Big(\max_{i\in[n]} |\mathcal{Z}_n(i) - \mathbb{E}_W[\mathcal{Z}_n(i)]| \geq g_n\Big) &\leq \sum_{i=1}^n 2\exp\{-g_n B_n\}\\ &= 2\exp\{\log n(1 - (g_n/\log n)B_n)\},\end{aligned} \tag{6.9}$$

As $g_n/\log n = \eta b_n$ diverges with n, it follows that this expression tends to zero in probability. For (6.3), we can also use (6.8) with $g_n = \eta a_{t_n n}\log(1/t_n)$ and we can write

$$\begin{aligned}&\frac{g_n^2}{2(\max_{i\in[n]}\mathbb{E}_W[\mathcal{Z}_n(i)]+g_n)}\\&=\frac{(\eta a_{t_n n}\log(1/t_n))^2}{2b_{t_n n}\log(1/t_n)(\max_{i\in[n]}\mathbb{E}_W[\mathcal{Z}_n(i)]/(b_{t_n n}\log(1/t_n))+\eta a_{t_n n}/b_{t_n n})}\\&=\frac{\eta^2}{2}\frac{a_{t_n n}^2\log(1/t_n)}{b_{t_n n}}B_n,\end{aligned}$$

where B_n converges in probability to a positive constant (see the proof of Proposition 5.5 and use the definition of a_n and b_n in Remark 2.4). Since $b_n/b_{t_n n}\to \mathrm{e}$ (again see the proof of Proposition 5.5) and by the definition of a_n and b_n in the (**Gumbel**)-(**RaV**) sub-case, it follows that the right-hand side is at least $Cb_n(\log n)^{1/\tau-1}$ with high probability, for some small constant $C>0$. Replacing g_n with $Cb_n(\log n)^{1/\tau-1}$, which grows faster than $\log n$, in (6.9) then yields the desired result. Finally, for (6.4) and (6.5), the same approach is valid with $g_n = u_n$ and $g_n = n$, respectively, though B_n now converges in distribution to some random variable (see Proposition 5.6). Still, it follows that $1-\eta^2(g_n/\log n)B_n<0$ with high probability, so that the right-hand side of (6.9) still converges to zero in probability. Then, in all the above cases, using the dominated convergence theorem yields (6.1), (6.3), (6.4) and (6.5).

Concentration under first order scaling, almost sure convergence. We now turn to the almost sure result for the (**Gumbel**)-(**RV**) sub-case, as in (6.2). Similar to (6.9), we write for any $\eta>0$,

$$\mathbb{P}_W\Big(\max_{i\in[n]}|\mathcal{Z}_n(i)-\mathbb{E}_W[\mathcal{Z}_n(i)]|\geq \eta g_n\Big)\leq 2\exp\Big\{\log n-\frac{\eta^2 g_n^2}{2(\max_{i\in[n]}\mathbb{E}_W[\mathcal{Z}_n(i)]+\eta g_n)}\Big\},$$

where $g_n = b_n\log n$. By Proposition 5.4, we can almost surely bound this from above by

$$2\exp\{\log n-\eta^2 Cb_n\log n\}\leq 2\exp\Big\{-\frac12\eta^2Cb_n\log n\Big\},$$

for some sufficiently small constant $C>0$ and when n is at least $N\in\mathbb{N}$, for some random N. Thus, we can conclude that this upper bound is almost surely summable in n, as b_n tends to infinity with n. Since η is arbitrary, the $\mathbb{P}_W$-almost sure convergence to 0 is established. Then, since

$$\begin{aligned}&\mathbb{P}\Big(\forall\eta>0\ \exists N\in\mathbb{N}\ \forall n\geq N:\max_{i\in[n]}|\mathcal{Z}_n(i)-\mathbb{E}_W[\mathcal{Z}_n(i)]|<\eta g_n\Big)\\&=\mathbb{E}\Big[\mathbb{P}_W\Big(\forall\eta>0\ \exists N\in\mathbb{N}\ \forall n\geq N:\max_{i\in[n]}|\mathcal{Z}_n(i)-\mathbb{E}_W[\mathcal{Z}_n(i)]|<\eta g_n\Big)\Big]\\&=\mathbb{E}[1]=1,\end{aligned}\tag{6.10}$$

the $\mathbb{P}$-almost sure convergence follows as well.

Concentration under second order scaling. We now prove (6.6), which holds when the following two claims are true:

$$\begin{aligned}&\mathbb{P}_W\Big(\max_{i\in[n]}\mathcal{Z}_n(i)\geq\max_{i\in[n]}\mathbb{E}_W[\mathcal{Z}_n(i)]+\eta a_n\log n\log\log n\Big)\xrightarrow{\mathbb{P}}0,\\&\mathbb{P}_W\Big(\max_{i\in[n]}\mathcal{Z}_n(i)\leq\max_{i\in[n]}\mathbb{E}_W[\mathcal{Z}_n(i)]-\eta a_n\log n\log\log n\Big)\xrightarrow{\mathbb{P}}0.\end{aligned}\tag{6.11}$$

As it turns out, we can prove the second claim with relative ease compared to the first claim, so we defer it to the end of the proof. We first focus on proving the first line of (6.11). Moreover, we first provide a proof for the case that $\tau\in(0,3/4)$, on which we base a more involved proof for all $\tau\in(0,1]$.

Let $\tau \in (0,3/4)$. We use a union bound for the first line of (6.11) and split the sum into two sets, defined as

$$C_n^1 := \{i \in [n] : W_i < (1-\sqrt{\varepsilon_n})b_i\}, \qquad C_n^2 := \{i \in [n] : W_i \geq (1-\sqrt{\varepsilon_n})b_i\},$$

where $\varepsilon_n = (\log n)^{-c}$, for some $c \in (0,2)$ to be determined later on. The size of C_n^2 can be controlled well enough, so that a precise union bound can be applied to this part. For the other set, we claim that with high probability,

$$C_n^1 \subseteq \{i \in [n] : \mathbb{E}_W[\mathcal{Z}_n(i)] \leq (1-\sqrt{\varepsilon_n}) \max_{i\in[n]} \mathbb{E}_W[\mathcal{Z}_n(i)]\} =: \widetilde{C}_n^1. \tag{6.12}$$

Then, on the event that $C_n^1 \subseteq \widetilde{C}_n^1$, we are able to manipulate terms in the probability to such an extent that we obtain an improved large deviation bound and show this part converges to zero in probability as well.

Let us start by showing the with high probability inclusion of C_n^1 in $\widetilde{C}_n^1$. Take $i \in C_n^1$. Then,

$$\mathbb{E}_W[\mathcal{Z}_n(i)] = W_i \sum_{j=i}^{n-1} \frac{1}{S_j} \leq (1-\sqrt{\varepsilon_n})b_i \log(n/i)\Big(1 + \frac{|Y_n - Y_i|}{\log(n/i)}\Big),$$

where $Y_n := \sum_{j=1}^{n-1} 1/S_j - \log n$. Furthermore, by Proposition 5.4, (5.11), with high probability

$$\max_{i\in[n]} \mathbb{E}_W[\mathcal{Z}_n(i)] \geq (1-\gamma)b_{n^\gamma} \log n\Big(1 + \frac{1/2-\eta}{1-\gamma}\frac{\log\log n}{\log n}\Big),$$

for any fixed $\eta > 0$. If thus suffices to show that when $i \in C_n^1$, then with high probability

$$b_i \log(n/i)\Big(1 + \frac{|Y_n - Y_i|}{\log(n/i)}\Big) \leq (1-\gamma)b_{n^\gamma} \log n\Big(1 + \frac{1/2-\eta}{1-\gamma}\frac{\log\log n}{\log n}\Big) \tag{6.13}$$

is satisfied for some $\eta > 0$. We show a stronger statement, namely that (6.13) holds with high probability for any $i \in [n]$. We recall from (5.2) that Y_n converges almost surely. In particular, we have that, with high probability, $\max_{i\in[n]} Y_i \leq (\log\log n)^{1/2}$ and $\max_{\log\log n \leq i \leq n} |Y_i - Y_n|$ converges to 0 in probability. Note first that for $i \leq (\log\log n)$,

$$b_i \log(n/i)\Big(1 + \frac{|Y_n - Y_i|}{\log(n/i)}\Big) \leq (1-\gamma)b_{n^\gamma} \log n\Big(1 + \frac{1/2-\eta}{1-\gamma}\frac{\log\log n}{\log n}\Big).$$

Next we consider $\log\log n \leq i \leq n^{\gamma-\varepsilon}$ or $i \geq n^{\gamma+\varepsilon}$ for some $\varepsilon > 0$, when we get that with high probability

$$b_i \log(n/i)\Big(1 + \frac{|Y_n - Y_i|}{\log(n/i)}\Big) \leq C(1-\gamma)b_{n^\gamma} \log n,$$

for some $C \in (0,1)$, as follows from the proof of Proposition 5.4 ((5.15) to be more precise), so that (6.13) is satisfied. It remains to prove that (6.13) is satisfied with high probability when $i = n^\gamma k_n$, where k_n is sub-polynomial, in the sense that $|\log k_n|/\log n \to 0$. First, as before, with high probability

$$1 + \frac{|Y_n - Y_i|}{\log(n/i)} \leq 1 + \eta\frac{\log\log(n^\gamma)}{\log n},$$

for any constant $\eta > 0$. Moreover, by Remark 2.4, for any $\eta > 0$,

$$(1-\gamma)b_{n^\gamma} \log n \geq c_1(1-\gamma)\log n \log(n^\gamma)^{1/\tau}\Big(1 + \frac{(b/\tau - \eta)\log\log(n^\gamma)}{(1-\gamma)\log n}\Big),$$

for all n large. Finally, for n large,

$$\begin{aligned}
b_i\log(n/i) &= c_1(1-\gamma)\log n\log(n^\gamma)^{1/\tau}\Big(1+\frac{(b/\tau)\log\log(n^\gamma k_n)+b\log c_1+\log\tau}{(1-\gamma)\log n}\Big)\\
&\quad\times\Big(1-\frac{\log k_n}{(1-\gamma)\log n}\Big)\Big(1+\frac{\log k_n}{\gamma\log n}\Big)^{1/\tau}\\
&\le c_1(1-\gamma)\log n\log(n^\gamma)^{1/\tau}\Big(1+\frac{(b/\tau)\log\log(n^{\gamma\pm\eta})}{(1-\gamma)\log n}\Big)\\
&\le c_1(1-\gamma)\log n\log(n^\gamma)^{1/\tau}\Big(1+\frac{(b/\tau+\eta)\log\log(n^\gamma)}{(1-\gamma)\log n}\Big)
\end{aligned}$$

as $(1-x/(1-\gamma))(1+x/\gamma)^{1/\tau}\le 1$ for $x\in[0,1-\gamma]$ and where the $\pm$ sign depends on the sign of b. Combining all of the above, we find for n large and with high probability,

$$\begin{aligned}
&b_i\log(n/i)\Big(1+\frac{|Y_n-Y_i|}{\log(n/i)}\Big)\\
&\quad\le c_1(1-\gamma)\log n\log(n^\gamma)^{1/\tau}\Big(1+\frac{(b/\tau+\eta)\log\log(n^\gamma)}{(1-\gamma)\log n}\Big)\Big(1+\eta\frac{\log\log(n^\gamma)}{\log n}\Big)\\
&\quad\le c_1(1-\gamma)\log n\log(n^\gamma)^{1/\tau}\Big(1+\Big(\frac{b/\tau+\eta}{1-\gamma}+2\eta\Big)\frac{\log\log(n^\gamma)}{\log n}\Big),
\end{aligned}$$

and

$$\begin{aligned}
&(1-\gamma)b_{n^\gamma}\log n\Big(1+\frac{1/2-\eta}{1-\gamma}\frac{\log\log n}{\log n}\Big)\\
&\quad\ge c_1(1-\gamma)\log n\log(n^\gamma)^{1/\tau}\Big(1+\Big(\frac{b/\tau-\eta}{1-\gamma}+\frac{1/2-\eta}{1-\gamma}-\eta\Big)\frac{\log\log(n^\gamma)}{\log n}\Big),
\end{aligned}$$

so that (6.13) is established with high probability for all $i\in[n-1]$, in particular for all $i\in C_n^1$, when η is sufficiently small.

As a second step, we control the size of C_n^2. First, we fix an $\eta>0$ and set $I=\max\{I_1,I_2,I_3\}$, where $I_1,I_2,I_3\in\mathbb{N}$ are such that

$$\begin{aligned}
\mathbb{P}(W\ge(1-\sqrt{\varepsilon_n})b_i)&\le(1+\eta)a((1-\sqrt{\varepsilon_n})b_i)^b\exp\{-((1-\sqrt{\varepsilon_n})b_i/c_1)^\tau\}, && i\ge I_1,\\
\mathbb{P}(W\ge b_i)&\ge(1-\eta)ab_i^b\exp\{-(b_i/c_1)^\tau\}, && i\ge I_2,\\
\mathbb{P}(W\ge b_i)&\le(1+\eta)/i, && i\ge I_3.
\end{aligned}$$

We note that I is well-defined, as b_i is (eventually) increasing in i and diverges with i and as b_i is such that $\lim_{i\to\infty}\mathbb{P}(W\ge b_i)\,i=1$. We then arrive at

$$\mathbb{E}\left[|C_n^2|\right]\le I+\sum_{i=I}^n(1+\eta)a((1-\sqrt{\varepsilon_n})b_i)^b\exp\{-((1-\sqrt{\varepsilon_n})b_i/c_1)^\tau\}.$$

By writing $(1-\sqrt{\varepsilon_n})^\tau=1-\tau\sqrt{\varepsilon_n}(1+o(1))$, and as we can bound $(1-\sqrt{\varepsilon_n})^b$ from above by some sufficiently large constant C which depends only on b, we obtain

$$\begin{aligned}
\mathbb{E}\left[|C_n^2|\right]&\le I+(1+\eta)C\sum_{i=I}^n ab_i^b\exp\{-(b_i/c_1)^\tau\}\exp\{\tau\sqrt{\varepsilon_n}(b_i/c_1)^\tau(1+o(1))\}\\
&\le I+(1+\eta)^2(1-\eta)^{-1}C\exp\{\tau\sqrt{\varepsilon_n}(b_n/c_1)^\tau(1+o(1))\}\sum_{i=I}^n i^{-1}\\
&\le I+(1+\eta)^2(1-\eta)^{-1}C\exp\{\tau\sqrt{\varepsilon_n}\log n(1+o(1))\}\log n\\
&= I+(1+\eta)^2(1-\eta)^{-1}C\exp\{\tau\sqrt{\varepsilon_n}\log n(1+o(1))\}.
\end{aligned}$$

Since $\sqrt{\varepsilon_n}\log n=(\log n)^{1-c/2}$ and $c\in(0,2)$, it follows by Markov's inequality that for any $C_3>\tau$,

$$|C_n^2|\exp\{-C_3(\log n)^{1-c/2}\}\xrightarrow{\mathbb{P}}0. \tag{6.14}$$

We are now ready to prove the first line of (6.11). We use the sets $C_n^1, \widetilde{C}_n^1$ and C_n^2, in particular (6.12) and (6.14) to obtain

$$
\begin{aligned}
&\mathbb{P}_W\Big(\max_{i\in[n]} \mathcal{Z}_n(i) \geq \max_{i\in[n]} \mathbb{E}_W[\mathcal{Z}_n(i)] + \eta a_n \log n \log\log n\Big)\\
&\quad\leq \mathbb{P}_W\Big(\Big\{\max_{i\in[n]} \mathcal{Z}_n(i) \geq \max_{i\in[n]} \mathbb{E}_W[\mathcal{Z}_n(i)] + \eta a_n \log n \log\log n\Big\} \cap \{C_n^1 \subseteq \widetilde{C}_n^1\}\Big)\\
&\qquad + \mathbb{P}\Big(C_n^1 \not\subseteq \widetilde{C}_n^1\Big)\\
&\quad\leq \sum_{i\in C_n^1} \mathbb{P}_W\Big(\{\mathcal{Z}_n(i) \geq \max_{i\in[n]} \mathbb{E}_W[\mathcal{Z}_n(i)] + \eta a_n \log n \log\log n\} \cap \{C_n^1 \subseteq \widetilde{C}_n^1\}\Big)\\
&\qquad + \sum_{i\in C_n^2} \mathbb{P}_W\Big(\mathcal{Z}_n(i) \geq \max_{i\in[n]} \mathbb{E}_W[\mathcal{Z}_n(i)] + \eta a_n \log n \log\log n\Big) + \mathbb{P}\Big(C_n^1 \not\subseteq \widetilde{C}_n^1\Big).
\end{aligned}
$$

As established in (6.12), the third probability converges to zero with n. For the first probability we use that on $\{C_n^1 \subseteq \widetilde{C}_n^1\}$, $\mathbb{E}_W[\mathcal{Z}_n(i)] \leq (1-\sqrt{\varepsilon_n})\max_{i\in[n]} \mathbb{E}_W[\mathcal{Z}_n(i)]$, and for the second probability we use that $\max_{i\in[n]} \mathbb{E}_W[\mathcal{Z}_n(i)] \geq \mathbb{E}_W[\mathcal{Z}_n(i)]$ for any $i \in [n]$ to find the upper bound

$$
\begin{aligned}
&\sum_{i\in C_n^1} \mathbb{P}_W\Big(\mathcal{Z}_n(i) - \mathbb{E}_W[\mathcal{Z}_n(i)] \geq \sqrt{\varepsilon_n}\max_{i\in[n]} \mathbb{E}_W[\mathcal{Z}_n(i)] + \eta a_n \log n \log\log n\Big)\\
&\qquad + \sum_{i\in C_n^2} \mathbb{P}_W(\mathcal{Z}_n(i) - \mathbb{E}_W[\mathcal{Z}_n(i)] \geq \eta a_n \log n \log\log n) + o(1).
\end{aligned} \tag{6.15}
$$

Now, applying a large deviation bound to (6.15) yields

$$
\begin{aligned}
&\sum_{i\in C_n^1} \exp\Big\{-\frac{(\sqrt{\varepsilon_n}\max_{i\in[n]} \mathbb{E}_W[\mathcal{Z}_n(i)] + \eta a_n \log n \log\log n)^2}{2(\mathbb{E}_W[\mathcal{Z}_n(i)] + \sqrt{\varepsilon_n}\max_{i\in[n]} \mathbb{E}_W[\mathcal{Z}_n(i)] + \eta a_n \log n \log\log n)}\Big\}\\
&\qquad + \sum_{i\in C_n^2} \exp\Big\{-\frac{(\eta a_n \log n \log\log n)^2}{2(\mathbb{E}_W[\mathcal{Z}_n(i)] + \eta a_n \log n \log\log n)}\Big\} + o(1).
\end{aligned} \tag{6.16}
$$

In both exponents we bound the conditional mean in the denominator by the maximum conditional mean. This yields the upper bound

$$
n\exp\{-\varepsilon_n b_n \log n A_n\} + |C_n^2|\exp\Big\{-\frac{\eta^2 a_n^2 \log n(\log\log n)^2}{2b_n}B_n\Big\},
$$

where both A_n, B_n converge in probability to positive constants. We now set $c < 1/\tau$ to ensure that $\varepsilon_n b_n$ diverges, so that the first term converges to zero in probability. Thus, $c < (1/\tau)\wedge 2$ is required. We can write the second term as

$$
|C_n^2|\exp\{-C_3(\log n)^{1-c/2}\}\exp\{C_3(\log n)^{1-c/2} - (\log n)^{1/\tau-1}(\log\log n)^2\widetilde{B}_n\},
$$

where $\widetilde{B}_n$ converges in probability to a positive constant. Now, by (6.14), the product of the first two terms converges to zero in probability and the last term converges to zero in probability when $1-c/2 < 1/\tau - 1$, or $c > 4-2/\tau$. We thus find that (6.11) is established when we can find a $c \in (0,2)$ such that $4-2/\tau < c < 1/\tau$, which holds for all $\tau \in (0,3/4)$.

We now extend this approach so that the first line of (6.11) can be achieved for all $\tau \in (0,1]$. To this end, we define the sequence $(p_k)_{k\in\mathbb{N}}$ as $p_k := (3/4)p_{k-1} + 1/(4c\tau), k \geq 1$, and $p_0 = 1/2$. We solve the recursion to obtain

$$
p_k = \frac{1}{c\tau} - \Big(\frac{1}{c\tau} - \frac{1}{2}\Big)\Big(\frac{3}{4}\Big)^k, \tag{6.17}
$$

from which it immediately follows that p_k is increasing when $c < (1/\tau)\wedge 2$. Moreover, we can rewrite the recursion as $p_k = p_{k-1}/2 + (p_{k-1}/2 + 1/(2c\tau))/2$, so that

$$
p_k \in (p_{k-1}, p_{k-1}/2 + 1/(2c\tau)), \qquad k \geq 1. \tag{6.18}
$$

We also define, for some $K\in\mathbb{N}_0$ to be specified later, the sets

$$\begin{aligned} C_n^1 &:= \{i\in[n] : W_i<(1-\varepsilon_n^{p_0})b_i\},\\ C_n^k &:= \{i\in[n] : W_i\in[(1-\varepsilon_n^{p_{k-1}})b_i,(1-\varepsilon_n^{p_k})b_i)\}, && k\in\{2,\dots,K\},\\ \widetilde{C}_n^k &:= \{i\in[n] : \mathbb{E}_W[\mathcal{Z}_n(i)]\leq(1-\varepsilon_n^{p_k})\max_{i\in[n]}\mathbb{E}_W[\mathcal{Z}_n(i)]\}, && k\in[K],\\ C_n^{K+1} &:= \{i\in[n] : W_i\geq(1-\varepsilon_n^{p_K})b_i\}. \end{aligned}$$

By the same argument as provided for the proof of (6.12), it follows that for any fixed $k\in\mathbb{N}$, with high probability $C_n^k\subseteq\widetilde{C}_n^k$. Similar to the approach for $\tau\in(0,3/4)$ we can then bound

$$\begin{aligned} &\mathbb{P}_W\Big(\max_{i\in[n]}\mathcal{Z}_n(i)\geq\max_{i\in[n]}\mathbb{E}_W[\mathcal{Z}_n(i)]+\eta a_n\log n\log\log n\Big)\\ &\leq\sum_{i\in C_n^1}\mathbb{P}_W\Big(\mathcal{Z}_n(i)-\mathbb{E}_W[\mathcal{Z}_n(i)]\geq\varepsilon_n^{p_0}\max_{i\in[n]}\mathbb{E}_W[\mathcal{Z}_n(i)]+\eta a_n\log n\log\log n\Big)\\ &\quad+\sum_{k=2}^K\sum_{i\in C_n^k}\mathbb{P}_W\Big(\mathcal{Z}_n(i)-\mathbb{E}_W[\mathcal{Z}_n(i)]\geq\varepsilon_n^{p_k}\max_{i\in[n]}\mathbb{E}_W[\mathcal{Z}_n(i)]+\eta a_n\log n\log\log n\Big)\\ &\quad+\sum_{i\in C_n^{K+1}}\mathbb{P}_W\Big(\mathcal{Z}_n(i)-\mathbb{E}_W[\mathcal{Z}_n(i)]\geq\eta a_n\log n\log\log n\Big)+\mathbb{P}\Big(\bigcup_{k=1}^K\{C_n^k\not\subseteq\widetilde{C}_n^k\}\Big). \end{aligned}\tag{6.19}$$

We do not include the sum over $i\in C_n^1$ in the double sum, as the upper bound we use is slightly different for these terms. The last term converges to zero by using a union bound, as established above (6.19) and since K is fixed. As in the simplified proof for $\tau\in(0,3/4)$ where $K=0$, we require $cp_k<1$ for all $k\in\{0,1,\dots,K\}$, so that $\eta a_n\log n\log\log n$ is negligible compared to $\varepsilon_n^{p_k}\max_{i\in[n]}\mathbb{E}_W[\mathcal{Z}_n(i)]$. Since p_k is increasing, $cp_K<1$ suffices. Using (6.17) yields that K cannot be too large, i.e. we need

$$\Big(\frac{3}{4}\Big)^K>\frac{1}{c}\Big(\frac{1}{\tau}-1\Big)\Big(\frac{1}{c\tau}-\frac{1}{2}\Big)^{-1}.\tag{6.20}$$

We now again apply a large deviation bound as in (6.16). Furthermore, with an equivalent approach that led to (6.14), we find with high probability an upper bound for (6.19) of the form

$$\begin{aligned} &n\exp\{-\varepsilon_n^{2p_0}b_n\log nA_{0,n}\}+\sum_{k=1}^K C\exp\{C_k(\log n)^{1-cp_{k-1}}-\eta^2\varepsilon_n^{2p_k}b_n\log nA_{k,n}\}\\ &\quad+C\exp\Big\{C_{K+1}(\log n)^{1-cp_K}-\eta^2\frac{a_n^2\log n(\log\log n)^2}{b_n}A_{K+1,n}\Big\}+o(1), \end{aligned}\tag{6.21}$$

where $C,C_1,\dots,C_{K+1}>0$ are suitable constants and the $A_{k,n}$ are random variables which converge in probability to some strictly positive constants $A_k,k\in\{0,1,\dots,K+1\}$. In order for all these terms to converge to zero in probability, the following conditions need to be met:

$$1/\tau-2p_0c>0,\qquad 1-cp_{k-1}<-2cp_k+1/\tau+1,\quad k\in[K],\qquad 1-cp_K\leq1/\tau-1.\tag{6.22}$$

Since $p_0=1/2$, it follows that $c<(1/\tau)\wedge 2$ still needs to be satisfied. By the bounds on p_k in (6.18) the second condition is satisfied and the final condition holds when $p_K\geq(2-1/\tau)/c$, or

$$\Big(\frac{3}{4}\Big)^K\leq 2\frac{1}{c}\Big(\frac{1}{\tau}-1\Big)\Big(\frac{1}{c\tau}-\frac{1}{2}\Big)^{-1}.$$

Together with (6.20) this yields

$$\frac{1}{c}\Big(\frac{1}{\tau}-1\Big)\Big(\frac{1}{c\tau}-\frac{1}{2}\Big)^{-1}<\Big(\frac{3}{4}\Big)^K\leq\frac{2}{c}\Big(\frac{1}{\tau}-1\Big)\Big(\frac{1}{c\tau}-\frac{1}{2}\Big)^{-1}.$$

Since the ratio of the lower and upper bound is exactly 2, such a $K \in \mathbb{N}_0$ can always be found, as long as

$$\frac{1}{c}\Big(\frac{1}{\tau}-1\Big)\Big(\frac{1}{c\tau}-\frac{1}{2}\Big)^{-1}<1,$$

which is satisfied for any $\tau \in (0,1)$ when $c<2$. It thus follows that (6.11) holds for all $\tau \in (0,1)$. When $\tau=1$, the condition $p_K \leq (2-1/\tau)c$ simplifies to $p_K \geq 1/c$, which, together with the condition $p_K < 1/c$ implies that $p_K = 1/c$ is required. However, when $\tau = 1$, the limit of p_K is $1/c$, so that K needs to tend to infinity with n. Therefore, we repeat the same arguments, but now take $K = K(n) = \lceil 1/|\log(3/4)|(\log\log\log n - \log\log\log\log n)\rceil$. We then need to check the following things:

(i) The conditions on p_k and c are met.
(ii) The final probability in (6.19) converges to zero in probability with n.
(iii) All terms in (6.21) individually converge to zero with n, as well as when summing them all together.

The first two conditions in (6.22) still need to be satisfied, and this is the case when K grows with n as well. Furthermore, as $\tau = 1$, $p_k < 1/c$ is satisfied for all $k \in \mathbb{N}$, establishing (i).

For (ii), we observe that by (6.13),

$$\begin{aligned}&\mathbb{P}\Big(\bigcup_{k=1}^{K}\{C_n^k \not\subseteq \widetilde{C}_n^k\}\Big)\\ &\leq \mathbb{P}\Bigg(\bigcup_{i=1}^{n-1}\Big\{b_i\log(n/i)\Big(1+\frac{|Y_n-Y_i|}{\log(n/i)}\Big) > (1-\gamma)b_{n^\gamma}\log n\Big(1+\frac{1/2-\eta}{\tau}\frac{\log\log n}{\log n}\Big)\Big\}\Bigg),\end{aligned}$$

for some small $\eta > 0$, and the decay of this probability to zero has been established in (6.13).

Finally, for (iii), we check the convergence of the terms in (6.21). The first term clearly still converges to zero in probability. Then, for each term in the sum we note that the constants C_k can all be chosen such that $C_k < \tau + \delta$ for all $k \in [K+1]$ and any $\delta > 0$. Similarly, the random terms $A_{k,n}$, which converge in probability to positive constants A_k, can also be shown to be bounded away from zero uniformly in k. This yields that we need only consider the rate of divergence of the remaining terms. We write,

$$\begin{aligned}(\log n)^{1-cp_{k-1}} &= (\log n)^{(2-c)(2/3)(3/4)^k} = \exp\Big\{\frac{2}{3}(2-c)\exp\{k\log(3/4)+\log\log\log n\}\Big\},\\ \varepsilon_n^{2p_k}b_n\log n &\sim (\log n)^{(2-c)(3/4)^k} = \exp\{(2-c)\exp\{k\log(3/4)+\log\log\log n\}\},\end{aligned}$$

where we recall that $\tau = 1$ and thus the expression of p_k is simplified. We note that both terms diverge with n for each $k \in [K]$ by the choice of K and since $\log(3/4) > -1$. Moreover, the latter term is dominant for every $k \in [K]$, so that each term in the sum in (6.21) tends to zero in probability. An upper bound for the entire sum is established when setting $p_{k-1} = p_{K-1}, p_k = p_K$ and bounding (with high probability) $C_k < \tau + \delta$ $A_{k,n} > \delta$, for some small $\delta > 0$ uniformly in k. We then obtain the upper bound

$$C\exp\big\{\log K - \eta^2\delta(\log\log n)^{3(2-c)/4}(1+o(1))\big\},$$

which converges to zero as $\log K$ is negligible compared to the double logarithmic term. Now, for the final term in (6.21), we write as before,

$$(\log n)^{1-cp_K} = \exp\{(1-c/2)\exp\{K\log(3/4)+\log\log\log n\}\} \leq \mathrm{e}^{(1-c/2)\log\log\log n},$$

and $(\log\log n)^2 = \exp\{2\log\log\log n\}$ so that the latter term dominates the former, which yields the desired result.

What remains is to prove the second line of (6.11) holds for any $\tau \in (0,1]$ and $\eta > 0$. We note that the event in the brackets occurs when

$$\mathcal{Z}_n(i) \leq \max_{i\in[n]} \mathbb{E}_W[\mathcal{Z}_n(i)] - \eta a_n \log n \log\log n \quad \forall i \in [n],$$

so that

$$\begin{aligned}
\mathbb{P}_W&\Big(\max_{i\in[n]} \mathcal{Z}_n(i) \leq \max_{i\in[n]} \mathbb{E}_W[\mathcal{Z}_n(i)] - \eta a_n \log n \log\log n\Big)\\
&\leq \mathbb{P}_W\Big(\bigcap_{i\in[n]} \Big\{\mathcal{Z}_n(i) \leq \max_{i\in[n]} \mathbb{E}_W[\mathcal{Z}_n(i)] - \eta a_n \log n \log\log n\Big\}\Big)\\
&\leq \mathbb{P}_W\Big(\mathcal{Z}_n(\widetilde{I}_n) \leq \max_{i\in[n]} \mathbb{E}_W[\mathcal{Z}_n(i)] - \eta a_n \log n \log\log n\Big),
\end{aligned}$$

where we recall that $\widetilde{I}_n := \inf\{i \in [n] : \mathbb{E}_W[\mathcal{Z}_n(i)] \geq \mathbb{E}_W[\mathcal{Z}_n(j)]$ for all $j \in [n]\}$. Since $\widetilde{I}_n$ is determined by $W_1, \ldots, W_n$, it follows that $\max_{i\in[n]} \mathbb{E}_W[\mathcal{Z}_n(i)] = \mathbb{E}_W[\mathcal{Z}_n(\widetilde{I}_n)]$. Thus,

$$\begin{aligned}
\mathbb{P}_W&\Big(\mathcal{Z}_n(\widetilde{I}_n) \leq \max_{i\in[n]} \mathbb{E}_W[\mathcal{Z}_n(i)] - \eta a_n \log n \log\log n\Big)\\
&\leq \mathbb{P}_W\Big(\big|\mathcal{Z}_n(\widetilde{I}_n) - \mathbb{E}_W[\mathcal{Z}_n(\widetilde{I}_n)]\big| \geq \eta a_n \log n \log\log n\Big)\\
&\leq \frac{\mathrm{Var}_W(\mathcal{Z}_n(\widetilde{I}_n))}{(\eta a_n \log n \log\log n)^2}\\
&\leq \frac{\max_{i\in[n]} \mathbb{E}_W[\mathcal{Z}_n(i)]}{b_n \log n} \frac{b_n}{\eta^2 a_n^2 \log n(\log\log n)^2},
\end{aligned}$$

which converges to zero almost surely, as the first fraction on the right-hand side converges to a positive constant almost surely and the second fraction converges to zero, since $\tau \in (0,1]$. Therefore, the second statement in (6.11) holds and combining this with the first statement of (6.11) and the dominated convergence theorem concludes the proof of (6.6).

Finally, (6.7) can be proved in a similar way as (6.6), though the case $\tau = 1$ no longer holds due to the absence of the $\log\log n$ term. □

7. Proof of the main theorems

We now prove the main theorems, Theorem 2.9, 2.12 and 2.14. For clarity, we split the proof of Theorem 2.9 into three parts, dealing with the **(Bounded)**, **(Gumbel)** and **(Fréchet)** cases separately, which all use somewhat different approaches. In all cases, the proof also holds for the model with a *random out-degree* as discussed in Remark 2.2(*ii*) when setting $m = 1$, as in this model the in-degree $\mathcal{Z}_n(i)$ of each vertex $i \in [n]$ can still be written as a sum of independent indicator random variables.

7.1. Proof of Theorem 2.9, Bounded case.

Before we prove the **(Bounded)** case of Theorem 2.9, we state an adaptation of [8, Lemma 1]:

Lemma 7.1. *Let $A_{n,i} := \{\mathcal{Z}_n(i) \geq a_n\}$ for some sequence $(a_n)_{n\in\mathbb{N}}$. Then,*

$$\mathbb{P}_W\left(\bigcup_{i=1}^n A_{n,i}\right) \leq \sum_{i=1}^n \mathbb{P}_W(A_{n,i}), \qquad \mathbb{P}_W\left(\bigcup_{i=1}^n A_{n,i}\right) \geq \frac{\sum_{i=1}^n \mathbb{P}_W(A_{n,i})}{1 + \sum_{i=1}^n \mathbb{P}_W(A_{n,i})},$$

and as a result,

$$\mathbb{P}_W\left(\bigcup_{i=1}^n A_{n,i}\right) \overset{\mathbb{P}/\text{a.s.}}{\longrightarrow} \begin{cases} 0, & \text{if } \sum_{i=1}^n \mathbb{P}_W(A_{n,i}) \overset{\mathbb{P}/\text{a.s.}}{\longrightarrow} 0,\\ 1, & \text{if } \sum_{i=1}^n \mathbb{P}_W(A_{n,i}) \overset{\mathbb{P}/\text{a.s.}}{\longrightarrow} \infty. \end{cases}$$

Proof. The result directly follows by applying [8, Lemma 1] to the conditional probability measure $\mathbb{P}_W$. □

*Proof of Theorem 2.9, (**Bounded**) case.* The proof heavily relies on the proof of [8, Theorem 2], which we adapt to work for WRGs. Before proving almost sure convergence, we prove convergence in probability. We do this by providing an upper and lower bound and show that these coincide. Then, using these bounds we prove almost sure convergence. Let us start with the upper bound. We set $a_n := c\log n$, with $c > 1/\log\theta_m$ and let $\varepsilon \in (0, \min\{m/\mathbb{E}[W] - c + c\log(c\mathbb{E}[W]/m), c\mathbb{E}[W]/(me^2), 1/2\})$. Note that the first argument of the minimum equals zero when $c = m/\mathbb{E}[W]$ and is positive otherwise. As $\theta_m = 1 + \mathbb{E}[W]/m$ and $c > 1/\log\theta_m > m/\mathbb{E}[W]$, this minimum is strictly positive. Then, we aim to show that

$$\sum_{i=1}^n \mathbb{P}_W(\mathcal{Z}_n(i) \geq a_n) \xrightarrow{a.s.} 0, \tag{7.1}$$

which implies via Lemma 7.1 and the dominated convergence theorem that

$$\mathbb{P}\Big(\max_{i\in[n]} \mathcal{Z}_n(i) \geq a_n\Big) \to 0. \tag{7.2}$$

Using a Chernoff bound and the fact that $\mathcal{Z}_n(i)$ is a sum of independent indicator random variables, we have for any $t > 0$,

$$\mathbb{P}_W(\mathcal{Z}_n(i) \geq a_n) \leq \mathrm{e}^{-ta_n} \prod_{j=i}^{n-1} \Big(\frac{W_i}{S_j}\mathrm{e}^t + \Big(1 - \frac{W_i}{S_j}\Big)\Big)^m \leq \mathrm{e}^{-ta_n + (\mathrm{e}^t - 1)mW_i(H_n - H_i)},$$

where $H_n := \sum_{j=1}^{n-1} 1/S_j$. This expression is minimised for $t = \log(a_n) - \log(mW_i(H_n - H_i))$, which yields the upper bound

$$\mathbb{P}_W(\mathcal{Z}_n(i) \geq a_n) \leq \mathrm{e}^{a_n(1 - u_i + \log u_i)}, \tag{7.3}$$

with $u_i = mW_i(H_n - H_i)/a_n$. We note that the mapping $x \mapsto 1 - x + \log x$ is increasing for $x \in (0,1)$. Moreover, by (5.2), $mH_n/a_n < 1$ holds almost surely for all sufficiently large n by the choice of c. Then, as we can bound W_i from above by 1 almost surely and $(H_n - H_i)$ is decreasing in i, we find, almost surely, for n large and uniformly in i,

$$\begin{aligned}
\mathbb{P}_W(\mathcal{Z}_n(i) \geq a_n) &\leq \exp\{a_n(1 - mH_n/a_n + \log(mH_n/a_n)\} \\
&= \exp\{c\log n(1 - m/(c\mathbb{E}[W]) + \log(m/(c\mathbb{E}[W])))(1 + o(1))\} \\
&= \exp\{-\log n(m/\mathbb{E}[W] - c + c\log(c\mathbb{E}[W]/m))(1 + o(1))\}.
\end{aligned}$$

Thus,

$$\sum_{i<n^\varepsilon} \mathbb{P}_W(\mathcal{Z}_n(i) \geq a_n) \leq \exp\{-\log n(m/\mathbb{E}[W] - c + c\log(c\mathbb{E}[W]/m) - \varepsilon)(1 + o(1))\}, \tag{7.4}$$

which tends to zero almost surely as $\varepsilon < m/\mathbb{E}[W] - c + c\log(c\mathbb{E}[W]/m)$. Similarly, again using that $W_i \leq 1, mH_n/a_n < 1$ almost surely for n large,

$$\sum_{i>n^{1-\varepsilon}} \mathbb{P}_W(\mathcal{Z}_n(i) \geq a_n) \leq n\exp\Big\{a_n\Big(1 - \frac{m(H_n - H_{\lceil n^{1-\varepsilon}\rceil})}{a_n} + \log\Big(\frac{m(H_n - H_{\lceil n^{1-\varepsilon}\rceil})}{a_n}\Big)\Big\}.$$

As $H_n - H_{\lceil n^{1-\varepsilon}\rceil} = \varepsilon\log n(1 + o(1))$ almost surely for n large,

$$\begin{aligned}
\sum_{i\geq n^{1-\varepsilon}} \mathbb{P}_W(\mathcal{Z}_n(i) \geq a_n) &\leq n\exp\Big\{c\log n\Big(1 - \frac{\varepsilon m}{c\mathbb{E}[W]} + \log\Big(\frac{\varepsilon m}{c\mathbb{E}[W]}\Big)\Big)(1 + o(1))\Big\} \\
&= n^{-(-c + \varepsilon m/\mathbb{E}[W] - c\log(\varepsilon m/(c\mathbb{E}[W])) - 1)(1 + o(1))},
\end{aligned} \tag{7.5}$$

which also tends to zero almost surely since $\varepsilon < c\mathbb{E}[W]/(me^2)$. It thus remains to prove that

$$\sum_{n^\varepsilon \leq i < n^{1-\varepsilon}} \mathbb{P}_W(\mathcal{Z}_n(i) \geq a_n) \xrightarrow{a.s.} 0. \tag{7.6}$$

We use the same bound as in (7.3), which holds uniformly in $i \in [n]$ and we recall that $u_i = mWi(H_n - H_i)/(c\log n)$. In fact, we bound (7.3) from above further by using that $u_i \leq m(H_n - H_i)/(c\log n) =: \widetilde{u}_i$ almost surely. Define $u : \mathbb{R} \to \mathbb{R}$ by $u(x) := m(1 -$

$\log x/\log n)/(c\mathbb{E}[W])$ and $\phi:\mathbb{R}\to\mathbb{R}$ by $\phi(x):=1-x+\log x$. Then, for $n^\varepsilon\le i<n^{1-\varepsilon}$ such that $i=n^{\beta+o(1)}$ for some $\beta\in[\varepsilon,1-\varepsilon]$ (where the $o(1)$ is independent of β) and $x\in[i,i+1)$,

$$\begin{aligned}|\phi(\widetilde{u}_i)-\phi(u(x))|&\le|\widetilde{u}_i-u(x)|+|\log(\widetilde{u}_i/u(x))|\\&=\Big|\frac{m}{c\mathbb{E}[W]}\Big(1-\frac{\log x}{\log n}\Big)-\frac{m}{c\log n}\sum_{j=i}^{n-1}\frac{1}{S_j}\Big|\\&\quad+\Big|\log\Big(\frac{\mathbb{E}[W]}{\log n-\log x}\sum_{j=i}^{n-1}\frac{1}{S_j}\Big)\Big|.\end{aligned}\tag{7.7}$$

By (5.2) and since i diverges with n, $\sum_{j=i}^{n-1}1/S_j-\log(n/i)/\mathbb{E}[W]=o(1)$ almost surely as $n\to\infty$. Applying this to the right-hand side of (7.7) yields

$$|\phi(\widetilde{u}_i)-\phi(u(x))|\le\frac{m}{c\mathbb{E}[W]}\Big|\frac{\log x-\log i}{\log n}\Big|+\Big|\log\Big(1+\frac{\log x-\log i+o(1)}{\log n-\log x}\Big)\Big|.$$

Since $x\ge i\ge n^\varepsilon$ and $|x-i|\le 1$, we thus obtain that, uniformly in $n^\varepsilon\le i<n^{1-\varepsilon}$ and $x\in[i,i+1)$, $|\phi(\widetilde{u}_i)-\phi(u(x))|=o(1/(n^\varepsilon\log n))$ almost surely as $n\to\infty$. Applying this to the left-hand side of (7.6) together with (7.3) (using $\widetilde{u}_i$ rather than u_i), we can bound the sum from above by

$$\begin{aligned}\sum_{n^\varepsilon\le i<n^{1-\varepsilon}}\mathbb{P}_W(\mathcal{Z}_n(i)\ge a_n)&\le\sum_{n^\varepsilon\le i<n^{1-\varepsilon}}\mathrm{e}^{a_n\phi(\widetilde{u}_i)}\\&\le\sum_{n^\varepsilon\le i<n^{1-\varepsilon}}\int_i^{i+1}\mathrm{e}^{a_n\phi(u(x))+a_n|\phi(\widetilde{u}_i)-\phi(u(x))|}\,\mathrm{d}x\\&\le(1+o(1))\int_{n^\varepsilon}^{n^{1-\varepsilon}+1}\mathrm{e}^{a_n\phi(u(x))}\,\mathrm{d}x.\end{aligned}\tag{7.8}$$

Recall that $\theta_m=1+\mathbb{E}[W]/m$ and set $\widetilde{\theta}_m:=1+m/\mathbb{E}[W]$. Using the variable transformation $w=\widetilde{\theta}_m(\log n-\log x)$ and Stirling's formula in the last line yields

$$\begin{aligned}&(1+o(1))\frac{n^{1+c-c\log\theta_m}}{\widetilde{\theta}_m(c\log n)^{c\log n}}\int_{\varepsilon\widetilde{\theta}_m\log n+o(1)}^{(1-\varepsilon)\widetilde{\theta}_m\log n}w^{c\log n}\mathrm{e}^{-w}\,\mathrm{d}w\\&\le o(1)+(1+o(1))\frac{n^{1+c-c\log\theta_m}}{\widetilde{\theta}_m(c\log n)^{c\log n}}\Gamma(1+c\log n)\\&\sim\frac{n^{1-c\log\theta_m}}{\widetilde{\theta}_m}\sqrt{2\pi c\log n},\end{aligned}$$

which tends to zero by the choice of c. Hence, combining the above with (7.4) and (7.5) yields (7.1) and hence (7.2).

Now, let $a_n:=\lceil c\log n\rceil, b_n:=\lceil\delta\log n\rceil$ with $c\in(0,1/\log\theta_m)$ and $\delta\in(0,1/\log\theta_m-c)$. For $i\in\mathbb{N}$ fixed, we couple $\mathcal{Z}_n(i)$ to a sequence of suitable random variables. Let $(P_j)_{j\ge2}$ be independent Poisson random variables with mean $mW_i/S_{j-1}, j\ge 2$. Then, we can couple $\mathcal{Z}_n(i)$ to the P_j's to obtain

$$\mathcal{Z}_n(i)\ge\sum_{j=i+1}^n P_j\mathbb{1}_{\{P_j\le 1\}}=\sum_{j=i+1}^n P_j-\sum_{j=i+1}^n P_j\mathbb{1}_{\{P_j>1\}}=:W_n(i)-Y_n(i).$$

By Lemma 7.1 and the inequality

$$\mathbb{P}_W(\mathcal{Z}_n(i)\ge a_n)\ge\mathbb{P}_W(W_n(i)\ge a_n+b_n)-\mathbb{P}_W(Y_n(i)\ge b_n),$$

it follows that we are required to prove that, for some $\varepsilon,\xi>0$ sufficiently small,

$$\sum_{\substack{n^\varepsilon\le i\le n^{1-\varepsilon}\\ W_i\ge\mathrm{e}^{-\xi}}}\mathbb{P}_W(W_n(i)\ge a_n+b_n)\xrightarrow{\mathbb{P}}\infty,\qquad\sum_{n^\varepsilon\le i\le n^{1-\varepsilon}}\mathbb{P}_W(Y_n(i)\ge b_n)\xrightarrow{\mathbb{P}}0,\tag{7.9}$$

as n tends to infinity, to obtain

$$\mathbb{P}_W\Big(\max_{i\in[n]}\mathcal{Z}_n(i)\geq a_n\Big)\xrightarrow{\mathbb{P}}1.$$

Using the uniform integrability of the conditional probability measure then yields

$$\lim_{n\to\infty}\mathbb{P}\Big(\max_{i\in[n]}\mathcal{Z}_n(i)\geq a_n\Big)=1,$$

which together with (7.2) proves convergence in probability of $\max_{i\in[n]}\mathcal{Z}_n(i)/\log n$ to $1/\log\theta_m$.

We first prove the first claim of (7.9). Note that $W_n(i)$ is a Poisson random variable with parameter $mW_i\sum_{j=i}^{n-1}1/S_j$. We note that by the strong law of large numbers, for some $\eta\in(0,\mathrm{e}^{1/(c+\delta)}-\theta_m)$,

$$mW_i\sum_{j=i}^{n-1}1/S_j\geq mW_i\sum_{j=i}^{n-1}1/(j(\mathbb{E}\left[W\right]+\eta))\geq(mW_i/(\mathbb{E}\left[W\right]+\eta))\log(n/i) \tag{7.10}$$

for all $n^\varepsilon\leq i\leq n$ almost surely when n is sufficiently large. We can thus, for n large, conclude that $W_n(i)$ stochastically dominates $X_n(i)$, where $X_n(i)$ is a Poisson random variable with a parameter equal to the right-hand side of (7.10). Then, also using that $W_i\leq 1$ almost surely, it follows that for $i\geq n^\varepsilon$,

$$\begin{aligned}\mathbb{P}_W(W_n(i)\geq a_n+b_n)&\geq\mathbb{P}_W(X_n(i)\geq a_n+b_n)\\&\geq\mathbb{P}_W(X_n(i)=a_n+b_n)\\&\geq\Big(\frac{i}{n}\Big)^{m/(\mathbb{E}[W]+\eta)}W_i^{a_n+b_n}\frac{((mW_i/(\mathbb{E}\left[W\right]+\eta))\log(n/i))^{a_n+b_n}}{(a_n+b_n)!}.\end{aligned}$$

We now sum over all $i\in[n]$ such that $n^\varepsilon\leq i\leq n^{1-\varepsilon},W_i\geq\mathrm{e}^{-\xi}$ for some sufficiently small $\varepsilon,\xi>0$. By the lower bound on the vertex-weight, we obtain the further lower bound

$$\begin{aligned}&\sum_{\substack{n^\varepsilon\leq i\leq n^{1-\varepsilon}\\W_i\geq\mathrm{e}^{-\xi}}}\mathbb{P}_W(W_n(i)\geq a_n+b_n)\\&\quad\geq\sum_{n^\varepsilon\leq i\leq n^{1-\varepsilon}}\mathbb{1}_{\{W_i\geq\mathrm{e}^{-\xi}\}}\mathrm{e}^{-\xi(a_n+b_n)}\Big(\frac{i}{n}\Big)^{m/(\mathbb{E}[W]+\eta)}\frac{((m/(\mathbb{E}\left[W\right]+\eta))\log(n/i))^{a_n+b_n}}{(a_n+b_n)!}\\&\quad=:T_n.\end{aligned} \tag{7.11}$$

We now claim that $T_n\xrightarrow{\mathbb{P}}\infty$. This follows from the fact that the mean of T_n diverges, and that T_n concentrates around the mean. We first show the former statement. Let $p=p(\xi)=\mathbb{P}\big(W\geq\mathrm{e}^{-\xi}\big)$. Note that, due to the fact that $x_0=\sup\{x\in\mathbb{R}:\mathbb{P}(W\leq x)<1\}=1$, $p>0$ for any $\xi>0$. Hence,

$$\mathbb{E}\left[T_n\right]=\frac{p}{\mathrm{e}^{\xi(a_n+b_n)}}\Big(\frac{m}{\mathbb{E}\left[W\right]+\eta}\Big)^{a_n+b_n}\frac{1}{(a_n+b_n)!}\sum_{n^\varepsilon\leq i\leq n^{1-\varepsilon}}\Big(\frac{i}{n}\Big)^{m/(\mathbb{E}[W]+\eta)}\log(n/i)^{a_n+b_n}. \tag{7.12}$$

Then, in a similar way as in (7.8), and applying a variable transformation $t=(1+m/(\mathbb{E}\left[W\right]+\eta))\log(n/x)$,

$$\begin{aligned}&\sum_{n^\varepsilon\leq i\leq n^{1-\varepsilon}}\Big(\frac{i}{n}\Big)^{m/(\mathbb{E}[W]+\eta)}(\log(n/i))^{a_n+b_n}\\&\quad=(1+o(1))\int_{n^\varepsilon}^{n^{1-\varepsilon}}\Big(\frac{x}{n}\Big)^{m/(\mathbb{E}[W]+\eta)}\log(n/x)^{a_n+b_n}\,\mathrm{d}x\\&\quad=(1+o(1))n(a_n+b_n)!\Big(1+\frac{m}{\mathbb{E}\left[W\right]+\eta}\Big)^{-(a_n+b_n+1)}\int_{\varepsilon(1+m/(\mathbb{E}[W]+\eta))\log n}^{(1-\varepsilon)(1+m/(\mathbb{E}[W]+\eta))}\frac{\mathrm{e}^{-t}t^{a_n+b_n}}{(a_n+b_n)!}\,\mathrm{d}t.\end{aligned}$$

We now identify the integral as the probability of the event $\{Y_n \in (\varepsilon(1+m/(\mathbb{E}[W]+\eta))\log n, (1-\varepsilon)(1+m/(\mathbb{E}[W]+\eta))\log n)\}$, where Y_n is a sum of a_n+b_n+1 rate one exponential random variables. Since $a_n+b_n=(1+o(1))(c+\delta)\log n$, it follows from the law of large numbers that this probability equals $1-o(1)$ when $c+\delta \in (\varepsilon(1+m/(\mathbb{E}[W]+\eta)), (1-\varepsilon)(1+m/(\mathbb{E}[W]+\eta)))$, which is the case for ε, η sufficiently small. Thus, combining the above with (7.12), we arrive at

$$\begin{aligned}\mathbb{E}[T_n] &\sim p\mathrm{e}^{-\xi(a_n+b_n)}\Big(\frac{m}{\mathbb{E}[W]+\eta}\Big)^{a_n+b_n}\Big(1+\frac{m}{\mathbb{E}[W]+\eta}\Big)^{-(a_n+b_n)+1}n\\ &= p\frac{\mathbb{E}[W]+\eta}{m\theta_m+\eta}\exp\{\log n(1-(1+o(1))(c+\delta)(\log(\theta_m+\eta/m)+\xi))\} \qquad (7.13)\\ &= p\frac{\mathbb{E}[W]+\eta}{m\theta_m+\eta}n^{1-(1+o(1))(c+\delta)(\log(\theta_m+\eta/m)+\xi)}.\end{aligned}$$

By the choice of c and δ, the exponent is positive when η and ξ are sufficiently small. What remains is to show that T_n concentrates around $\mathbb{E}[T_n]$. Using a Chebyshev bound yields for any $\zeta>0$ fixed,

$$\mathbb{P}(|T_n/\mathbb{E}[T_n]-1|\geq \zeta)\leq \frac{\mathrm{Var}(T_n)}{(\zeta\mathbb{E}[T_n])^2}, \qquad (7.14)$$

so that the result follows if $\mathrm{Var}(T_n)=o(\mathbb{E}[T_n]^2)$. Since T_n is a sum of weighted, independent Bernoulli random variables, we readily have

$$\mathrm{Var}(T_n)=\Big(\frac{m}{\mathbb{E}[W]+\eta}\Big)^{2(a_n+b_n)}\frac{p(1-p)\mathrm{e}^{-2\xi(a_n+b_n)}}{((a_n+b_n)!)^2}\sum_{n^\varepsilon\leq i\leq n^{1-\varepsilon}}\Big(\frac{i}{n}\Big)^{2m/(\mathbb{E}[W]+\eta)}\log(n/i)^{2(a_n+b_n)}.$$

Again writing the sum as an integral over x instead of i, and now using the variable transformation $t=(1+2m/(\mathbb{E}[W]+\eta))\log(n/x)$, we obtain that the sum equals

$$(1+o(1))n(2(a_n+b_n))!\Big(1+\frac{2m}{\mathbb{E}[W]+\eta}\Big)^{-2(a_n+b_n)-1}\int_{\varepsilon(1+2m/(\mathbb{E}[W]+\eta))\log n}^{(1-\varepsilon)(1+2m/(\mathbb{E}[W]+\eta))\log n}\frac{\mathrm{e}^{-t}t^{2(a_n+b_n)}}{(2(a_n+b_n))!}\,\mathrm{d}t.$$

We again interpret the integral as the probability of the event $\{\widetilde{Y}_n \in (\varepsilon(1+2m/(\mathbb{E}[W]+\eta))\log n, (1-\varepsilon)(1+2m/(\mathbb{E}[W]+\eta))\log n)\}$, where $\widetilde{Y}_n$ is a sum of $2(a_n+b_n)$ rate one exponential random variables. Again, for η and ε sufficiently small, this probability is $1-o(1)$ by the law of large numbers. Thus, we obtain,

$$\mathrm{Var}(T_n)=(1+o(1))\frac{p(1-p)(\mathbb{E}[W]+\eta)}{m(\theta_m+1)+\eta}\mathrm{e}^{-2\xi(a_n+b_n)}n\frac{(2(a_n+b_n))!}{((a_n+b_n)!)^2}(1+\theta_m+\eta/m)^{-2(a_n+b_n)}.$$

Using Stirling's approximation for the factorial terms then yields

$$\begin{aligned}\mathrm{Var}(T_n) &\sim \frac{p(1-p)(\mathbb{E}[W]+\eta)}{(m(\theta_m+1)+\eta)\sqrt{\pi(a_n+b_n)}}\mathrm{e}^{\log n+2(a_n+b_n)(\log 2-\log(1+\theta_m+\eta/m)-\xi)}\\ &= \frac{p(1-p)(\mathbb{E}[W]+\eta)}{(m(\theta_m+1)+\eta)\sqrt{\pi(a_n+b_n)}}n^{1+2(c+\delta)(1+o(1))(\log 2-\log(1+\theta_m+\eta/m)-\xi)}.\end{aligned}$$

Combining this with (7.13), we find that

$$\frac{\mathrm{Var}(T_n)}{\mathbb{E}[T_n]^2}\leq \frac{K}{\sqrt{a_n+b_n}}\exp\Big\{\log n\Big(2(c+\delta)(1+o(1))\log\Big(\frac{2(\theta_m+\eta/m)}{1+\theta_m+\eta/m}\Big)-1\Big)\Big\}, \qquad (7.15)$$

where $K>0$ is a suitable constant. The exponential terms decays with n when

$$c+\delta<\Big(\log\Big(4\Big(\frac{\theta_m+\eta/m}{1+\theta_m+\eta/m}\Big)^2\Big)\Big)^{-1},$$

which is satisfied for any $\theta_m\in(1,2]$ when η is sufficiently small, since $c+\delta<1/\log\theta_m$ and

$$\log\theta_m>\log\Big(4\Big(\frac{\theta_m+\eta/m}{1+\theta_m+\eta/m}\Big)^2\Big)$$

holds for any $\theta_m \in (1,2]$ when η is sufficiently small. Therefore, $T_n/\mathbb{E}[T_n] \xrightarrow{\mathbb{P}} 1$, so that $T_n \xrightarrow{\mathbb{P}} \infty$. This then implies the first statement in (7.9).

We now prove the second statement of (7.9). For a Poisson random variable P with mean λ, we find that

$$\mathbb{E}\left[P\mathbb{1}_{\{P>1\}}\right] = \mathbb{E}[P] - \mathbb{P}(P=1) = \lambda\big(1-\mathrm{e}^{-\lambda}\big) \le \lambda^2, \tag{7.16}$$

and, for any $t \in \mathbb{R}$, it follows from [8, Page 8] that

$$\mathbb{E}\left[\mathrm{e}^{t(P\mathbb{1}_{\{P>1\}}-\mathbb{E}[P\mathbb{1}_{\{P>1\}}])}\right] \le \mathrm{e}^{\lambda^2\mathrm{e}^{2t}}. \tag{7.17}$$

Now, since $Y_n(i) = \sum_{j=i+1}^n P_j\mathbb{1}_{\{P_j>1\}}$, where the P_j's are independent Poisson random variables with mean mW_i/S_{j-1}, using an upper bound inspired by (7.10) and using (7.16), we obtain that almost surely for all n large and $i \ge n^\varepsilon$,

$$\mathbb{E}_W[Y_n(i)] \le \frac{m^2}{(\mathbb{E}[W]-\eta)^2}\sum_{j=i}^{n-1} 1/j^2 \le \frac{m^2}{(\mathbb{E}[W]-\eta)^2(i-1)}.$$

Then, for $i \ge n^\varepsilon$ and n large enough so that $b_n(i-1) \ge 4(m/(\mathbb{E}[W]-\eta))^2$, we write for any $t>0$ using (7.17),

$$\begin{aligned}\mathbb{P}_W(Y_n(i) \ge b_n) &\le \mathbb{P}_W(Y_n(i) - \mathbb{E}_W[Y_n(i)] \ge b_n/2)\\ &\le \mathrm{e}^{-tb_n/2}\mathbb{E}_W[\mathrm{e}^{t(Y_n(i)-\mathbb{E}_W[Y_n(i)])}]\\ &\le \exp\{-tb_n/2 + \mathrm{e}^{2t}m^2/((\mathbb{E}[W]-\eta)^2(i-1))\}.\end{aligned}$$

This upper bound is smallest for $t = \log(b_n(i-1)(\mathbb{E}[W]-\eta)^2/(4m^2))/2$, which yields

$$\begin{aligned}\mathbb{P}_W(Y_n(i) \ge b_n) &\le \exp\Big\{\frac{b_n}{4}(1-\log(b_n(i-1)(\mathbb{E}[W]-\eta)^2/(4m^2)))\Big\}\\ &= \Big(\frac{4\mathrm{e}m^2}{b_n(\mathbb{E}[W]-\eta)^2(i-1)}\Big)^{b_n/4} \le n^{-\varepsilon b_n/4},\end{aligned}$$

when n is large enough such that $b_n \ge 8\mathrm{e}m^2/(\mathbb{E}[W]-\eta)^2$ and $n^\varepsilon \le 2(n^\varepsilon - 1)$. It then follows that

$$\sum_{n^\varepsilon \le i \le n^{1-\varepsilon}} \mathbb{P}_W(Y_n(i) \ge b_n) \le n^{1-\varepsilon\delta\log n/4}, \tag{7.18}$$

which tends to zero with n almost surely. This finishes the proof of

$$\max_{i\in[n]} \mathcal{Z}_n(i)/\log n \xrightarrow{\mathbb{P}} 1/\log\theta_m.$$

We now turn to the almost sure convergence. Let $Z_n := \max_{i\in[n]}\mathcal{Z}_n(i)$. Similar to [8], we use the bounds

$$\inf_{N\le n}\frac{Z_{2^n}}{(n+1)\log 2} \le \inf_{2^N\le n}\frac{Z_n}{\log n} \le \sup_{2^N\le n}\frac{Z_n}{\log n} \le \sup_{N\le n}\frac{Z_{2^{n+1}}}{n\log 2}.$$

It thus follows that to prove the almost sure convergence of the rescaled maximum degree $Z_n/\log n$, it suffices to do so for the subsequence $Z_{2^n}/(n\log 2)$, for which we can obtain stronger bounds due to the fact that 2^n grows exponentially. To prove the almost sure convergence of $Z_{2^n}/(n\log 2)$ to $1/\log\theta_m$, it thus suffices to prove

$$\liminf_{n\to\infty}\frac{Z_{2^n}}{(n+1)\log 2} \ge \frac{1}{\log\theta_m}, \qquad \limsup_{n\to\infty}\frac{Z_{2^{n+1}}}{n\log 2} \le \frac{1}{\log\theta_m}, \tag{7.19}$$

almost surely, which can be achieved with the bounds used to prove the convergence in probability. Namely, for the upper bound, using (7.4), (7.5) and (7.8), we obtain for any $c > 1/\log\theta_m$, a sufficiently small $\xi > 0$ and some large constant $C > 0$,

$$\sum_{i=1}^{2^{n+1}} \mathbb{P}_W(\mathcal{Z}_{2^{n+1}}(i)/(n\log 2) \ge c) \le 2\mathrm{e}^{-\xi n(1+o(1))} + (1+o(1))C\sqrt{n}\mathrm{e}^{-\xi n},$$

which is summable, so it follows from the Borel-Cantelli lemma that the upper bound in (7.19) holds $\mathbb{P}_W$-almost surely. A similar approach as in (6.10) then yields the $\mathbb{P}$-almost sure convergence.

Similarly, for the lower bound in (7.19), we have for $c < 1/\log\theta_m$ by Lemma 7.1,

$$\begin{aligned}\mathbb{P}_W(Z_{2^n}/((n+1)\log 2) < c) &\leq 1 - \frac{\sum_{i=1}^{2^n}\mathbb{P}_W(\mathcal{Z}_{2^n}(i)\geq c(n+1)\log 2)}{1+\sum_{i=1}^{2^n}\mathbb{P}_W(\mathcal{Z}_{2^n}(i)\geq c(n+1)\log 2)}\\ &= \frac{1}{1+\sum_{i=1}^{2^n}\mathbb{P}_W(\mathcal{Z}_{2^n}(i)\geq c(n+1)\log 2)}.\end{aligned} \tag{7.20}$$

Similar to (7.9), we again bound the sum from below by

$$\begin{aligned}\sum_{i=1}^{2^n}\mathbb{P}_W(\mathcal{Z}_{2^n}(i)\geq c(n+1)\log 2) \geq &\sum_{\substack{2^{\varepsilon n}\leq i\leq 2^{(1-\varepsilon)n}\\ W_i\geq \mathrm{e}^{-\xi}}}\mathbb{P}_W(W_{2^n}(i)\geq (c+\delta)(n+1)\log 2)\\ &-\sum_{2^{\varepsilon n}\leq i\leq 2^{(1-\varepsilon)n}}\mathbb{P}_W(Y_{2^n}(i)\geq \delta(n+1)\log 2)\,.\end{aligned}$$

First, by the bound in (7.18), we find that

$$1-\sum_{2^{\varepsilon n}\leq i\leq 2^{(1-\varepsilon)n}}\mathbb{P}_W(Y_{2^n}(i)\geq \delta(n+1)\log 2)\geq 0,$$

almost surely for all n large. Then, we bound the sum of tail probabilities of the $W_{2^n}(i)$ from below by T_{2^n}, where we recall the definition of T_n from (7.11). Combining (7.14) and (7.15), we find that $T_{2^n}\geq \mathbb{E}[T_{2^n}]/2$ almost surely for all n large. Together with (7.13) and (7.20), we thus obtain

$$\mathbb{P}_W(Z_{2^n}/((n+1)\log 2) < c)\leq C\exp\{-n\log 2(1-(1+o(1))(c+\delta)(\log(\theta_m+\eta/m)+\xi))\},$$

for some constant $C>0$, which is summable when η and ξ are sufficiently small, so that the lower bound holds for all but finitely many n $\mathbb{P}_W$-almost surely by the Borel-Cantelli lemma. Again, a similar argument as in (6.10) allows us to extend this to $\mathbb{P}$-almost surely, which concludes the proof. □

7.2. **Proof of Theorem 2.9, (Gumbel) case.**

In this subsection we prove the **(Gumbel)** case of Theorem 2.9, which combines the results obtained in Sections 5 and 6.

Proof of Theorem 2.9, **(Gumbel)** *case.* We only discuss the case $m=1$, as the proof for $m>1$ follows analogously. Most of the proof directly follows by combining Propositions 5.2, 5.4 and 5.5 with Proposition 6.1, (6.1), (6.2) and (6.3). The one statement that remains to be proved is that $\log(I_n)/\log n$ converges, either in probability or almost surely, depending on the class of distributions the vertex-weight distribution is in, to some constant. Let us start with the **(Gumbel)**-**(RV)** sub-case, as in (2.16).

We recall the sequence $(\widetilde{\varepsilon}_k)_{k\in\mathbb{N}_0}$ in (5.13) and note that $\widetilde{\varepsilon}_k$ is decreasing and tends to 0 with k. Now, fix $\varepsilon>0$ and let k be large enough such that $\widetilde{\varepsilon}_k<\varepsilon$. We are required to show that

$$\sum_{n=1}^{\infty}\mathbb{1}_{\{\log I_n/\log n\leq \gamma-\varepsilon\}}<\infty,\ \text{a.s.}$$

First, we note that

$$\begin{aligned}\Big\{\frac{\log I_n}{\log n}\leq\gamma-\varepsilon\Big\}&\subseteq\Big\{\frac{\log I_n}{\log n}\leq\gamma-\widetilde{\varepsilon}_k\Big\}\\ &\subseteq\Big\{\max_{n^{\gamma-\widetilde{\varepsilon}_k}\leq i\leq n}\frac{\mathcal{Z}_n(i)}{(1-\gamma)b_{n^\gamma}\log n}\leq\max_{i<n^{\gamma-\widetilde{\varepsilon}_k}}\frac{\mathcal{Z}_n(i)}{(1-\gamma)b_{n^\gamma}\log n}\Big\}.\end{aligned}$$

Let us denote the event on the right-hand side by A_n and, for $\eta > 0$, define the event

$$C_n := \Big\{ \max_{i\in[n]} |\mathcal{Z}_n(i) - \mathbb{E}_W[\mathcal{Z}_n(i)]| \leq \frac{\eta}{2}(1-\gamma)b_{n^\gamma}\log n \Big\}.$$

We then have that $A_n \subseteq (A_n \cap C_n) \cup C_n^c$, so that

$$\sum_{n=1}^\infty \mathbb{1}_{\{\log I_n/\log n \leq \gamma-\varepsilon\}} \leq \sum_{n=1}^\infty \mathbb{1}_{A_n} \leq \sum_{n=1}^\infty \mathbb{1}_{A_n}\mathbb{1}_{C_n} + \mathbb{1}_{C_n^c}.$$

We now use the concentration argument used to prove (6.2) in Proposition 6.1. It already follows from (6.2) that only finitely many of the $\mathbb{1}_{C_n^c}$ equal 1, so what remains is to show that, almost surely, only finitely many of the product of indicators equal 1 as well. On C_n,

$$A_n \cap C_n \subseteq \Big\{ \max_{n^{\gamma-\tilde\varepsilon_k}\leq i\leq n} \frac{\mathbb{E}_W[\mathcal{Z}_n(i)]}{(1-\gamma)b_{n^\gamma}\log n} \leq \max_{i<n^{\gamma-\tilde\varepsilon_k}} \frac{\mathbb{E}_W[\mathcal{Z}_n(i)]}{(1-\gamma)b_{n^\gamma}\log n} + \eta \Big\}.$$

The limsup of the second maximum in the event on the right-hand side is almost surely at most $c_k < 1$, where c_k is the quantity defined in (5.15). Then, we can directly bound the first maximum from below by

$$\max_{n^{\gamma-\tilde\varepsilon_k}\leq i\leq n} \frac{\mathbb{E}_W[\mathcal{Z}_n(i)]}{(1-\gamma)b_{n^\gamma}\log n} \geq \max_{n^{\gamma-\tilde\varepsilon_k}\leq i\leq n^\gamma} \frac{\mathbb{E}_W[\mathcal{Z}_n(i)]}{(1-\gamma)b_{n^\gamma}\log n} \geq \max_{n^{\gamma-\varepsilon_k}\leq i\leq n^\gamma} \frac{W_i}{b_{n^\gamma}} \frac{\sum_{i=n^\gamma}^{n-1} 1/S_j}{(1-\gamma)\log n},$$

and the lower bound converges almost surely to 1 by (5.2) and Lemma 5.3. Hence, if we take some $\delta \in (0, 1-c_k)$ and set $\eta < \delta/3$, then almost surely there exists an $N \in \mathbb{N}$ such that for all $n \geq N$,

$$\max_{i<n^{\gamma-\tilde\varepsilon_k}} \frac{\mathbb{E}_W[\mathcal{Z}_n(i)]}{(1-\gamma)b_{n^\gamma}\log n} < c_k + \delta/3, \qquad \max_{n^{\gamma-\tilde\varepsilon_k}\leq i\leq n} \frac{\mathbb{E}_W[\mathcal{Z}_n(i)]}{(1-\gamma)b_{n^\gamma}\log n} > 1 - \delta/3.$$

It follows that that the event

$$\Big\{ \max_{n^{\gamma-\tilde\varepsilon_k}\leq i\leq n} \frac{\mathbb{E}_W[\mathcal{Z}_n(i)]}{(1-\gamma)b_{n^\gamma}\log n} \leq \max_{i<n^{\gamma-\tilde\varepsilon_k}} \frac{\mathbb{E}_W[\mathcal{Z}_n(i)]}{(1-\gamma)b_{n^\gamma}\log n} + \eta \Big\}$$

almost surely does not hold for all $n \geq N$. Thus, for any $\varepsilon > 0$ and for all n large, $\log I_n/\log n \geq \gamma-\varepsilon$ almost surely. With a similar approach, we can prove that $\log I_n/\log n \leq \gamma + \varepsilon$, so that the almost sure convergence to γ is established.

For the **(Gumbel)**-**(SV)** and **(Gumbel)**-**(RaV)** sub-cases, we intend to prove the convergence of $\log I_n/\log n$ to 0 and 1 in probability, respectively. We provide a proof for the former sub-case, and note that the proof for the latter sub-case follows in a similar way.

Let $\varepsilon > 0$. Then,

$$\mathbb{P}(\log I_n/\log n \geq \varepsilon) = \mathbb{P}(I_n \geq n^\varepsilon) = \mathbb{P}\Big(\max_{n^\varepsilon\leq i\leq n} \frac{\mathcal{Z}_n(i)}{b_n\log n} > \max_{1\leq i<n^\varepsilon} \frac{\mathcal{Z}_n(i)}{b_n\log n} \Big). \tag{7.21}$$

Again, we define the event, for $\eta \in (\varepsilon/3)$ small,

$$C_n := \Big\{ \max_{i\in[n]} |\mathcal{Z}_n(i) - \mathbb{E}_W[\mathcal{Z}_n(i)]| \leq \eta b_n \log n/2 \Big\},$$

which holds with high probability due to (6.1). We can then further bound the right-hand side in (7.21) from above by

$$\begin{aligned} &\mathbb{P}\Big(\Big\{ \max_{n^\varepsilon\leq i\leq n} \frac{\mathcal{Z}_n(i)}{b_n\log n} > \max_{1\leq i<n^\varepsilon} \frac{\mathcal{Z}_n(i)}{b_n\log n} \Big\} \cap C_n \Big) + \mathbb{P}(C_n^c) \\ &\leq \mathbb{P}\Big(\max_{n^\varepsilon\leq i\leq n} \frac{\mathbb{E}_W[\mathcal{Z}_n(i)]}{b_n\log n} > \max_{1\leq i<n^\varepsilon} \frac{\mathbb{E}_W[\mathcal{Z}_n(i)]}{b_n\log n} - \eta \Big) + \mathbb{P}(C_n^c). \end{aligned} \tag{7.22}$$

In a similar way as in the proof of Proposition 5.2, as well as in the proof above for the **(Gumbel)**-**(RV)** sub-case, we can bound the maximum of the conditional mean in-degrees on the left-hand side from above and the maximum on the right-hand side from below by random quantities that converge in probability to fixed constants. Namely, for $\varepsilon > 0$ and $\beta \in (0, \varepsilon/3)$,

$$\max_{n^\varepsilon \le i \le n} \frac{\mathbb{E}_W[\mathcal{Z}_n(i)]}{b_n \log n} \le \max_{n^\varepsilon \le i \le n} \frac{W_i \sum_{j=n^\varepsilon}^{n-1} 1/S_j}{b_n \log n} \xrightarrow{\mathbb{P}} 1-\varepsilon,$$

$$\max_{1 \le i < n^\varepsilon} \frac{\mathbb{E}_W[\mathcal{Z}_n(i)]}{b_n \log n} - \eta \ge \max_{i \in [n^\beta]} \frac{W_i \sum_{j=n^\beta}^{n-1} 1/S_j}{b_n \log n} - \eta \xrightarrow{\mathbb{P}} 1-\beta-\eta,$$

so that, with high probability, the first quantity is at most $1-5\varepsilon/6$ and the second quantity is at least $1-(\beta+\eta)-\varepsilon/6 > 1-5\varepsilon/6$ by the choice of β and η. It thus follows that the first probability on the right-hand side of (7.22) tends to zero with n, and so does the second probability (C_n holds with high probability), so that the claim follows. $\square$

7.3. **Proof of Theorem 2.9, (Fréchet) case.** In this subsection we prove the **(Fréchet)** case of Theorem 2.9, which combines the results obtained in Sections 5 and 6.

Proof of Theorem 2.9, **(Fréchet)** *case.* The proof of the convergence of $\max_{i\in[n]} \mathcal{Z}_n(i)/u_n$ and $\max_{i\in[n]} \mathcal{Z}_n(i)/n$ as in (2.18) and (2.19), respectively, follows directly from Proposition 5.6 combined with (6.4) and (6.5) in Proposition 6.1. Then, the distributional convergence of I_n/n to I_α and I as in (2.18) and (2.19), respectively, follows from the same argument as in the proof of [19, Theorem 2.7]. In particular, we can conclude from that proof that, when $\alpha > 2$,

$$\mathbb{P}(I_\alpha \le t) = \frac{g(0,t)}{g(0,1)}, \quad \text{where } g(a,b) := \int_a^b \log(1/x)^{\alpha-1}\,\mathrm{d}x.$$

Finally, we note that when using the variable transformation $w = \log(1/x)$,

$$g(a,b) = \int_{\log(1/b)}^{\log(1/a)} w^{\alpha-1}\mathrm{e}^{-w}\,\mathrm{d}w = \Gamma(\alpha)\mathbb{P}\big(\mathrm{e}^{-W_\alpha} \in (a,b)\big),$$

where W_α is a $\Gamma(\alpha,1)$ random variable. Thus,

$$\mathbb{P}(I_\alpha \le t) = \frac{g(0,t)}{g(0,1)} = \mathbb{P}\big(\mathrm{e}^{-W_\alpha} \le t\big),$$

and $m\max_{(t,f)\in\Pi} f\log(1/t)$ follows a Fréchet distribution with shape parameter $\alpha-1$ and scale parameter $mg(0,1)^{1/(\alpha-1)} = m\Gamma(\alpha)^{1/(\alpha-1)}$, which concludes the proof. $\square$

7.4. **Second order behaviour of the maximum degree in the (Gumbel)- (RV) and (Gumbel)-(RaV) sub-cases.** In this subsection, we prove Theorems 2.12 and 2.14:

Proof of Theorem 2.12. We only discuss the case $m = 1$, as the proof for $m > 1$ follows analogously. Let us first deal with the results for the **(Gumbel)-(RV)** sub-case. The distributional convergence of the rescaled maximum degree to the correct limit, as in (2.21), and the convergence result in (2.23), directly follow by combining Proposition 5.4 with Proposition 6.1, (6.6) and (6.7). The one thing that remains to be proved is: $I_n(\gamma,s,t,\ell)/(\ell(n)n^\gamma)$ converges in distribution, jointly with the maximum degree of vertices $i \in C_n(\gamma,s,t,\ell)$, as in (2.21).

The distributional convergence of $I_n(\gamma,s,t,\ell)/(\ell(n)n^\gamma)$ follows from the same argument as in the proof of [19, Theorem 2.7], where now

$$g(a,b) := \log(b/a). \tag{7.23}$$

The joint convergence of I_n and $\max_{i\in C_n} \mathcal{Z}_n(i)$ when properly rescaled follows from [19, Theorem 2.7] as well.

In a similar way as in the proof of [19, Theorem 2.7], with g as in (7.23), for $x \in (s,t)$,

$$\mathbb{P}(I_\gamma \in (s,x)) = \frac{g(s,x)}{g(s,t)} = \frac{\log(x/s)}{\log(t/s)} = \mathbb{P}\big(\mathrm{e}^U \in (s,x)\big),$$

where $U \sim \text{Unif}(\log s, \log t)$. Finally, for any $x \in \mathbb{R}$,

$$\mathbb{P}\Big(\max_{\substack{(v,w)\in\Pi \\ v\in(s,t)}} w - \log v \leq x \Big) = \exp\Big\{ - \int_s^t \int_{x+\log v}^\infty \mathrm{e}^{-w}\,\mathrm{d}w\mathrm{d}v \Big\} = \exp\Big\{ - \mathrm{e}^{-(x-\log\log(t/s))} \Big\},$$

which proves that the distributional limits as described in (2.21) and (2.22) have the desired distributions. □

References

[1] L. Addario-Berry and L. Eslava. High degrees in random recursive trees. *Random Structures Algorithms*, 52(4):560–575, 2018.

[2] K. B. Athreya and S. Karlin. Limit theorems for the split times of branching processes. *J. Math. Mech.*, 17:257–277, 1967.

[3] S. Banerjee and S. Bhamidi. Persistence of hubs in growing random networks. *Preprint, arXiv:2004.13785*, 2020.

[4] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular variation*, volume 27 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1989.

[5] K. A. Borovkov and V. Vatutin. Trees with product-form random weights. *Discrete Mathematics & Theoretical Computer Science*, 2006.

[6] K. A. Borovkov and V. A. Vatutin. On the asymptotic behaviour of random recursive trees in random environments. *Adv. in Appl. Probab.*, 38(4):1047–1070, 2006.

[7] S. Dereich and M. Ortgiese. Robust analysis of preferential attachment models with fitness. *Combin. Probab. Comput.*, 23(3):386–411, 2014.

[8] L. Devroye and J. Lu. The strong convergence of maximal degrees in uniform random recursive trees and dags. *Random Structures Algorithms*, 7(1):1–14, 1995.

[9] M. Drmota. *Random trees*. SpringerWienNewYork, Vienna, 2009. An interplay between combinatorics and probability.

[10] L. Eslava. Depth of vertices with high degree in random recursive trees. *ArXiv*, abs/1611.07466, 2016.

[11] N. Fountoulakis and T. Iyer. Condensation phenomena in preferential attachment trees with neighbourhood influence. 2021.

[12] J. L. Gastwirth and P. K. Bhattacharya. Two probability models of pyramid or chain letter schemes demonstrating that their promotional claims are unreliable. *Oper. Res.*, 32(3):527–536, 1984.

[13] W. Goh and E. Schmutz. Limit distribution for the maximum degree of a random recursive tree. *J. Comput. Appl. Math.*, 142(1):61–82, 2002. Probabilistic methods in combinatorics and combinatorial optimization.

[14] A. Gut. *Probability: a graduate course*. Springer Texts in Statistics. Springer, New York, second edition, 2013.

[15] E. Hiesmayr and U. Işlak. Asymptotic results on Hoppe trees and their variations. *J. Appl. Probab.*, 57(2):441–457, 2020.

[16] R. van der Hofstad. *Random graphs and complex networks. Vol. 1*. Cambridge Series in Statistical and Probabilistic Mathematics, [43]. Cambridge University Press, Cambridge, 2017.

[17] R. van der Hofstad, P. Mörters, and N. Sidorova. Weak and almost sure limits for the parabolic Anderson model with heavy tailed potentials. *Ann. Appl. Probab.*, 18(6):2450–2494, 2008.

[18] T. Iyer. Degree distributions in recursive trees with fitnesses. *Preprint, arXiv:2005.02197*, 2020.

[19] B. Lodewijks and M. Ortgiese. A phase transition for preferential attachment models with additive fitness. *Electron. J. Probab.*, 25:Paper No. 146, 54, 2020.

[20] H. M. Mahmoud. *Evolution of random search trees*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons, Inc., New York, 1992. A Wiley-Interscience Publication.

[21] C. Mailler and G. Uribe Bravo. Random walks with preferential relocations and fading memory: a study through random recursive trees. *J. Stat. Mech. Theory Exp.*, (9):093206, 49, 2019.

[22] A. Meir and J. W. Moon. Recursive trees with no nodes of out-degree one. *Congr. Numer.*, 66:49–62, 1988. Nineteenth Southeastern Conference on Combinatorics, Graph Theory, and Computing (Baton Rouge, LA, 1988).

[23] H. S. Na and A. Rapoport. Distribution of nodes of a tree by degree. *Math. Biosci.*, 6:313–329, 1970.

[24] M. Pain and D. Sénizergues. Correction terms for the height of weighted recursive trees. *arXiv preprint arXiv:2101.01156*, 2021.

[25] S. I. Resnick. *Extreme values, regular variation and point processes*. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2008. Reprint of the 1987 original.

[26] D. Sénizergues. Geometry of weighted recursive and affine preferential attachment trees. *Preprint, arXiv:1904.07115*, 2019.

[27] R. T. Smythe and H. M. Mahmoud. A survey of recursive trees. *Teor. Ĭmovīr. Mat. Stat.*, (51):1–29, 1994.

[28] J. Szymański. On the maximum degree and the height of a random recursive tree. In *Random graphs '87 (Poznań, 1987)*, pages 313–324. Wiley, Chichester, 1990.
[29] R. Takahashi. Normalizing constants of a distribution which belongs to the domain of attraction of the Gumbel distribution. *Statist. Probab. Lett.*, 5(3):197–200, 1987.

Department of Mathematical Sciences, University of Bath, Claverton Down, Bath, BA2 7AY, United Kingdom.

Email address: b.lodewijks@bath.ac.uk, m.ortgiese@bath.ac.uk

3.2 Conclusion

We studied the weighted recursive graph in this chapter, in particular its degree distribution and the maximum degree. For the degree distribution, we have proved in Theorem 2.5 that a limiting degree distribution exists when the vertex-weight distribution has a finite mean, and in the case of an infinite mean power-law vertex-weight distribution we show a typical vertex is with high probability a leaf. The latter result is very comparable to the result in the extreme disorder regime provided in the previous chapter for preferential attachment models with fitness. Moreover, we provide an asymptotic expression for the limiting degree distribution in Theorem 2.7 for many classes of vertex-weight distributions, which are closely linked to maximum domains of attractions. This will prove useful in the upcoming chapter as well.

For the maximum degree, we establish first-order behaviour for a wide range of vertex-weight distributions in Theorem 2.8 and also establish the asymptotic size of the vertex label which attains the maximum degree in most cases. As discussed at the start of this chapter, the omission of the preferential mechanism in the weighted recursive graph allows for much richer behaviour compared to the preferential attachment models studied in the previous chapter, as the more subtle effects of the vertex-weights are able to appear. The main point of interest in Theorem 2.8 is the difference in the behaviour of the maximum degree when the vertex-weights are almost surely bounded or almost surely unbounded. In the latter case, the vertex label (its age) and the size of vertex-weights determine the growth of degrees and which vertices have large degrees. In the former case, the size of the vertex-weight is less significant, at least to first order, and we observe behaviour that is much closer to what can be observed for the random recursive tree and directed acyclic graph, see [44]. This behaviour is induced by competition between vertices to acquire edges and vertices getting sufficiently lucky by obtaining high degrees.

An interesting observation in Theorem 2.8 is the behaviour of the maximum degree in the case the vertex-weights follow a power-law distribution with an infinite mean, as in Equation (2.20). Here, we see that the joint limiting distribution of the rescaled maximum degree and its rescaled vertex label are the same as in the case of the preferential attachment models discussed in the previous chapter (Theorem 2.7, Equation (2.12)). Intuitively, this is due to the fact that the effect of the fitness is able to overpower the preferential attachment mechanism completely, in the sense that the influence of the in-degree of vertices on evolution of the maximum degree is negligible compared to the effect of the fitness, and can be observed in the degree distribution as well, as discussed above. This can be interpreted as the opposite of what happens in the weak disorder regime for these preferential attachment models, where the influence of the fitness is almost negligible.

Finally, Theorems 2.11 and 2.13 provide insight in the second-order behaviour of the maximum degree for particular vertex-weight distributions in the Gumbel maximum domain of attraction. A somewhat surprising but interesting result here is that the different rescaling and vertex ranges used lead to random and deterministic limits Theorem 2.11 and 2.13, respectively.

Chapter 4

Fine asymptotics for the maximum degree in weighted recursive trees with bounded random weights

In this chapter we develop new methods to obtain more refined asymptotics for the maximum degree in weighted recursive trees. This extends the result presented in Theorem 2.8, the Bounded case, in Chapter 3 and also extends known results for random recursive trees. The following draft is joint work with Laura Eslava and Marcel Ortgiese.

4.1 Outline of the article

Properties of the weighted recursive graph model are studied in the previous chapter for a wide range of vertex-weight distributions. In this chapter we extend these results for the weighted recursive tree model by investigating the finer asymptotics of the maximum degree when the vertex-weights are almost surely bounded. The weighted recursive tree model is a specific case of the weighted recursive graph model studied in the previous chapter, where every vertex connects to exactly one predecessor, yielding a tree. Here, we develop new methods to analyse the degree distribution of the weighted recursive tree model that allow us to obtain more detailed results when the vertex-weights are almost surely bounded.

As is proved in the previous chapter, the maximum degree in the weighted recursive tree is asymptotically $\log_\theta n$, where $\theta := 1 + \mathbb{E}[W]$ with $\mathbb{E}[W]$ the mean of the vertex-weight distribution, and this result can be obtained by assuming the vertex-weights are almost surely bounded only. In this chapter we discuss the finer asymptotics of the maximum degree for which the underlying vertex-weight distribution plays a more important role.

The methods used in the previous chapter to prove the first-order asymptotics of the maximum degree make use of the conditional negative correlation between degrees. These methods are not sufficient to obtain more detailed results, however. Instead, asymptotic independence of the degrees of vertices selected uniformly at random is proved in this chapter, which follows from a more precise understanding of the convergence of the degree distribution. Moreover, we prove that the fraction of vertices with degree k in the weighted recursive tree of size n is asymptotically equal to the limit p_k (which was identified in the previous chapter) for fixed k as well as for k that diverge with n and provide an error rate for the asymptotic expression uniformly in $k = k(n)$, under certain conditions on how quickly $k(n)$ diverges.

Similarly to the previous chapter, we are then able to identify several classes of vertex-weight distributions for which different behaviour of the finer asymptotics of the maximum degree can be proved. This is carried out by combining the asymptotic independence with precise asymptotic expression of the limiting degree distribution, which are outlined and proved in the previous chapter, and with techniques developed for random recursive trees in [3].

The main results of the paper are formulated in Section 2. Section 4 presents two examples for which more detailed results are obtained. The main technical results are developed in Section 5, which are then used in Sections 6 and 7 to prove the main results and the results presented in Section 4.

Appendix 6B: Statement of Authorship

<table>
<tr><td colspan="10">This declaration concerns the article entitled:</td></tr>
<tr><td colspan="10">Fine asymptotics for the maximum degree in weighted recursive trees with bounded random weights</td></tr>
<tr><td colspan="10">Publication status (tick one)</td></tr>
<tr><td>Draft manuscript</td><td>x</td><td>Submitted</td><td></td><td>In review</td><td></td><td>Accepted</td><td></td><td>Published</td><td></td></tr>
<tr><td>Publication details (reference)</td><td colspan="9">Authors: Laura Eslava, Bas Lodewijks, Marcel Ortgiese</td></tr>
<tr><td colspan="10">Copyright status (tick the appropriate statement)</td></tr>
<tr><td colspan="3">I hold the copyright for this material</td><td>x</td><td colspan="5">Copyright is retained by the publisher, but I have been given permission to replicate the material here</td><td></td></tr>
<tr><td>Candidate's contribution to the paper (provide details, and also indicate as a percentage)</td><td colspan="9">The candidate predominantly executed the…

Formulation of ideas:

80%. The candidate drove the formulation of ideas to go into this paper for a large part.

Design of methodology:

80%. The candidate has largely contributed to the development of the theoretical methodology in this paper, working out most proofs in detail to a large extent.

Experimental work:

N/A

Presentation of data in journal format:

80%. The candidate has written the initial draft, after which adjustments and changes were proposed by the other authors.</td></tr>
<tr><td>Statement from Candidate</td><td colspan="9">This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.</td></tr>
<tr><td>Signed</td><td colspan="5">Bas Lodewijks</td><td colspan="2">Date</td><td colspan="2">01/07/2021</td></tr>
</table>

FINE ASYMPTOTICS FOR THE MAXIMUM DEGREE IN WEIGHTED RECURSIVE TREES WITH BOUNDED RANDOM WEIGHTS

LAURA ESLAVA[†], BAS LODEWIJKS[⋆], AND MARCEL ORTGIESE[⋆]

Abstract. A weighted recursive tree is an evolving tree in which vertices are assigned random vertex-weights and new vertices connect to a predecessor with a probability proportional to its weight. Here, we study the maximum degree and near-maximum degrees in weighted recursive trees when the vertex-weights are almost surely bounded. We are able to specify higher-order corrections to the first order growth of the maximum degree established in prior work. The accuracy of the results depends on the behaviour of the weight distribution near the largest possible value and in some cases we manage to find the corrections up to random order. Additionally, we describe the tail distribution of the maximum degree and establish asymptotic normality of the number of vertices with near-maximum degree. Our analysis extends the results proved for random recursive trees (where the weights are constant) to the case of random weights. The main technical result shows that the degrees of several uniformly chosen vertices are asymptotically independent with explicit error corrections.

1. Introduction

The Weighted Recursive Tree model (WRT), first introduced by Borovkov and Vatutin [4], is a recursive tree process $(T_n, n \in \mathbb{N})$ and a generalisation of the random recursive tree model. Here we consider a variation, first studied by Hiesmayr and Işlak [10], where the first vertex does not necessarily have weight one. Let $(W_i)_{i\in\mathbb{N}}$ be a sequence of positive vertex-weights. Initialise the process with the tree T_1, which consists of the vertex 1 (which denotes the root), and assign vertex-weight W_1 to it. Recursively, at every step $n \geq 2$, we obtain T_n by adding to T_{n-1} the vertex n, assigning vertex-weight W_n to it and connecting n to a vertex $i \in [n-1]$, which, conditionally on the vertex-weights $W_1, \ldots, W_{n-1}$, is selected with a probability proportional to W_i. In this paper, we consider edges to be directed towards the vertex with the smaller label. We note that allowing every vertex to connect to $m \in \mathbb{N}$ many predecessors, each one selected independently, yields the more general Weighted Recursive Graph model (WRG) introduced in [13]. The focus of this paper is the WRT model in the case which the vertex-weights are *bounded random variables.*

Lodewijks and Ortgiese [13] established that, in the case of positive, bounded random vertex-weights, the maximum degree Δ_n of the WRG model grows logarithmically and that $\Delta n/\log n \xrightarrow{a.s.} 1/\log\theta_m$, where $\theta_m := 1+\mathbb{E}[W]/m$ with $\mathbb{E}[W]$ the mean of the vertex-weight distribution and $m \in \mathbb{N}$ the out-degree of each vertex. Note that setting $m = 1$ yields the result for the WRT model. In this paper, we improve on this result by describing higher-order asymptotic behaviour of the maximum degree when the vertex-weights are almost surely bounded. In this case we are able to distinguish several classes of vertex-weight distributions for which different higher-order behaviour can be observed.

Beyond the initial work of Borovkov and Vatutin and also Hiesmayr and Işlak studying the height, depth and size of branches of the WRT model, other properties such as the degree distribution, large and maximum degrees, and weighted profile and height of the tree have been studied. Mailler and Uribe Bravo [15], as well as Sénizergues [17] and Sénizergues and Pain [16] study the weighted profile and height of the WRT model. Mailler and Uribe

Date: August 27, 2021.

Key words and phrases. Weighted recursive graph, Random recursive graph, Uniform DAG, Maximum degree, Degree distribution, Random environment.

Bravo consider random vertex-weights with particular distributions, whereas Sénizergues and Pain allow for a more general model with both sequences of deterministic as well as random weights.

Iyer [11] and the more general work by Fountoulakis and Iyer [7] study the degree distribution of a large class of evolving weighted random trees, and Lodewijks and Ortgiese [13] study the degree distribution of the WRG model. In both cases, the WRT model is a particular example of the models studied and all results prove the existence of an almost sure limiting degree distribution for the empirical degree distribution.

Finally, Lodewijks and Ortgiese [13] also study the maximum degree and the labels of the maximum degree vertices of the WRG model for a large range of vertex-weight distributions. In particular, a distinction between distributions with unbounded support and bounded support is observed. In the former case the behaviour and size of the label of the maximum degree is mainly controlled by a balance of vertices being old (i.e. having a small label) and having a large vertex-weight. In the latter case, due to the fact that the vertex-weights are bounded, the behaviour is instead controlled by a balance of vertices being old and having a degree which significantly exceeds their mean degree.

A particular case of the WRT model is the Random Recursive Tree (RRT) model, which is obtained when each vertex-weight equals one almost surely. As a result, techniques used to study the maximum degree in the RRT model can be adapted to analyse the maximum degree in the WRT model. Lodewijks and Ortgiese [13] demonstrate this by adapting the approach of Devroye and Lu [5] for proving the almost sure convergence of the rescaled maximum degree in the Directed Acyclic Graphs model (DAG) (the multigraph case of the RRT model) and using it for the analysis of the maximum degree in the WRG model, as discussed above. Hence, we survey the development of the properties of the maximum degree of the RRT model.

Szymańsky was the first to study the maximum degree of the RRT model and proved its convergence of the mean; $\mathbb{E}[\Delta_n/\log n] \to 1/\log 2$. Later, Devroye and Lu [5] extend this to almost sure convergence and extended this to the DAG model as well. Goh and Schmutz [9] showed that $\Delta_n - \lfloor \log_2 n \rfloor$ converges in distribution along suitable subsequences and identified possible distributions for the limit. Adarrio-Berry and Eslava [1] provide a precise characterisation of the subsequential limiting distribution of rescaled large degrees in terms of a Poisson point process as well as a central limit theorem result for near-maximum degrees (of order $\log_2 n - i_n$ where $i_n \to \infty, i_n = o(\log n)$). Eslava [6] extends this to the joint convergence of the degree and depth of high degree vertices.

In this paper we adapt part of the techniques developed by Adarrio-Berry and Eslava in [1]. They consist of two main components: First, they establish an equivalence between the RRT model and the Kingman n-coalescent and use this to provide a detailed asymptotic description of the tail distribution of the degrees of k vertices selected uniformly at random, for any $k \in \mathbb{N}$. The Kingman n-coalescent is a process which starts with n trees, each consisting of only a single root. Then, at every step 1 through $n-1$, a pair of roots is selected uniformly at random and independently of this selection, each possibility with probability 1/2, one of the two roots is connected to the other with a directed edge. This reduces the number of trees by one and, after $n-1$ steps, yields a directed tree. It turns out that this directed tree is equal in law to the random recursive tree. In the n-coalescent all n roots in the initialisation are equal in law and the degrees of the vertices are exchangeable. This allows Adarrio-Berry and Eslava to obtain the degree tail distribution with a precise error rate. Second, this precise tail distribution is used to obtain joint factorial moments of the quantities

$$\begin{aligned} X_i^{(n)} &:= |\{j \in [n] : \mathcal{Z}_n(j) = \lfloor \log_2 n \rfloor + i\}|, \quad i \in \mathbb{Z}, \\ X_{\geq i}^{(n)} &:= |\{j \in [n] : \mathcal{Z}_n(j) \geq \lfloor \log_2 n \rfloor + i\}|, \quad i \in \mathbb{Z}, \end{aligned} \tag{1.1}$$

where $\mathcal{Z}_n(j)$ denotes the in-degree of vertex j in the tree of size n. The joint factorial moments of these $X_i^{(n)}, X_i^{(n)}$ are used to identify the limiting distribution of high degrees in the tree. The sub-sequential convergence, as mentioned above, is due to the floor function applied to $\log_2 n$ and the integer-valued in-degrees $\mathcal{Z}_n(j)$.

For the WRT model, however, it provides no advantage to construct a 'weighted' Kingman n-coalescent in order to obtain precise asymptotic expression for the tail distribution of vertex degrees. As pairs of roots in the Kingman n-coalescent are selected uniformly at random and hence the roots are equal in law, it is not necessary to keep track of which roots are selected at what step. In a weighted version of the Kingman n-coalescent, pairs of roots would have to be selected with probabilities proportional to their weights, so that it is necessary to record which roots are selected at which step. As a result, a weighted Kingman n-coalescent is not (more) useful in analysing the tail distribution of vertex degrees.

Instead, we improve results on the convergence of the empirical degree distribution of the WRT model obtained by Iyer [11] and Lodewijks and Ortgiese [13]. We obtain a convergence rate to the limiting degree distribution, the asymptotic empirical degree distribution for degrees $k = k(n)$ which diverge with n, as well as asymptotic independence of degrees of vertices selected uniformly at random. We combine this with the joint factorial moments of quantities similar to (1.1) and use the techniques developed by Adarrio-Berry and Eslava [1] to derive fine asymptotics of the maximum degree in the WRT model.

Notation. Throughout the paper we use the following notation: we let $\mathbb{N} := \{1, 2, \ldots\}$ be the natural numbers, set $\mathbb{N}_0 := \{0, 1, \ldots\}$ to include zero and let $[t] := \{i \in \mathbb{N} : i \leq t\}$ for any $t \geq 1$. For $x \in \mathbb{R}$, we let $\lceil x \rceil := \inf\{n \in \mathbb{Z} : n \geq x\}$ and $\lfloor x \rfloor := \sup\{n \in \mathbb{Z} : n \leq x\}$, and for $x \in \mathbb{R}, k \in \mathbb{N}$, let $(x)_k := x(x-1)(x-2)\cdots(x-(k-1))$ and $(x)^{(k)} := x(x+1)(x+2)\cdots(x+(k-1))$. Moreover, for sequences $(a_n, b_n)_{n\in\mathbb{N}}$ such that b_n is positive for all n we say that $a_n = o(b_n), a_n \sim b_n, a_n = \mathcal{O}(b_n)$ if $\lim_{n\to\infty} a_n/b_n = 0, \lim_{n\to\infty} a_n/b_n = 1$ and if there exists a constant $C > 0$ such that $|a_n| \leq Cb_n$ for all $n \in \mathbb{N}$, respectively. For random variables $X, (X_n)_{n\in\mathbb{N}}$ we denote $X_n \xrightarrow{d} X, X_n \xrightarrow{\mathbb{P}} X$ and $X_n \xrightarrow{a.s.} X$ for convergence in distribution, probability and almost sure convergence of X_n to X, respectively. Also, we write $X_n = o_{\mathbb{P}}(1)$ if $X_n \xrightarrow{\mathbb{P}} 0$. Furthermore, we say a sequence of random variables $(X_n)_{n\in\mathbb{N}}$ is tight if for any $\varepsilon > 0$ there exists a $K_\varepsilon > 0$ such that $\limsup_{n\to\infty} \mathbb{P}(|X_n| \geq K_\varepsilon) < \varepsilon$. Finally, we use the conditional probability measure $\mathbb{P}_W(\cdot) := \mathbb{P}(\,\cdot\,|(W_i)_{i\in\mathbb{N}})$ and conditional expectation $\mathbb{E}_W[\cdot] := \mathbb{E}[\,\cdot\,|(W_i)_{i\in\mathbb{N}}]$, where the $(W_i)_{i\in\mathbb{N}}$ are the i.i.d. vertex-weights of the WRT model.

2. Definitions and main results

The weighted recursive tree (WRT) model is a growing random tree model that generalises the random recursive tree (RRT), in which vertices are assigned (random) weights and new vertices connect with existing vertices with a probability proportional to the vertex-weights.

The definition of the WRT model follows the one in [10]:

Definition 2.1 (Weighted Recursive Tree)**.** Let $(W_i)_{i\geq 1}$ be a sequence of i.i.d. copies of a positive random variable W such that $\mathbb{P}(W > 0) = 1$ and set

$$S_n := \sum_{i=1}^{n} W_i.$$

We construct the *weighted recursive tree* as follows:

1) Initialise the tree with a single vertex 1, denoted as the root, and assign to the root a vertex-weight W_1. Denote this tree by T_1.
2) For $n \geq 1$, introduce a new vertex $n+1$ and assign to it the vertex-weight W_{n+1}. Conditionally on T_n, connect to some $i \in [n]$ with probability W_i/S_n. Denote the resulting tree by T_{n+1}.

We treat T_n as a directed tree, where edges are directed from new vertices towards old vertices.

Remark 2.2. (i) Note that the edge connection probabilities are invariant under a rescaling of the vertex-weights. In particular, we may without loss of generality assume for vertex-weight distributions with bounded support that $x_0 := \sup\{x \in \mathbb{R} \,|\, \mathbb{P}(W \le x) < 1\} = 1$.

Lodewijks and Ortgiese studied certain properties of the Weighted Recursive Graph (WRG) model in [13]. This is a more general version of the WRT model that allows every vertex to connect to $m \in \mathbb{N}$ vertices when introduced, yielding a multigraph when $m > 1$. This paper aims to recover and extend some of these results in the tree case ($m = 1$) when the vertex-weights are almost surely bounded, i.e. $x_0 < \infty$. As stated in Remark 2.2(i), we can set $x_0 = 1$ without loss of generality. To formulate the results we need to assume that the distribution of the weights is sufficiently regular, allowing us to control their extreme value behaviour. In certain cases it is more convenient to formulate the assumptions in terms of the distribution of the random variable $(1 - W)^{-1}$:

Assumption 2.3 (Vertex-weight distribution)**.** The vertex-weights $W, (W_i)_{i\in\mathbb{N}}$ are i.i.d. strictly positive random variables, which are:

- Bounded from above almost surely, such that $x_0 := \sup\{x \in \mathbb{R} | \mathbb{P}(W \le x) < 1\} = 1$.
- Bounded away from zero almost surely: $\exists\, w^* \in (0, 1)$ such that $\mathbb{P}(W \ge w^*) = 1$.

Furthermore, the vertex-weights satisfy one of the following conditions:

(**Atom**) The vertex weights follow a distribution that has an atom at one, i.e. there exists a $q_0 \in (0, 1]$ such that $\mathbb{P}(W = 1) = q_0$. (Note that $q_0 = 1$ recovers the RRT model)

(**Weibull**) The vertex-weights follow a distribution that belongs to the Weibull maximum domain of attraction (MDA). This implies that there exist some $\alpha > 1$ and positive function ℓ which is slowly varying at infinity, such that

$$\mathbb{P}(W \ge 1 - 1/x) = \mathbb{P}\big((1 - W)^{-1} \ge x\big) = \ell(x)x^{-(\alpha-1)}, \qquad x \ge (1 - w^*)^{-1}.$$

(**Gumbel**) The distribution belongs to the Gumbel maximum domain of attraction (MDA) (and $x_0 = 1$). This implies that there exist sequences $(a_n, b_n)_{n\in\mathbb{N}}$, such that

$$\frac{\max_{i\in[n]} W_i - b_n}{a_n} \xrightarrow{d} \Lambda,$$

where Λ is a Gumbel random variable.

Within this class, we further distinguish the following two sub-classes:

(**RV**) There exist $a, c_1, \tau > 0$, and $b \in \mathbb{R}$ such that

$$\mathbb{P}(W > 1 - 1/x) = \mathbb{P}\big((1 - W)^{-1} > x\big) \sim a x^b \mathrm{e}^{-(x/c_1)^\tau} \quad \text{as } x \to \infty.$$

(**RaV**) There exist $a, c_1 > 0, b \in \mathbb{R}$, and $\tau > 1$ such that

$$\mathbb{P}(W > 1 - 1/x) = \mathbb{P}\big((1 - W)^{-1} > x\big) \sim a(\log x)^b \mathrm{e}^{-(\log(x)/c_1)^\tau} \quad \text{as } x \to \infty.$$

Remark 2.4. The assumption that the vertex-weights are bounded away from zero is required only for a very specific part of the proof of Proposition 5.1. Though we were unable to omit this assumption, we believe it is a mere technicality that can be overcome or at the very least replaced by weaker conditions.

Throughout, we will write

$$\mathcal{Z}_n(i) := \text{ in-degree of vertex } i \text{ in } T_n.$$

Working with the in-degree allows us to (in principle) generalise our methods to graphs with random out-degree, as mentioned in Remark 2.2. Obviously, if the out-degree is fixed, we can recover the results for the degree from our results on the $\mathcal{Z}_n(i)$.

In [13], the following results are obtained for the WRG model: if we let $\theta_m := 1+\mathbb{E}[W]/m$, and

$$p_k := \mathbb{E}\left[\frac{\theta_m - 1}{\theta_m - 1 + W}\Big(\frac{W}{\theta_m - 1 + W}\Big)^k\right], \qquad p_{\geq k} := \sum_{j=k}^{\infty} p_k = \mathbb{E}\left[\Big(\frac{W}{\theta_m - 1 + W}\Big)^k\right], \quad (2.1)$$

then almost surely for any $k \in \mathbb{N}$ fixed,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{\mathcal{Z}_n(i)=k\}} = p_k, \qquad \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{\mathcal{Z}_n(i)\geq k\}} = p_{\geq k}, \quad (2.2)$$

whenever W follows a distribution with a finite mean. In particular, the above is satisfied for all cases in Assumption 2.3. Moreover, if the vertex-weights are bounded almost surely (without loss of generality $x_0 = 1$),

$$\max_{i\in[n]} \frac{\mathcal{Z}_n(i)}{\log_{\theta_m} n} \xrightarrow{a.s.} 1.$$

In this paper we improve these results when considering the WRT model with almost surely bounded weights (so that $m = 1$ and thus $\theta := \theta_1 = 1 + \mathbb{E}[W] \in (1, 2]$).

First, we are able to extend the result in (2.2) to the case when $k = k(n)$ that diverges with n in the sense that the difference between both sides converges to zero in mean, under certain constraints on $k(n)$, and we obtain a convergence rate as well. Combining this result with techniques developed by Addario-Berry and Eslava in [1] for random recursive trees we are then able to identify the higher-order asymptotic behaviour of the maximum degree depending on the cases in Assumption 2.3. Additionally, in certain cases we are able to derive an asymptotic tail distribution for the maximum degree and obtain an asymptotic normality result for the number of vertices with 'near-maximal' degrees (in certain cases). These results can be extended to the model with a *random out-degree* as mentioned in Remark 2.2 as well.

Define $\theta := 1 + \mathbb{E}[W]$ and

$$\begin{aligned} X_i^{(n)} &:= |\{j \in [n] : \mathcal{Z}_n(j) = \lfloor \log_\theta n\rfloor + i\}|, \\ X_{\geq i}^{(n)} &:= |\{j \in [n] : \mathcal{Z}_n(j) \geq \lfloor \log_\theta n\rfloor + i\}|. \end{aligned} \quad (2.3)$$

For certain classes of vertex-weight distributions, we can prove the distributional convergence of these quantities along subsequences, as is the case for the RRT model in [1]. This result can be formulated in terms of convergence of point processes. Let $\mathbb{Z}^* := \mathbb{Z} \cup \{\infty\}$ and endow $\mathbb{Z}^*$ with the metric $d(i,j) = |2^{-i} - 2^{-j}|$ and $d(i,\infty) = 2^{-i}, i,j \in \mathbb{Z}$, and let $\mathcal{M}^\#_{\mathbb{Z}^*}$ be the space of bounded finite measures on $\mathbb{Z}^*$. If we let $\mathcal{P}$ be a Poisson point process on $\mathbb{R}$ with intensity measure $\lambda(\mathrm{d}x) := q_0\theta^{-x}\log\theta\,\mathrm{d}x, q_0 \in (0,1]$, and define

$$\mathcal{P}^\varepsilon := \sum_{x\in\mathcal{P}} \delta_{\lfloor x+\varepsilon\rfloor}, \qquad \mathcal{P}^{(n)} := \sum_{i\in[n]} \delta_{\mathcal{Z}_n(i) - \lfloor \log_\theta n\rfloor}, \qquad \varepsilon_n := \log_\theta n - \lfloor \log_\theta n\rfloor, \quad (2.4)$$

then we can provide conditions such that $\mathcal{P}^{(n_\ell)}$ converges weakly to $\mathcal{P}^\varepsilon$, for subsequences $(n_\ell)_{\ell\in\mathbb{N}}$ such that $\varepsilon_{n_\ell} \to \varepsilon$ as $\ell \to \infty$. We abuse notation to write $\mathcal{P}^\varepsilon(i) = \mathcal{P}^\varepsilon(\{i\}) = |\{x \in \mathcal{P} : \lfloor x+\varepsilon\rfloor = i\}| = |\{x \in \mathcal{P} : x \in [i-\varepsilon, i+1-\varepsilon)\}|$.

We now state our main results, which we split into several theorems based on the cases in Assumption 2.3.

Theorem 2.5 (High degrees in WRTs **(Atom)** case)**.** *Consider the WRT model in Definition 2.1 with vertex-weights $(W_i)_{i\in[n]}$ that satisfy the* **(Atom)** *case in Assumption 2.3 for some $q_0 \in (0,1]$. Fix $\varepsilon \in [0,1]$. Let $(n_\ell)_{\ell\in\mathbb{N}}$ be a positive integer sequence such that $\varepsilon_{n_\ell} \to \varepsilon$ as $\ell \to \infty$. Then $\mathcal{P}^{(n_\ell)}$ converges weakly in $\mathcal{M}^\#_{\mathbb{Z}^*}$ to $\mathcal{P}^\varepsilon$ as $\ell \to \infty$. Equivalently, for any $i < i' \in \mathbb{Z}$, jointly as $\ell \to \infty$,*

$$(X_i^{(n_\ell)}, X_{i+1}^{(n_\ell)}, \ldots, X_{i'-1}^{(n_\ell)}, X_{\geq i'}^{(n_\ell)}) \xrightarrow{d} (\mathcal{P}^\varepsilon(i), \mathcal{P}^\varepsilon(i+1), \ldots, \mathcal{P}^\varepsilon(i'-1), \mathcal{P}^\varepsilon([i',\infty)).$$

We note that this result recovers and extends [1, Theorem 1.2], in which an equivalent result is presented which only holds for the random recursive tree, i.e. the particular case of the weighted recursive tree in which $q_0 := \mathbb{P}(W=1)=1$. In Theorem 2.5 we allow $q_0 \in (0,1)$ as well, under the additional assumption that $\mathbb{P}(W \geq w^*)=1$ for some $w^* \in (0,1)$.

When the vertex-weight distribution belongs to the Weibull MDA, we can prove convergence in probability under a deterministic second-order scaling, but are unable to obtain what we conjecture to be a random third-order term similar to the result in Theorem 2.5:

Theorem 2.6 (High degrees in WRTs, (**Weibull**) case). *Consider the WRT model in Definition 2.1 with vertex-weights $(W_i)_{i\in[n]}$ that satisfy the (**Weibull**) case in Assumption 2.3 for some $\alpha>1$. Then,*

$$\max_{i\in[n]} \frac{\mathcal{Z}_n(i)-\log_\theta n}{\log_\theta \log_\theta n} \xrightarrow{\mathbb{P}} -(\alpha-1).$$

Finally, when the vertex-weight distribution belongs to the Gumbel MDA, we have similar results compared to the Weibull MDA case in the above theorem. Here we are also able to obtain a deterministic second-order scaling and we provide bounds for the third- and fourth-order behaviour of the maximum degree in a particular sub-case as well:

Theorem 2.7 (High degrees in WRTs, (**Gumbel**) case). *Consider the WRT model in Definition 2.1 with vertex-weights $(W_i)_{i\in[n]}$ that satisfy the (**Gumbel**) case in Assumption 2.3. In the (**RV**) sub-case, recall $\gamma := 1/(1+\tau)$. Then,*

$$\max_{i\in[n]} \frac{\mathcal{Z}_n(i)-\log_\theta n}{(\log_\theta n)^{1-\gamma}} \xrightarrow{\mathbb{P}} -\frac{\tau^\gamma}{(1-\gamma)\log\theta}\Big(\frac{1-\theta^{-1}}{c_1}\Big)^{1-\gamma} =: -C_{\theta,\tau,c_1}. \tag{2.5}$$

*In the (**RaV**) sub-case,*

$$\max_{i\in[n]} \frac{\mathcal{Z}_n(i)-\log_\theta n + C_1(\log_\theta\log_\theta n)^\tau - C_2(\log_\theta\log_\theta n)^{\tau-1}\log_\theta\log_\theta\log_\theta n}{(\log_\theta\log_\theta n)^{\tau-1}} \xrightarrow{\mathbb{P}} C_3, \tag{2.6}$$

where

$$\begin{aligned} &C_1 := (\log\theta)^{\tau-1}c_1^{-\tau}, \qquad C_2 := (\log\theta)^{\tau-1}\tau(\tau-1)c_1^{-\tau},\\ &C_3 := \big(\log_\theta(\log\theta)(\tau-1)\log\theta - \log(\mathrm{e}c_1^\tau(1-\theta^{-1})/\tau)\big)(\log\theta)^{\tau-2}\tau c_1^{-\tau}. \end{aligned} \tag{2.7}$$

We see that only in the (**Atom**) case we are able to obtain the higher-order asymptotics up to random order. This is due to the fact that, in this particular case, the vertices with high degree all have vertex-weight one. In the other classes covered in Theorems 2.6 and 2.7 vertices with high degrees have a vertex-weight close to one, which causes their degrees to grow slightly slower. This results in the higher-order asymptotics as observed in these theorems.

We are able to obtain more precise results related to the maximum and near-maximum degree vertices in the (**Atom**) case as well, which again recover and extend the results in [1].

Theorem 2.8 (Asymptotic tail distribution for maximum degree in (**Atom**) case). *Consider the WRT model in Definition 2.1 with vertex-weights $(W_i)_{i\in[n]}$ that satisfy the (**Atom**) case in Assumption 2.3 for some $q_0\in(0,1]$. Then, for any $i=i(n)$ with $i+\log_\theta n < (\theta/(\theta-1))\log n$ and $\liminf_{n\to\infty} i > -\infty$,*

$$\mathbb{P}\Big(\max_{j\in[n]} \mathcal{Z}_n(j) \geq \lfloor \log_\theta n\rfloor + i\Big) = \big(1-\exp\{-q_0\theta^{-i+\varepsilon_n}\}\big)(1+o(1)).$$

Finally, we establish an asymptotic normality result for the number of vertices which have 'near-maximum' degrees. For a precise definition of 'near-maximum', we define sequences $(s_k, r_k)_{k\in\mathbb{N}}$ as

$$\begin{aligned} s_k &:= \inf\big\{x\in(0,1) : \mathbb{P}(W\in(x,1)) \leq \exp\{-(1-\theta^{-1})(1-x)k\}\big\},\\ r_k &:= \exp\{-(1-\theta^{-1})(1-s_k)k\}. \end{aligned} \tag{2.8}$$

As a result, r_k can be used as the error term in the asymptotic expression of $p_{\geq k}$ (as in (2.1)) when the weight distribution satisfies the **(Atom)** case (see Theorem 5.3) and is essential in quantifying how much smaller 'near-maximum' degrees are relative to the maximum degree of the graph in this case. We note that r_k is decreasing and converges to zero with k (see Lemma 8.3), and that in the definition of s_k, r_k we can allow the index to be continuous rather than just an integer (the proof of Lemma 8.3 can be adapted to still hold in this case). We can then formulate the following theorem:

Theorem 2.9 (Asymptotic normality of near-maximum degree vertices, **(Atom)** case)**.** *Consider the WRT model in Definition 2.1 with vertex-weights $(W_i)_{i\in[n]}$ that satisfy the* **(Atom)** *case in Assumption 2.3 for some $q_0 \in (0,1]$. Then, for $i = i(n) \to -\infty$ such that $i = o(\log n \wedge |\log r_{\log_\theta n}|)$,*

$$\frac{X_i^{(n)} - q_0(1-\theta^{-1})\theta^{-i+\varepsilon_n}}{\sqrt{q_0(1-\theta^{-1})\theta^{-i+\varepsilon_n}}} \xrightarrow{d} N(0,1).$$

Remark 2.10. The constraint $i = o(\log n \wedge \log r_{\log_\theta n})$ can be simplified by providing more information on the tail of the weight distribution. Only when W has an atom at one and support bounded away from one do we have that $o(\log n \wedge \log r_{\log_\theta n}) = o(\log n)$. That is, when there exists an $s \in (0,1)$ such that $\mathbb{P}(W \in (s,1)) = 0$. In that case, we can set $s_k = s$ and $r_k = \exp\{-(1-\theta^{-1})(1-s)k\}$ for all k large, so that

$$\log r_{\log_\theta n} = -(1-\theta^{-1})(1-s)\log_\theta n,$$

so that indeed $o(\log n \wedge \log r_{\log_\theta n}) = o(\log n)$. In all other cases it follows that $s_k \uparrow 1$, so that $\log r_{\log_\theta n} = o(\log n)$ and the constraint simplifies to $i = o(\log r_{\log_\theta n})$.

Outline of the paper

In Section 3 we provide a short overview and intuitive idea of the proofs of Theorems 2.5, 2.6, 2.7, 2.8 and 2.9. In Section 4 we discuss two examples of vertex-weight distributions which satisfy the **(Weibull)** and **(Gumbel)** cases, respectively, for which more precise results can be obtained. We then provide the key concepts and results that are used in the proofs of the main theorems discussed in Section 2 in Section 5 and use these results to prove the main theorems in Section 6. Finally, in Section 7 we provide the necessary techniques and results, comparable to what is presented in Section 5, to prove the statements regarding the examples of Section 4.

3. Intuitive idea of (the proof of) the main theorems

We provide a short intuitive idea as to why the results stated in Section 2 hold.

The main elements in obtaining a more precise understanding of the behaviour of the maximum degree of the WRT are the following:

(i) A precise expression of the tail distribution of the in-degree of uniformly at random selected vertices $(v_\ell)_{\ell\in[k]}$, for any $k \in \mathbb{N}$. That is,

$$\mathbb{P}(\mathcal{Z}_n(v_\ell) \geq m_\ell, \ell \in [k]) = \prod_{\ell=1}^{k} p_{\geq m_\ell}(1 + o(n^{-\beta})),$$

for some $\beta > 0$ and where the $m_\ell \in \mathbb{N}$ are such that $m_\ell < c\log n$ for some $c \in (0, \theta/(\theta-1))$. This extends (2.2) in the sense of convergence in mean to $k \in \mathbb{N}$ many uniformly at random selected vertices rather than just one, and allows the m_ℓ to grow with n rather than being fixed. Moreover, the error term $1 + o(n^{-\beta})$ extends previously known results as well, for which no convergence rate was known.

(ii) The asymptotic behaviour of $p_{\geq k}$, as defined in (2.1), as $k \to \infty$ for each case in Assumption 2.3.

(i), which is proved in Proposition 5.1, allows us to obtain bounds on the probability of the event $\{\max_{j\in[n]}\mathcal{Z}_n(j)\geq k_n\}$ for any sequence $k_n\to\infty$ as $n\to\infty$. These probabilities can be expressed in terms of $np_{\geq k_n}$ (using union bounds and the Chung-Erdős inequality), as is shown in Lemma 5.8. By (ii) we can then precisely quantify k_n such that these bounds either tend to zero or one, which implies whether $\{\max_{j\in[n]}\mathcal{Z}_n(i)\geq k_n\}$ does or does not hold with high probability. This is the main approach for Theorems 2.6 and 2.7.

To obtain the random limits described in terms of the Poisson process $\mathcal{P}^\varepsilon$, as in Theorem 2.5, we use a similar approach as in [1]. Both (i) and (ii) are still essential, but are now used to obtain factorial moments of the quantities $X_i^{(n)}$ and $X_{\geq i}^{(n)}$, defined in (2.3), as shown in Proposition 5.6. More specifically, for any $i<i'\in\mathbb{Z}$ and $a_i,\dots,a_{i'}\in\mathbb{N}_0$, and recalling that $(x)_k:=x(x-1)\dots(x-(k-1))$,

$$\mathbb{E}\left[\left(X_{\geq i'}^{(n)}\right)_{a_{i'}}\prod_{k=i}^{i'-1}\left(X_i^{(n)}\right)_{a_i}\right]=\left(q_0\theta^{-i'+\varepsilon_n}\right)^{a_{i'}}\prod_{k=i}^{i'-1}\left(q_0(1-\theta^{-1})\theta^{-k+\varepsilon_n}\right)^{a_k}(1+o(1)). \tag{3.1}$$

We stress that the specific form of the right-hand side is due to the underlying assumption in Theorem 2.5 that the vertex-weight distribution has an atom at one, as in the **(Atom)** case of Assumption 2.3. The error term can be specified in more detail, but we omit this as it serves no further purpose here. The result essentially follows directly from these estimates by observing that the right-hand side of (3.1) can be understood as the factorial moment of the Poisson random variables $\mathcal{P}^\varepsilon([i-\varepsilon,i+1-\varepsilon)),\dots,\mathcal{P}^\varepsilon([i'-\varepsilon,\infty))$, when ε_n converges to some ε.

The equality in (3.1) follows from the fact that $X_i^{(n)}$ and $X_{\geq i}^{(n)}$ can be expressed as sums of indicator random variables of disjoint events, so that their factorial means can be understood via the probabilities in (i). Then, again using the asymptotic behaviour of $p_{\geq k}$ (as in (ii)), allows us to obtain the right-hand side of (3.1).

Finally, Theorems 2.8 and 2.9 are also a result of (3.1). This is due to the fact that the events $\{\max_{j\in[n]}\mathcal{Z}_n(j)\geq\lfloor\log_\theta n\rfloor+i\}$ can be understood via the events $\{X_{\geq i}^{(n)}>0\}$. Again using ideas similar to ones developed in [1] then allow us to obtain the results.

4. Examples

In this section we discuss some particular choices of distributions for the vertex-weights for which more precise statements can be made compared to those stated in Section 2. The reason we can improve on these more general results is due to a better understanding of the asymptotic behaviour of p_k and $p_{\geq k}$ (see (2.1)) as $k\to\infty$. As discussed in Section 3, to understand the asymptotic behaviour of the (near-)maximum degree(s) up to random order a very precise asymptotic expression for $p_{\geq k}$ is required. Though not possible in general in the **(Weibull)** and **(Gumbel)** cases of Assumption 2.3, certain choices of vertex-weight distributions do allow for a more explicit formulation of $p_{\geq k}$, yielding improved asymptotics. The proofs of the results presented here are deferred to Section 7.

Example 4.1 ('Beta' distribution bounded away from zero)**.** We consider a random variable W with a tail distribution

$$\mathbb{P}(W\geq x)=Z_{w^*}\int_x^1\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}s^{\alpha-1}(1-s)^{\beta-1}\,\mathrm{d}s,\qquad x\in[w^*,1), \tag{4.1}$$

for some $\alpha,\beta>0,w^*\in(0,1)$ and where Z_{w^*} is a normalising term to ensure that $\mathbb{P}(W\geq w^*)=1$. W can be interpreted as a beta random variable, conditionally on $\{W\geq w^*\}$. We set, for $\theta:=1+\mathbb{E}[W]\in(1,2)$,

$$\begin{aligned}X_i^{(n)}&:=|\{j\in[n]:\mathcal{Z}_n(j)=\lfloor\log_\theta n-\beta\log_\theta\log_\theta n\rfloor+i\}|,\\ X_{\geq i}^{(n)}&:=|\{j\in[n]:\mathcal{Z}_n(j)\geq\lfloor\log_\theta n-\beta\log_\theta\log_\theta n\rfloor+i\}|,\\ \varepsilon_n&:=(\log_\theta n-\beta\log_\theta\log_\theta n)-\lfloor\log_\theta n-\beta\log_\theta\log_\theta n\rfloor.\end{aligned}$$

Then, we can formulate the following results.

Theorem 4.2. *Consider the WRT model in Definition 2.1 with vertex-weights $(W_i)_{i\in[n]}$ whose distribution satisfies (4.1) for some $\alpha,\beta>0, w^*\in(0,1)$ and fix $\varepsilon\in(0,1)$. Let $(n_\ell)_{\ell\in\mathbb{N}}$ be an increasing integer sequence satisfying $\varepsilon_{n_\ell}\to\varepsilon$ as $\ell\to\infty$ and let $\mathcal{P}$ be a Poisson point process on $\mathbb{R}$ with intensity measure $\lambda(x)=Z_{w^*}(\Gamma(\alpha+\beta)/\Gamma(\alpha))(1-\theta^{-1})^{-\beta}\theta^{-x}\log\theta\,\mathrm{d}x$. Define*

$$\mathcal{P}^\varepsilon:=\sum_{x\in\mathcal{P}}\delta_{\lfloor x+\varepsilon\rfloor},\qquad \mathcal{P}^{(n)}:=\sum_{i\in[n]}\delta_{\mathcal{Z}_n(i)-\lfloor\log_\theta n-\beta\log_\theta\log_\theta n\rfloor}.$$

Then in $\mathcal{M}^\#_{\mathbb{Z}^*}$ (the space of bounded finite measures on $\mathbb{Z}^*=\mathbb{Z}\cup\{\infty\}$), $\mathcal{P}^{(n_\ell)}$ converges weakly to $\mathcal{P}^\varepsilon$ as $\ell\to\infty$. Equivalently, for any $i<i'\in\mathbb{Z}$, jointly as $\ell\to\infty$,

$$(X_i^{(n_\ell)},X_{i+1}^{(n_\ell)},\ldots,X_{i'-1}^{(n_\ell)},X_{\geq i'}^{(n_\ell)})\xrightarrow{d}(\mathcal{P}^\varepsilon(i),\mathcal{P}^\varepsilon(i+1),\ldots,\mathcal{P}^\varepsilon(i'-1),\mathcal{P}^\varepsilon([i',\infty)).$$

We remark that the second-order term $\beta\log_\theta\log_\theta n$ is established in Theorem 2.6 as well and that the above theorem recovers this result and extends it to the random third-order term, which is similar to the result in Theorem 2.5.

Theorem 4.3. *Consider the WRT model in Definition 2.1 with vertex-weights $(W_i)_{i\in[n]}$ whose distribution satisfies (4.1) for some $\alpha,\beta>0, w^*\in(0,1)$. Then, for any $i=i(n)$ with $i\sim\delta\log_\theta n$ for some $\delta\in[0,1/(\theta-1))$ ($\delta=0$ denotes $i=o(\log n)$) and $\liminf_{n\to\infty}i>-\infty$,*

$$\mathbb{P}\Big(\max_{j\in[n]}\mathcal{Z}_n(j)\geq\lfloor\log_\theta n-\beta\log_\theta\log_\theta n\rfloor+i\Big)$$
$$=\Big(1-\exp\Big\{-Z_{w^*}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)}\frac{(1-\theta^{-1})^{1-\beta}}{(\theta-1)(1+\delta)^\beta}\theta^{-i+1+\varepsilon_n}\Big\}\Big)(1+o(1)).$$

Theorem 4.4. *Consider the WRT model in Definition 2.1 with vertex-weights $(W_i)_{i\in[n]}$ whose distribution satisfies (4.1) for some $\alpha,\beta>0, w^*\in(0,1)$, and set $c_{\alpha,\beta,\theta}:=Z_{w^*}(\Gamma(\alpha+\beta)/\Gamma(\alpha))(1-\theta^{-1})^{1-\beta}$. Then, for $i=i(n)\to-\infty$ such that $i=o(\log\log n)$,*

$$\frac{X_i^{(n)}-c_{\alpha,\beta,\theta}\theta^{-i+\varepsilon_n}}{\sqrt{c_{\alpha,\beta,\theta}\theta^{-i+\varepsilon_n}}}\xrightarrow{d}N(0,1).$$

The three theorems are the analogue of Theorems 2.5, 2.8 and 2.9, respectively, where we now consider vertex-weights distributed according to a distribution as in (4.1) rather than a distribution with an atom at one.

Example 4.5 (Fraction of 'gamma' random variables)**.** We consider a random variable W with a tail distribution

$$\mathbb{P}(W\geq x)=Z_{w^*}(1-x)^{-b}\mathrm{e}^{-x/(c_1(1-x))},\qquad x\in[w^*,1),\tag{4.2}$$

for some $b\in\mathbb{R}, c_1>0, w^*\in(0,1)$ and where Z_{w^*} is a normalising term to ensure that $\mathbb{P}(W\geq w^*)=1$. $(1-W)^{-1}$ belongs to the Gumbel maximum domain of attraction, as

$$\mathbb{P}\big((1-W)^{-1}\geq x\big)=\mathbb{P}(W\geq 1-1/x)=Z_{w^*}\mathrm{e}^{1/c_1}x^b\mathrm{e}^{-x/c_1},\qquad x\geq(1-w^*)^{-1},$$

so that W belongs to the Gumbel MDA as well by [13, Lemma 2.6], and satisfies the **(Gumbel)**-**(RV)** sub-case with $a=Z_{w^*}\mathrm{e}^{1/c_1}, b\in\mathbb{R}, c_1>0, \tau=1$. As a result $X:=(1-W)^{-1}$ is a 'gamma' random variable in the sense that its tail distribution is asymptotically equal to that of a gamma random variable, up to constants. W can then be written as $W=(X-1)/X$, a fraction of these 'gamma' random variables.

Recall C_{θ,τ,c_1} from (2.5). We set, for $\theta:=1+\mathbb{E}[W]\in(1,2)$,

$$C:=\mathrm{e}^{c_1^{-1}(1-\theta^{-1})/2}\sqrt{\pi}c_1^{-1/4+b/2}(1-\theta^{-1})^{1/4+b/2},\tag{4.3}$$

and

$$\begin{aligned}
X_i^{(n)} &:= \big|\big\{j\in[n]: \mathcal{Z}_n(j)=\big\lfloor \log_\theta n - C_{\theta,1,c_1}\sqrt{\log_\theta n}+(b/2+1/4)\log_\theta\log_\theta n\big\rfloor+i\big\}\big|,\\
X_{\geq i}^{(n)} &:= \big|\big\{j\in[n]: \mathcal{Z}_n(j)\geq\big\lfloor \log_\theta n - C_{\theta,1,c_1}\sqrt{\log_\theta n}+(b/2+1/4)\log_\theta\log_\theta n\big\rfloor+i\big\}\big|,\\
\varepsilon_n &:= \big(\log_\theta n - C_{\theta,1,c_1}\sqrt{\log_\theta n}+(b/2+1/4)\log_\theta\log_\theta n\big)\\
&\quad -\big\lfloor \log_\theta n - C_{\theta,1,c_1}\sqrt{\log_\theta n}+(b/2+1/4)\log_\theta\log_\theta n\big\rfloor.
\end{aligned}$$

Then, we can formulate the following results.

Theorem 4.6. *Consider the WRT model in Definition 2.1 with vertex-weights $(W_i)_{i\in[n]}$ whose distribution satisfies (4.2) for some $b\in\mathbb{R}, c_1>0, w^*\in(0,1)$ and recall C_{θ,τ,c_1} and C from (2.6) and (4.3), respectively. Then,*

$$\max_{i\in[n]}\frac{\mathcal{Z}_n(i)-\log_\theta n + C_{\theta,1,c_1}\sqrt{\log_\theta n}}{\log_\theta\log_\theta n}\overset{\mathbb{P}}{\longrightarrow}\frac{b}{2}+\frac{1}{4}. \tag{4.4}$$

Furthermore, fix $\varepsilon\in(0,1)$ and let $(n_\ell)_{\ell\in\mathbb{N}}$ be an increasing integer sequence satisfying $\varepsilon_{n_\ell}\to\varepsilon$ as $\ell\to\infty$. Let $\mathcal{P}$ be a Poisson point process on $\mathbb{R}$ with intensity measure $\lambda(x)=Z_{w^*}C\theta^{C^2_{\theta,1,c_1}/2-x}\log\theta\,\mathrm{d}x$. Define

$$\mathcal{P}^\varepsilon:=\sum_{x\in\mathcal{P}}\delta_{\lfloor x+\varepsilon\rfloor},\qquad \mathcal{P}^{(n)}:=\sum_{i\in[n]}\delta_{\mathcal{Z}_n(i)-\lfloor \log_\theta n - C_{\theta,1,c_1}\sqrt{\log_\theta n}+(b/2+1/4)\log_\theta\log_\theta n\rfloor}.$$

Then in $\mathcal{M}^\#_{\mathbb{Z}^*}$ (the space of bounded finite measures on $\mathbb{Z}^*=\mathbb{Z}\cup\{\infty\}$), $\mathcal{P}^{(n_\ell)}$ converges weakly to $\mathcal{P}^\varepsilon$ as $\ell\to\infty$. Equivalently, for any $i<i'\in\mathbb{Z}$, jointly as $\ell\to\infty$,

$$(X_i^{(n_\ell)},X_{i+1}^{(n_\ell)},\dots,X_{i'-1}^{(n_\ell)},X_{\geq i'}^{(n_\ell)})\overset{d}{\longrightarrow}(\mathcal{P}^\varepsilon(i),\mathcal{P}^\varepsilon(i+1),\dots,\mathcal{P}^\varepsilon(i'-1),\mathcal{P}^\varepsilon([i',\infty))).$$

We remark that the second-order term in (4.4) is established in Theorem 2.7, (2.5), as well. The above theorem recovers this former result and extends it to the third-order rescaling and to the random fourth-order term, which is similar to the result in Theorem 2.5.

Theorem 4.7. *Consider the WRT model in Definition 2.1 with vertex-weights $(W_i)_{i\in[n]}$ whose distribution satisfies (4.2) for some $b\in\mathbb{R}, c_1>0, w^*\in(0,1)$ and recall C_{θ,τ,c_1} and C from (2.6) and (4.3), respectively. Then, for any $i=i(n)$ with $i\sim\delta\sqrt{\log_\theta n}$ for some $\delta\geq 0$ ($\delta=0$ denotes $i=o(\sqrt{\log_\theta n})$) and $\liminf_{n\to\infty} i>-\infty$,*

$$\begin{aligned}
&\mathbb{P}\Big(\max_{j\in[n]}\mathcal{Z}_n(j)\geq\lfloor\log_\theta n - C_{\theta,1,c_1}\sqrt{\log_\theta n}+(b/2+1/4)\log_\theta\log_\theta n\rfloor+i\Big)\\
&\quad=\Big(1-\exp\Big\{-Z_{w^*}\frac{C}{\theta-1}\theta^{-i+1+\varepsilon_n+C_{\theta,1,c_1}(C_{\theta,1,c_1}-\delta)/2}\Big\}\Big)(1+o(1)).
\end{aligned}$$

Theorem 4.8. *Consider the WRT model in Definition 2.1 with vertex-weights $(W_i)_{i\in[n]}$ whose distribution satisfies (4.2) for some $b\in\mathbb{R},_1>0,w^*\in(0,1)$, recall C_{θ,τ,c_1} and C from (2.6) and (4.3), respectively, and set $c_{\theta,c_1}:=Z_{w^*}C\theta^{C^2_{\theta,1,c_1}/2}$. Then, for $i=i(n)\to-\infty$ such that $i=o(\log\log n)$,*

$$\frac{X_i^{(n)}-c_{\theta,c_1}\theta^{-i+\varepsilon_n}}{\sqrt{c_{\theta,c_1}\theta^{-i+\varepsilon_n}}}\overset{d}{\longrightarrow}N(0,1).$$

The three theorems are the analogue of Theorems 2.5, 2.8 and 2.9, respectively, where we now consider vertex-weights distributed according to a distribution as in (4.2) rather than a distribution with an atom at one.

In both examples we see that a better understanding of the asymptotic behaviour of the tail of the degree distribution, $(p_{\geq k})_{k\in\mathbb{N}_0}$, allows us to identify the higher-order asymptotic behaviour of the (near-)maximum degree(s). It also shows that a higher order random limit as in the sense of Theorems 4.2 and 4.6 is not expressed just by vertex-weights whose distribution has an atom at one, and we conjecture that this result is in fact universal for *all* vertex-weights distributions with bounded support.

5. Degree tail distributions and factorial moments

In this section we state and prove the key elements required to prove the main results as stated in Section 2. We stress that the results presented and proved in this section cover all the classes introduced in Assumption 2.3 (in fact, they cover any vertex-weight W such that $\sup\{x > 0 : \mathbb{P}(W \leq x) < 1\} = 1, \mathbb{P}(W \geq w^*) = 1$ for some $w^* \in (0,1)$) and that the distinction between the classes of Assumption 2.3 follows in Section 6.

5.1. **Statement of results and main ideas.** As discussed in Section 3, to understand the asymptotic behaviour of the maximum degree and near-maximum degrees we require a more precise understanding of the convergence in mean of the empirical degree distribution. To that end, we present the following result:

Proposition 5.1 (Distribution of typical vertex degrees)**.** *Let W be a positive random variable such that $x_0 := \sup\{x > 0 : \mathbb{P}(W \leq x) < 1\} = 1$ and such that there exists a $w^* \in (0,1)$ so that $\mathbb{P}(W \geq w^*) = 1$. Consider the WRT model in Definition 2.1 with vertex-weights $(W_i)_{i\in[n]}$ which are i.i.d. copies of W, fix $k \in \mathbb{N}$ and let $(v_\ell)_{\ell\in[k]}$ be k vertices selected uniformly at random without replacement from $[n]$. For a fixed $c \in (0, \theta/(\theta-1))$, there exist $\beta \geq \beta' > 0$ such that uniformly over non-negative integers $m_\ell < c\log n, \ell \in [k]$,*

$$\mathbb{P}(\mathcal{Z}_n(v_\ell) = m_\ell, \ell \in [k]) = \prod_{\ell=1}^{k} \mathbb{E}\left[\frac{\mathbb{E}[W]}{\mathbb{E}[W]+W}\Big(\frac{W}{\mathbb{E}[W]+W}\Big)^{m_\ell}\right](1+o(n^{-\beta})), \quad (5.1)$$

and

$$\mathbb{P}(\mathcal{Z}_n(v_\ell) \geq m_\ell, \ell \in [k]) = \prod_{\ell=1}^{k} \mathbb{E}\left[\Big(\frac{W}{\mathbb{E}[W]+W}\Big)^{m_\ell}\right](1+o(n^{-\beta'})). \quad (5.2)$$

Remark 5.2. *(i)* In [5, Lemma 1], it is proved that the degrees $(\mathcal{Z}_n(j))_{j\in[n]}$ are negative quadrant dependent when considering the RRT model (the WRT with deterministic weights, all equal to 1). That is, for any $k \in \mathbb{N}$ and $j_1 \neq \ldots \neq j_k \in [n], m_1, \ldots, m_k \in \mathbb{N}$,

$$\mathbb{P}\left(\bigcap_{\ell=1}^{k} \mathcal{Z}_n(j_\ell) \geq m_\ell\right) \leq \prod_{j=1}^{k} \mathbb{P}(\mathcal{Z}_n(j_\ell) \geq m_\ell).$$

This property only holds for the conditional probability measure $\mathbb{P}_W$ when considering the WRT (or, more generally, the WRG) model, as follows from [13, Lemma 7.1], and can be obtained 'asymptotically' for the probability measure $\mathbb{P}$, as in the proof of [13, Theorem 2.8, Bounded case]. Proposition 5.1 improves on this by establishing *asymptotic independence* under the non-conditional probability measure $\mathbb{P}$ of the degrees of typical vertices, which allows us to extend the results in [13] to more precise asymptotics.

(ii) We note that the result only requires the two main conditions in Assumption 2.3. Hence, results for other vertex-weight distributions that do not satisfy any of the particular cases outlined in this assumption can be obtained as well using the methods presented in this paper.

(iii) The result in Proposition 5.1 improves on known results, especially those in [8, 11]. In these papers similar techniques are used to prove a weaker result, in which the m_ℓ are not allowed to diverge with n and where no convergence rate is provided.

To use this (tail) distribution of k typical vertices $v_1, \ldots, v_k$, a precise expression for the expected values on the right-hand side in Proposition 5.1 is required. Recall p_k from (2.1). The following theorem comes from [13, Theorem 2.7], in which the maximum degree of weighted recursive graphs is studied for a large class of vertex-weight distribution and in which asymptotic expressions of p_k are presented.

Theorem 5.3 ([13], Asymptotic behaviour of p_k)**.** *Recall that $\theta := 1+\mathbb{E}[W]$. We consider the different cases with respect to the vertex-weights as in Assumption 2.3, and can relax the assumption that W is bounded away from zero, i.e. $w^* = 0$ is allowed.*

(**Atom**) *Recall that $q_0 = \mathbb{P}(W = 1) > 0$ and recall r_k from (2.8). Then,*

$$p_k = q_0(1-\theta^{-1})\theta^{-k}\big(1+\mathcal{O}(r_k)\big).$$

(**Weibull**) *Recall that $\alpha > 1$ is the power-law exponent. Then, for all $k > 1/\mathbb{E}[W]$,*

$$\underline{L}(k)k^{-(\alpha-1)}\theta^{-k} \le p_k \le \overline{L}(k)k^{-(\alpha-1)}\theta^{-k}, \tag{5.3}$$

where $\underline{L}, \overline{L}$ are slowly varying at infinity.

(**Gumbel**) *(i) If W satisfies the* (**RV**) *sub-case with parameter $\tau > 0$, set $\gamma := 1/(\tau+1)$. Then,*

$$p_k = \exp\Big\{-\frac{\tau^\gamma}{1-\gamma}\Big(\frac{(1-\theta^{-1})k}{c_1}\Big)^{1-\gamma}(1+o(1))\Big\}\theta^{-k}. \tag{5.4}$$

(ii) If W satisfies the (**RaV**) *sub-case with parameter $\tau > 1$,*

$$p_k = \exp\Big\{-\Big(\frac{\log k}{c_1}\Big)^\tau\Big(1-\tau(\tau-1)\frac{\log\log k}{\log k}+\frac{K_{\tau,c_1,\theta}}{\log k}(1+o(1))\Big)\Big\}\theta^{-k}. \tag{5.5}$$

where $K_{\tau,c_1,\theta} := \tau\log(\mathrm{e}c_1^\tau(1-\theta^{-1})/\tau)$.

Remark 5.4. Equivalent upper and lower bounds can be obtained for $p_{\ge k}$, as in (2.1), by adjusting constants only.

We also provide less precise but more general bounds on the degree distribution.

Lemma 5.5. *Let W be a positive random variable with $x_0 := \sup\{x > 0 : \mathbb{P}(W \le x) < 1\} = 1$. Then, for any $\xi > 0$ and k sufficiently large,*

$$(\theta+\xi)^{-k} \le p_k \le p_{\ge k} \le \theta^{-k}.$$

Proof. The upper bound on $p_{\ge k}$ directly follows from the fact that $x \mapsto (x/(\theta-1+x))^k$ is increasing in x, so that

$$p_{\ge k} = \mathbb{E}\Big[\Big(\frac{W}{\theta-1+W}\Big)^k\Big] \le \Big(\frac{x_0}{\theta-1+x_0}\Big)^k = \theta^{-k},$$

when $x_0 = 1$. For the lower bound, let us take some $\delta \in (0, \xi/(\theta-1+\xi))$ and define

$$f_k(\theta, x) := \frac{\mathbb{E}[W]}{\mathbb{E}[W]+x}\Big(\frac{x}{\mathbb{E}[W]+x}\Big)^k = \frac{\theta-1}{\theta-1+x}\Big(\frac{x}{\theta-1+x}\Big)^k. \tag{5.6}$$

Note that $p_k = \mathbb{E}[f_k(\theta, W)]$. Then, since $f_k(\theta, x)$ is increasing in x on $(0,1]$ for k sufficiently large,

$$\mathbb{E}[f_k(\theta, W)] \ge \mathbb{E}\big[f_k(\theta, W)\mathbb{1}_{\{W>1-\delta\}}\big] \ge \mathbb{P}(W > 1-\delta)\,\frac{\theta-1}{\theta-\delta}\Big(\frac{1-\delta}{\theta-\delta}\Big)^k.$$

We note that, since $x_0 = 1$, $\mathbb{P}(W > 1-\delta) > 0$ for any $\delta > 0$. Now, by the choice of δ, $(\theta+\xi)(1-\delta)/(\theta-\delta) > 1$, so we can find some $\gamma > 0$ sufficiently small so that

$$\mathbb{E}[f_k(\theta, W)]\,(\theta+\xi)^k \ge \mathbb{P}(W > 1-\delta)\,\frac{\theta-1}{\theta-\delta}\Big(\frac{(\theta+\xi)(1-\delta)}{\theta-\delta}\Big)^k \ge (1+\gamma)^k \ge 1,$$

as required. $\square$

Recall the definition of $X_i^{(n)}, X_{\ge i}^{(n)}$ and ε_n from (2.3) and (2.4), respectively. Proposition 5.1 combined with Theorem 5.3 then allows us to obtain the following result.

Proposition 5.6 (Factorial moments for vertex-weights satisfying the (**Atom**) case)**.** *Consider the WRT model as in Definition 2.1 with vertex-weights $(W_i)_{i\in[n]}$ that satisfy the* (**Atom**) *case in Assumption 2.3 for some $q_0 \in (0,1]$. Recall r_k from (2.8), recall that $\theta := 1+\mathbb{E}[W]$ and that $(x)_k := x(x-1)\cdots(x-(k-1))$ for $x\in\mathbb{R}, k\in\mathbb{N}_0$. For a fixed $K\in\mathbb{N}, c\in(0,\theta/(\theta-1))$, there exists a $\beta>0$ such that the following holds. For any $i=i(n), i'=i'(n)$ in $\mathbb{Z}$ such that $0<i+\log_\theta n<i'+\log_\theta n<c\log n$ and $a_j\in\mathbb{N}_0, j\in\{i,\dots,i'\}$ such that $\sum_{j=i}^{i'} a_j = K$,*

$$\mathbb{E}\left[\left(X_{\geq i'}^{(n)}\right)_{a_{i'}}\prod_{k=i}^{i'-1}\left(X_k^{(n)}\right)_{a_k}\right] = \left(q_0\theta^{-i'+\varepsilon_n}\right)^{a_{i'}}\prod_{k=i}^{i'-1}\left(q_0(1-\theta^{-1})\theta^{-k+\varepsilon_n}\right)^{a_k} \times\left(1+\mathcal{O}\left(r_{\lfloor\log_\theta n\rfloor+i}\vee n^{-\beta}\right)\right).$$

Remark 5.7. Related to Remark 2.10, the error term decays polynomially only if W has an atom at one and support bounded away from one and $\log_\theta n+i>\eta\log n$ for some $\eta>0$. That is, when there exists an $s\in(0,1)$ such that $\mathbb{P}(W\in(s,1))=0$. In that case, $s_k\leq s$ and $r_k\leq\exp\{-(1-\theta^{-1})(1-s)k\}$ for all k large, so that

$$r_{\lfloor\log_\theta n\rfloor+i}\vee n^{-\beta}\leq\exp\{-(1-\theta^{-1})(1-s)\eta\log n\}\vee n^{-\beta} = n^{-\min\{\eta(1-\theta^{-1})(1-s),\beta\}}.$$

In all other cases, the error term decays slower than polynomially.

Proof of Proposition 5.6 subject to Proposition 5.1. We closely follow the approach in [1, Proposition 2.1], where an analogue result in presented and proved for the case $q_0=1$, i.e. for the random recursive tree. Set $K':=K-a_{i'}$ and for each $i\leq k\leq i'$ and each $u\in\mathbb{N}$ such that $\sum_{\ell=i}^{k-1}a_\ell<u\leq\sum_{\ell=i}^{k}a_\ell$, let $m_u=\lfloor\log_\theta n\rfloor+k$. We note that $m_u<\log_\theta n+i'<c\log n$, so that the results in Proposition 5.1 can be used. Also, let $(v_u)_{u\in[K]}$ be K vertices selected uniformly at random without replacement from $[n]$, and define $I:=[K]\backslash[K']$. Then, as the $X_{\geq k}^{(n)}$ and $X_k^{(n)}$ can be expressed as sums of indicators,

$$\begin{aligned}\mathbb{E}\Big[\Big(X_{\geq i'}^{(n)}\Big)_{a_{i'}}\prod_{k=i}^{i'-1}\Big(X_k^{(n)}\Big)_{a_k}\Big] &= (n)_K\mathbb{P}\big(\mathcal{Z}_n(v_u)=m_u, \mathcal{Z}_n(v_w)\geq m_w, u\in[K'], w\in I\big)\\ &= (n)_K\sum_{\ell=0}^{K'}\sum_{\substack{S\subseteq[K']\\|S|=\ell}}(-1)^\ell\mathbb{P}\big(\mathcal{Z}_n(v_u)\geq m_u+\mathbb{1}_{\{u\in S\}}, u\in[K]\big),\end{aligned}\tag{5.7}$$

where the second step follows from [1, Lemma 5.1] and is based on an inclusion-exclusion argument. We can now use Proposition 5.1. First, we note that there exists a $\beta>0$ such that for non-negative integers $m_1,\dots,m_K<c\log n$,

$$\mathbb{P}\big(\mathcal{Z}_n(v_u)\geq m_u+\mathbb{1}_{\{u\in S\}}, u\in[K]\big) = \prod_{u=1}^K\mathbb{E}\Big[\Big(\frac{W}{\mathbb{E}[W]+W}\Big)^{m_u+\mathbb{1}_{\{u\in S\}}}\Big]\big(1+o\big(n^{-\beta}\big)\big).\tag{5.8}$$

Now, by Theorem 5.3 and the definition of r_k in (2.8) and as r_k is decreasing by Lemma 8.3 in the Appendix, when $|S|=\ell$,

$$\prod_{u=1}^K\mathbb{E}\left[\Big(\frac{W}{\mathbb{E}[W]+W}\Big)^{m_u+\mathbb{1}_{\{u\in S\}}}\right] = q_0^K\theta^{-\ell-\sum_{u=1}^K m_u}\big(1+\mathcal{O}\big(r_{\lfloor\log_\theta n\rfloor+i}\vee n^{-\beta'}\big)\big),\tag{5.9}$$

as the smallest m_u equals $\lfloor\log_\theta n\rfloor+i$. We have

$$\begin{aligned}(n)_K\sum_{\ell=0}^{K'}\sum_{\substack{S\subseteq[K']\\|S|=\ell}}(-1)^\ell q_0^K\theta^{-\ell-\sum_{u=1}^K m_u} &= (n)_K q_0^K\theta^{-\sum_{u=1}^K m_u}\sum_{\ell=0}^{K'}\binom{K'}{\ell}(-1)^\ell\theta^{K'-\ell}\\ &= (n)_K q_0^K(1-\theta^{-1})^{K'}\theta^{-\sum_{u=1}^K m_u}.\end{aligned}\tag{5.10}$$

We then observe that $(n)_K = \theta^{K\log_\theta n}(1+\mathcal{O}(1/n))$. Moreover, we recall that $K = \sum_{k=i}^{i'} a_k, K' = \sum_{k=i}^{i'-1} a_k$ and $m_u = \lfloor \log_\theta n\rfloor + k$ if $\sum_{\ell=i}^{k-1} a_\ell \leq u < \sum_{\ell=i}^{k} a_\ell$ for $i \leq \ell \leq i'$, and recall ε_n from (2.4). Using (5.10) combined with (5.8) and (5.9) in (5.7), we obtain

$$\begin{aligned}\mathbb{E}\Bigg[\Big(X^{(n)}_{\geq i'}\Big)_{a_{i'}}\prod_{k=i}^{i'-1}\Big(X^{(n)}_k\Big)_{a_k}\Bigg] &= q_0^K(1-\theta^{-1})^{K'}\theta^{K\log_\theta n-\sum_{u=1}^K m_u}\big(1+\mathcal{O}\big(r_{\lfloor \log_\theta n\rfloor+i}\vee n^{-\beta}\big)\big)\\ &= \Big(q_0\theta^{-i'+\varepsilon_n}\Big)^{a_{i'}}\prod_{k=i}^{i'-1}\Big(q_0(1-\theta^{-1})\theta^{-k+\varepsilon_n}\Big)^{a_k}\\ &\quad\times\big(1+\mathcal{O}\big(r_{\lfloor \log_\theta n\rfloor+i}\vee n^{-\beta}\big)\big),\end{aligned}$$

as desired. □

The next lemma builds on [13, Lemma 7.1] and [5, Lemma 1] and provides bounds on the maximum degree that hold with high probability.

Lemma 5.8. *Consider the WRT model in Definition 2.1. Fix $c \in (0,\theta/(\theta-1))$ and let $(k_n)_{n\in\mathbb{N}}$ be a non-negative, diverging integer sequence such that $k_n < c\log n$ and let v_1 be a vertex selected uniformly at random from $[n]$. If $\lim_{n\to\infty} n\mathbb{P}(\mathcal{Z}_n(v_1)\geq k_n)=0$, then*

$$\lim_{n\to\infty}\mathbb{P}\Big(\max_{i\in[n]}\mathcal{Z}_n(i)\geq k_n\Big)=0.$$

Similarly, when instead $\lim_{n\to\infty} n\mathbb{P}(\mathcal{Z}_n(v_1)\geq k_n)=\infty$,

$$\lim_{n\to\infty}\mathbb{P}\Big(\max_{i\in[n]}\mathcal{Z}_n(i)\geq k_n\Big)=1.$$

Remark 5.9. Similar to what is discussed in Remark 5.2(i), the result in this lemma is stronger than the results presented in [13, Lemma 7.1] and [5, Lemma 1]. It extends the latter to the WRT model rather than just the RRT model, and improves the former as the result holds for the non-conditional probability measure $\mathbb{P}$ rather than $\mathbb{P}_W$, which is what is used in [13]. Due to the difficulties of working with the conditional probability measure, only a first order asymptotic result can be proved there. With the improved understanding of the degree distribution, as in Proposition 5.1, the above result can be obtained, which allows for finer asymptotics to be proved.

Proof of Lemma 5.8 subject to Proposition 5.1. The first result immediately follows from a union bound and the fact that

$$n\mathbb{P}(\mathcal{Z}_n(v_1)\geq k_n)=\sum_{i=1}^n\mathbb{P}(\mathcal{Z}_n(i)\geq k_n)\,. \tag{5.11}$$

For the second result, let $A_{n,i} := \{\mathcal{Z}_n(i)\geq k_n\}, i\in[n]$. Then, by the Chung-Erdős inequality,

$$\mathbb{P}\Big(\max_{i\in[n]}\mathcal{Z}_n(i)\geq k_n\Big)=\mathbb{P}(\cup_{i=1}^n A_{n,i})\geq\frac{\Big(\sum_{i=1}^n\mathbb{P}(A_{n,i})\Big)^2}{\sum_{i\neq j}\mathbb{P}(A_{n,i}\cap A_{n,j})+\sum_{i=1}^n\mathbb{P}(A_{n,i})}.$$

By (5.11) it follows that $\sum_{i=1}^n\mathbb{P}(A_{n,i})=n\mathbb{P}(\mathcal{Z}_n(v_1)\geq k_n)$. Furthermore, by Proposition 5.1,

$$\sum_{i\neq j}\mathbb{P}(A_{n,i}\cap A_{n,j})=n(n-1)\mathbb{P}(A_{n,v_1}\cap A_{n,v_2})=(n\mathbb{P}(A_{n,v_1}))^2(1+o(1)),$$

where v_2 is another vertex selected uniformly at random, unequal to v_1. Note that the condition that $k_n < c\log n$ is required for this to hold. Together with the above lower bound, these two observations yield

$$\mathbb{P}(\cup_{i=1}^n A_{n,i})\geq\frac{(n\mathbb{P}(A_{n,v_1}))^2}{(n\mathbb{P}(A_{n,v_1}))^2(1+o(1))+n\mathbb{P}(A_{n,v_1})}=\frac{n\mathbb{P}(A_{n,v_1})}{n\mathbb{P}(A_{n,v_1})\,(1+o(1))+1}.$$

Hence, when $n\mathbb{P}(A_{n,v_1}) = n\mathbb{P}(\mathcal{Z}_n(v_1) \geq k_n)$ diverges with n, we obtain the desired result. $\square$

What remains is to prove Proposition 5.1. As the proof is rather long and involved, we discuss the strategy of the proof first and take care of certain parts of the proof in separate lemmas. The main part of the proof is dedicated to proving (5.1). Once this is established, (5.2) follows without much effort. We thus focus on discussing the proof of (5.1).

The left-hand side of (5.1) can be expressed by conditioning on the values of the typical vertices, and splitting between cases of young and old vertices. That is,

$$\begin{aligned}\mathbb{P}(\mathcal{Z}_n(v_\ell) = m_\ell, \ell \in [k]) &= \frac{1}{(n)_k} \sum_{1 \leq j_1 \neq \ldots \neq j_k \leq n} \mathbb{E}\left[\mathbb{P}_W(\mathcal{Z}_n(j_\ell) = m_\ell, \ell \in [k])\right] \\ &= \frac{1}{(n)_k} \sum_{n^{1-\varepsilon} \leq j_1 \neq \ldots \neq j_k \leq n} \mathbb{E}\left[\mathbb{P}_W(\mathcal{Z}_n(j_\ell) = m_\ell, \ell \in [k])\right] \\ &\quad + \frac{1}{(n)_k} \sum_{\mathbf{j} \in I_n(\varepsilon)} \mathbb{E}\left[\mathbb{P}_W(\mathcal{Z}_n(j_\ell) = m_\ell, \ell \in [k])\right],\end{aligned} \tag{5.12}$$

where for any $\varepsilon \in (0,1)$, $I_n(\varepsilon) := \{\mathbf{j} = (j_1, \ldots, j_k) : 1 \leq j_1 \neq \ldots \neq j_k \leq n,\ \exists i \in [k]\ j_i < n^{1-\varepsilon}\}$. Splitting the sum on the first line into the two sums on the second and third line allows us to deal with them in a different way. In the sum on the second line, in which all indices are at least $n^{1-\varepsilon}$, we can apply the law of large numbers on sums of vertex-weights to gain more control over the conditional probability of the event $\{\mathcal{Z}_n(j_\ell) = m_\ell, \ell \in [k]\}$. The aim is to show that this first sum has the desired form, as on the right-hand side of (5.1).

The sum on the third line, in which at least one of the indices takes on values strictly smaller than $n^{1-\varepsilon}$ can be shown to be negligible compared to the first sum. Especially when m_ℓ is large, this is non-trivial. To do this, we consider the tail events $\{\mathcal{Z}_n(j_\ell) \geq m_\ell, \ell \in [k]\}$ and use the negative quadrant dependence of the degrees (see Remark 5.2 and [13, Lemma 7.1]), so that we can deal with the more tractable probabilities $\mathbb{P}_W(\mathcal{Z}_n(j_\ell) \geq m_\ell)$ for $\ell \in [k]$, rather than the probability of all tail degree events. Depending on whether the indices in $I_n(\varepsilon)$ are at most or at least $n^{1-\varepsilon}$, we then use bounds similar to one developed in the proof of [13, Lemma 7.1] or use an approach similar to what we use to bound the sum on the second line of (5.12), respectively.

In the following lemma, we deal with the sum on the second line of (5.12).

Lemma 5.10. *Let W be a positive random variable such that $x_0 := \sup\{x > 0 : \mathbb{P}(W \leq x) < 1\} = 1$. Consider the WRT model in Definition 2.1 with vertex-weights $(W_i)_{i\in[n]}$ which are i.i.d. copies of W and fix $k \in \mathbb{N}, c \in (0, \theta/(\theta-1))$. Then, there exist a $\beta > 0$ and $\varepsilon \in (0,1)$ such that uniformly over non-negative integers $m_\ell < c\log n, \ell \in [k]$,*

$$((n)_k)^{-1} \sum_{n^{1-\varepsilon} \leq j_1 \neq \ldots \neq j_k \leq n} \mathbb{E}\left[\mathbb{P}_W(\mathcal{Z}_n(j_\ell) = m_\ell, \ell \in [k])\right] = \prod_{\ell=1}^{k} \mathbb{E}\left[\frac{\mathbb{E}[W]}{\mathbb{E}[W]+W}\Big(\frac{W}{\mathbb{E}[W]+W}\Big)\right]\big(1+o(n^{-\beta})\big).$$

We note that the assumption $\mathbb{P}(W \geq w^*) = 1$ for some $w^* \in (0,1)$ is not required for this result to hold. To prove this lemma, we sum over all possible m_ℓ vertices that connect to j_ℓ for each $\ell \in [k]$ and use the fact that the $j_1, \ldots, j_k$ are at least $n^{1-\varepsilon}$ to precisely control the connection probabilities and to evaluate the sums over all the possible m_ℓ vertices, $\ell \in [k]$, as well as the sum over the indices $j_1, \ldots, j_k$.

In the following lemma, we show the sum on the third line of (5.12) is negligible compared to the sum on the second line.

Lemma 5.11. *Let W be a positive random variable such that $x_0 := \sup\{x > 0 : \mathbb{P}(W \le x) < 1\} = 1$ and such that there exists a $w^* \in (0,1)$ so that $\mathbb{P}(W \ge w^*) = 1$. Consider the WRT model in Definition 2.1 with vertex-weights $(W_i)_{i\in[n]}$ which are i.i.d. copies of W and fix $k \in \mathbb{N}, c \in (0, \theta/(\theta-1))$. Then, there exist an $\eta > 0$ and an $\varepsilon \in (0,1)$ such that uniformly over non-negative integers $m_\ell < c\log n, \ell \in [k]$,*

$$\frac{1}{(n)_k}\sum_{\boldsymbol{j}\in I_n(\varepsilon)}\mathbb{E}\left[\mathbb{P}_W(\mathcal{Z}_n(j_\ell) = m_\ell, \ell\in[k])\right] = o\bigg(\prod_{\ell=1}^k \mathbb{E}\left[\frac{\mathbb{E}[W]}{\mathbb{E}[W]+W}\Big(\frac{W}{\mathbb{E}[W]+W}\Big)^{m_\ell}\right]n^{-\eta}\bigg),$$

where $I_n(\varepsilon) := \{\boldsymbol{j} = (j_1,\ldots,j_k) : 1 \le j_1 \ne \ldots \ne j_k \le n, \exists\, i \in [k]\ j_i < n^{1-\varepsilon}\}$.

Note that the assumption that the vertex-weights are bounded away from zero is required in this lemma, where it was not necessary in Lemma 5.10. In fact, it is required for one inequality in the proof only, which convinces us that it could be omitted with more work or at the very least replaced by weaker assumptions.

It is clear that (5.2) in Proposition 5.1 immediately follows from using the results of Lemmas 5.10 and 5.11 in (5.12). In what follows we first prove Lemma 5.10 in Section 5.2, prove Lemma 5.11 in Section 5.3 and finally complete the proof of Proposition 5.1 in Section 5.4.

5.2. **Proof of Lemma 5.10.**

Proof of Lemma 5.10. To prove the result we provide a matching upper bound and lower bound (up to error terms) for

$$((n)_k)^{-1}\sum_{n^{1-\varepsilon}\le j_1\ne\ldots\ne j_k\le n}\mathbb{E}\left[\mathbb{P}_W(\mathcal{Z}_n(j_\ell) = m_\ell, \ell\in[k])\right].$$

Upper bound
Let us introduce the event

$$E_n := \bigg\{\sum_{\ell=1}^j W_\ell \in ((1-\zeta_n)\mathbb{E}[W]\,j, (1+\zeta_n)\mathbb{E}[W]\,j),\ \forall\ n^{1-\varepsilon} \le j \le n\bigg\}, \tag{5.13}$$

where $\zeta_n = n^{-\delta(1-\varepsilon)}/\mathbb{E}[W]$ for some $\delta \in (0, 1/2)$. By noting that $\widetilde{S}_j := \sum_{\ell=1}^j W_\ell - j\mathbb{E}[W]$ is a martingale, that $|\widetilde{S}_j - \widetilde{S}_{j-1}| \le 1 + \mathbb{E}[W] = \theta$ and that $\zeta_n \ge j^{-\delta}/\mathbb{E}[W]$ for $j \ge n^{1-\varepsilon}$, we can use the Azuma-Hoeffding inequality to obtain

$$\mathbb{P}(E_n^c) \le \sum_{j\ge n^{1-\varepsilon}}\mathbb{P}\Big(|\widetilde{S}_j| \ge \zeta_n j\mathbb{E}[W]\Big) \le 2\sum_{j\ge n^{1-\varepsilon}}\exp\Big\{-\frac{j^{1-2\delta}}{2\theta^2}\Big\}. \tag{5.14}$$

Writing $c_\theta := 1/(2\theta^2)$, we further bound the sum from above by

$$2\int_{\lfloor n^{1-\varepsilon}\rfloor}^\infty \exp\big\{-c_\theta x^{1-2\delta}\big\}\,\mathrm{d}x = 2\frac{c_\theta^{-1/(1-2\delta)}}{1-2\delta}\Gamma\Big(\frac{1}{1-2\delta}, c_\theta\lfloor n^{1-\varepsilon}\rfloor^{1-2\delta}\Big), \tag{5.15}$$

where $\Gamma(a,x)$ is the incomplete Gamma function, and we note that the right-hand side is $o(n^{-\gamma})$ for any $\gamma > 0$ (and thus of smaller order than $\prod_{\ell=1}^k p_{m_\ell} n^{-\beta}$ for any $\beta > 0$ and uniformly in $m_1, \ldots m_k < (\theta/(\theta-1))\log n$). This yields the upper bound

$$\begin{aligned}&\frac{1}{(n)_k}\sum_{n^{1-\varepsilon}\le j_1\ne\ldots\ne j_k\le n}\mathbb{E}\left[\mathbb{P}_W(\mathcal{Z}_n(j_\ell) = m_\ell, \ell\in[k])\right]\\ &\le ((n)_k)^{-1}\sum_{n^{1-\varepsilon}\le j_1\ne\ldots\ne j_k\le n}\mathbb{E}[\mathbb{P}_W(\mathcal{Z}_n(j_\ell) = m_\ell, \ell\in[k])\,\mathbb{1}_{E_n}] + \mathcal{O}\Big(\Gamma\Big(\frac{1}{1-2\delta}, c_\theta\lfloor n^{1-\varepsilon}\rfloor^{1-2\delta}\Big)\Big),\end{aligned} \tag{5.16}$$

Now, to express the first term in (5.16) we consider ordered indices $j_\ell, \ell \in [k]$, rather than unordered ones. We provide details for the case $n^{1-\varepsilon} \le j_1 < j_2 < \ldots < j_k \le n$ and discuss later on how the other permutations of $j_1, \ldots, j_k$ can be dealt with.

Moreover, for every $\ell \in [k]$, we introduce the ordered indices $j_\ell < i_{1,\ell} < \ldots < i_{m_\ell,\ell} \leq n, \ell \in [k]$, which denote the steps at which vertex ℓ increases it degree by one. Note that for every $\ell \in [k]$ these indices are distinct by definition, but we also require that $i_{s,\ell} \neq i_{t,j}$ for any $\ell, j \in [k], s \in [m_\ell], t \in [m_j]$ (equality is allowed only when $\ell = j$ and $s = t$). We denote this constraint by adding a $*$ on the summation symbol.

Finally, we define $j_{k+1} := n$. Combining these additional steps, we arrive at

$$\begin{aligned}
&\frac{1}{(n)_k} \sum_{n^{1-\varepsilon} \leq j_1 < \ldots < j_k \leq n} \mathbb{E}\left[\mathbb{P}_W(\mathcal{Z}_n(j_\ell) = m_\ell, \ell \in [k])\, \mathbb{1}_{E_n}\right] \\
&= \frac{1}{(n)_k} \sum_{n^{1-\varepsilon} \leq j_1 < \ldots < j_k \leq n} \sideset{}{^*}\sum_{\substack{j_\ell < i_{1,\ell} < \ldots < i_{m_\ell,\ell} \leq n, \\ \ell \in [k]}} \mathbb{E}\Bigg[\prod_{t=1}^{k} \prod_{s=1}^{m_t} \frac{W_{j_t}}{\sum_{\ell=1}^{i_{s,t}-1} W_\ell} \\
&\qquad \times \prod_{u=1}^{k} \prod_{\substack{s=j_u+1 \\ s \neq i_{\ell,t}, \ell \in [m_t], t \in [k]}}^{j_{u+1}} \left(1 - \frac{\sum_{\ell=1}^{u} W_{j_\ell}}{\sum_{\ell=1}^{s-1} W_\ell}\right) \mathbb{1}_{E_n}\Bigg].
\end{aligned}$$

We then include the terms where $s = i_{\ell,t}$ for $\ell \in [m_t], t \in [k]$ in the second double product. To do this, we need to change the first double product to

$$\prod_{t=1}^{k} \prod_{s=1}^{m_t} \frac{W_{j_t}}{\sum_{\ell=1}^{i_{s,t}-1} W_\ell - \sum_{\ell=1}^{k} W_{j_\ell} \mathbb{1}_{\{i_{s,t} > j_\ell\}}} \leq \prod_{t=1}^{k} \prod_{s=1}^{m_t} \frac{W_{j_t}}{\sum_{\ell=1}^{i_{s,t}-1} W_\ell - k}, \tag{5.17}$$

that is, we subtract the vertex-weight W_{j_ℓ} in the numerator when the vertex j_ℓ has already been introduced by step $i_{s,t}$. In the upper bound we use that the weights are bounded from above by one. We thus arrive at the upper bound

$$\begin{aligned}
&\frac{1}{(n)_k} \sum_{n^{1-\varepsilon} \leq j_1 < \ldots < j_k \leq n} \sideset{}{^*}\sum_{\substack{j_\ell < i_{1,\ell} < \ldots < i_{m_\ell,\ell} \leq n, \\ \ell \in [k]}} \mathbb{E}\Bigg[\prod_{t=1}^{k} \prod_{s=1}^{m_t} \frac{W_{j_t}}{\sum_{\ell=1}^{i_{s,t}-1} W_\ell - k} \\
&\qquad\qquad \times \prod_{u=1}^{k} \prod_{s=j_u+1}^{j_{u+1}} \left(1 - \frac{\sum_{\ell=1}^{u} W_{j_\ell}}{\sum_{\ell=1}^{s-1} W_\ell}\right) \mathbb{1}_{E_n}\Bigg].
\end{aligned}$$

For ease of writing, we omit the first sum until we actually intend to sum over the indices $j_1, \ldots, j_k$. We use the bounds from the event E_n to bound

$$\sum_{\ell=1}^{i_{s,t}-1} W_\ell \geq (i_{s,t} - 1)\mathbb{E}[W](1 - \zeta_n), \qquad \sum_{\ell=1}^{s-1} W_\ell \leq s\mathbb{E}[W](1 + \zeta_n).$$

For n sufficiently large, we observe that $(i_{s,t} - 1)\mathbb{E}[W](1 - \zeta_n) - k \geq i_{s,t}\mathbb{E}[W](1 - 2\zeta_n)$, so that we obtain

$$((n)_k)^{-1} \sideset{}{^*}\sum_{\substack{j_\ell < i_{1,\ell} < \ldots < i_{m_\ell,\ell} \leq n, \\ \ell \in [k]}} \mathbb{E}\Bigg[\prod_{t=1}^{k} \prod_{s=1}^{m_t} \frac{W_{j_t}}{i_{s,t})\mathbb{E}[W](1 - 2\zeta_n)} \prod_{u=1}^{k} \prod_{s=j_u+1}^{j_{u+1}} \left(1 - \frac{\sum_{\ell=1}^{u} W_{j_\ell}}{s\mathbb{E}[W](1 + \zeta_n)}\right) \mathbb{1}_{E_n}\Bigg].$$

Moreover, relabelling the vertex-weights W_{j_t} to W_t for $t \in [k]$ does not change the distribution of the terms within the expected value, so that the expected value remains unchanged. We can also bound the indicator from above by one, to arrive at the upper bound

$$\frac{1}{(n)_k} \sideset{}{^*}\sum_{\substack{j_\ell < i_{1,\ell} < \ldots < i_{m_\ell,\ell} \leq n, \\ \ell \in [k]}} \mathbb{E}\Bigg[\prod_{t=1}^{k} \prod_{s=1}^{m_t} \frac{W_t}{i_{s,t}\mathbb{E}[W](1 - 2\zeta_n)} \prod_{u=1}^{k} \prod_{s=j_u+1}^{j_{u+1}} \left(1 - \frac{\sum_{\ell=1}^{u} W_\ell}{s\mathbb{E}[W](1 + \zeta_n)}\right)\Bigg].$$

We bound the final product from above by

$$\begin{aligned}\prod_{s=j_u+1}^{j_{u+1}}\Big(1-\frac{\sum_{\ell=1}^u W_\ell}{s\mathbb{E}[W](1+\zeta_n)}\Big) &\leq \exp\Big\{-\frac{1}{\mathbb{E}[W](1+\zeta_n)}\sum_{s=j_u+1}^{j_{u+1}}\frac{\sum_{\ell=1}^u W_\ell}{s}\Big\}\\ &\leq \exp\Big\{-\frac{1}{\mathbb{E}[W](1+\zeta_n)}\sum_{\ell=1}^{u}W_\ell\log\Big(\frac{j_{u+1}}{j_u+1}\Big)\Big\}\\ &= \Big(\frac{j_{u+1}}{j_u+1}\Big)^{-\sum_{\ell=1}^u W_\ell/(\mathbb{E}[W](1+\zeta_n))}.\end{aligned} \tag{5.18}$$

As the weights are almost surely bounded by one, we thus find

$$\begin{aligned}\prod_{s=j_u+1}^{j_{u+1}}\Big(1-\frac{\sum_{\ell=1}^u W_\ell}{s\mathbb{E}[W](1+\zeta_n)}\Big) &\leq \Big(\frac{j_{u+1}}{j_u}\Big)^{-\sum_{\ell=1}^u W_\ell/(\mathbb{E}[W](1+\zeta_n))}\Big(1+\frac{1}{j_u}\Big)^{k/(\mathbb{E}[W](1+\zeta_n))}\\ &= \Big(\frac{j_{u+1}}{j_u}\Big)^{-\sum_{\ell=1}^u W_\ell/(\mathbb{E}[W](1+\zeta_n))}\Big(1+\mathcal{O}\big(n^{-(1-\varepsilon)}\big)\Big).\end{aligned}$$

As a result, we obtain the upper bound

$$\begin{aligned}&((n)_k)^{-1}\sum^{*}_{\substack{j_\ell<i_{1,\ell}<\ldots<i_{m_\ell,\ell}\leq n,\\ \ell\in[k]}}\mathbb{E}\Bigg[\prod_{t=1}^k\Big(\Big(\frac{W_t}{\mathbb{E}[W]}\Big)^{m_t}\prod_{s=1}^{m_t}\frac{1}{i_{s,t}(1-2\zeta_n)}\Big)\prod_{u=1}^k\Big(\frac{j_{u+1}}{j_u}\Big)^{-\sum_{\ell=1}^u W_\ell/(\mathbb{E}[W](1+\zeta_n))}\Bigg]\\ &\quad\times\Big(1+\mathcal{O}\big(n^{-(1-\varepsilon)}\big)\Big)\\ &= ((n)_k)^{-1}\sum^{*}_{\substack{j_\ell<i_{1,\ell}<\ldots<i_{m_\ell,\ell}\leq n,\\ \ell\in[k]}}(1-2\zeta_n)^{-\sum_{t=1}^k m_t}\mathbb{E}\Bigg[\prod_{t=1}^k\Big(\frac{W_t}{\mathbb{E}[W]}\Big)^{m_t}\\ &\quad\times\prod_{t=1}^k\Big(j_t^{W_t/(\mathbb{E}[W](1+\zeta_n))}\prod_{s=1}^{m_t}i_{s,t}^{-1}\Big)n^{-\sum_{\ell=1}^k W_\ell/(\mathbb{E}[W](1+\zeta_n))}\Bigg]\Big(1+\mathcal{O}\big(n^{-(1-\varepsilon)}\big)\Big),\end{aligned}$$

where in the last step we recall that $j_{k+1}=n$. We then bound this from above even further by no longer constraining the indices $i_{s,t}$ to be distinct. That is, for different $t_1,t_2\in[k]$, we allow $i_{s_1,t_1}=i_{s_2,t_2}$ to hold for any $s_1\in[m_{t_1}], s_2\in[m_{t_2}]$. This yields

$$\begin{aligned}&\frac{1}{(n)_k}\sum_{\substack{j_\ell<i_{1,\ell}<\ldots<i_{m_\ell,\ell}\leq n,\\ \ell\in[k]}}(1-2\zeta_n)^{-\sum_{t=1}^k m_t}\mathbb{E}\Bigg[\prod_{t=1}^k\Big(\frac{W_t}{\mathbb{E}[W]}\Big)^{m_t}\\ &\quad\times\prod_{t=1}^k\Big(j_t^{W_t/(\mathbb{E}[W](1+\zeta_n))}\prod_{s=1}^{m_t}i_{s,t}^{-1}\Big)n^{-\sum_{\ell=1}^k W_\ell/(\mathbb{E}[W](1+\zeta_n))}\Bigg]\Big(1+\mathcal{O}\big(n^{-(1-\varepsilon)}\big)\Big).\end{aligned} \tag{5.19}$$

We set

$$a_t := W_t/(\mathbb{E}[W](1+\zeta_n)),$$

and look at the terms

$$\frac{n^{-\sum_{t=1}^k a_t}}{(n)_k}\sum_{\substack{j_\ell<i_{1,\ell}<\ldots<i_{m_\ell,\ell}\leq n,\\ \ell\in[k]}}\prod_{t=1}^k\Big((a_t(1+\zeta_n))^{m_t}j_t^{a_t}\prod_{s=1}^{m_t}i_{s,t}^{-1}\Big). \tag{5.20}$$

We bound the sums from above by multiple integrals, almost surely, which yields

$$\frac{n^{-\sum_{t=1}^k a_t}}{(n)_k}\prod_{t=1}^k(a_t(1+\zeta_n))^{m_t}j_t^{a_t}\int_{j_t}^n\int_{x_{1,t}}^n\cdots\int_{x_{m_t-1,t}}^n\prod_{s=1}^{m_t}x_{s,t}^{-1}\,\mathrm{d}x_{m_t,t}\ldots\mathrm{d}x_{1,t}. \tag{5.21}$$

By repeated substitutions of the form $u_{i,t} = \log(n/x_{i,t}), i \in [m_t - 1]$, we obtain

$$\frac{n^{-\sum_{t=1}^k a_t}}{(n)_k} \prod_{t=1}^k (a_t(1+\zeta_n))^{m_t} j_t^{a_t} \frac{\log(n/j_t)^{m_t}}{m_t!}.$$

Substituting this in (5.20) and reintroducing the sum over the indices $j_1, \ldots, j_k$, we arrive at

$$\frac{(1+\zeta_n)^{\sum_{t=1}^k m_t}}{(n)_k} n^{-\sum_{t=1}^k a_t} \prod_{t=1}^k a_t^{m_t} \sum_{n^{1-\varepsilon} \leq j_1 < \ldots < j_k \leq n} \prod_{t=1}^k j_t^{a_t} \frac{\log(n/j_t)^{m_t}}{m_t!}. \tag{5.22}$$

We observe that switching the order of the indices $j_1, \ldots, j_k$ achieves the same result as permuting the $m_1, \ldots, m_k$ and $a_1, \ldots, a_k$. Hence, if we let $\pi : [k] \to [k]$ be a permutation, then considering the indices $n^{1-\varepsilon} \leq j_{\pi(1)} < j_{\pi(2)} < \ldots < j_{\pi(k)} \leq n$ yields a similar result as in (5.22) but with a term $j_{\pi(t)}^{a_{\pi(t)}} \log(n/j_{\pi(t)})^{m_{\pi(t)}}/m_{\pi(t)}!$ in the final product. Since this product is invariant to such permutations of the m_t and a_t, the only thing that would change is the summation order of the indices $j_1, \ldots, j_k$. We will use this further on.

We then bound the sum over $n^{1-\varepsilon} \leq j_1 < \ldots < j_k \leq n$ from above by multiple integrals as well. First, we note that $j_t^{a_t} \log(n/j_t)^{m_t}$ is increasing up to $j_t = n\exp\{-m_t/a_t\}$, at which it is maximised, and decreasing for $n\exp\{-m_t/a_t\} < j_t \leq n$ for all $t \in [k]$. To provide the optimal bound, we want to know whether this maximum is attained in $[n^{1-\varepsilon}, n]$ or not. That is, whether $n\exp\{-m_t/a_t\} \in [n^{1-\varepsilon}, n]$ or not. To this end, we consider two cases:

(1) $m_t = c_t \log n(1+o(1))$ with $c_t \in [0, 1/(\theta-1)]$, $t \in [k]$ ($c_t = 0$ denotes $m_t = o(\log n)$).
(2) $m_t = c_t \log n(1+o(1))$ with $c_t \in (1/(\theta-1), c)$, $t \in [k]$.

Clearly, when $c \leq 1/(\theta-1)$ the second case can be omitted, so that without loss of generality we can assume $c > 1/(\theta-1)$. Moreover, we can assume without loss of generality that all terms $m_1, \ldots, m_k$ satisfy the same case, as a mixture of different cases can be dealt with in the same way, as will become clear in what follows. In the second case, it directly follows that the maximum is almost surely attained at

$$n\exp\{-m_t/a_t\} \leq n\exp\{-c_t \log n(\theta-1)(1+o(1))\} = n^{1-c_t(\theta-1)(1+o(1))} = o(1),$$

so that the summand $j_t^{a_t} \log(n/j_t)^{m_t}$ is almost surely decreasing in j_t when $n^{1-\varepsilon} \leq j_t \leq n$. In the first case, such a conclusion cannot be made in general and depends on the precise value of W_t. Therefore, the first case requires a more involved approach. Throughout the rest of the proof of the upper bound, we assume case (1) holds and discuss as we go along what alterations to make when case (2) holds instead. In the first case, we use Corollary 8.2 (with $g \equiv 1$) to obtain the upper bound

$$\sum_{j_k > j_{k-1}}^n \prod_{t=1}^k j_t^{a_t} \frac{\log(n/j_t)^{m_t}}{m_t!} \leq \prod_{t=1}^{k-1} j_t^{a_t} \frac{\log(n/j_t)^{m_t}}{m_t!} \left[\int_{j_{k-1}}^n x_k^{a_k} \frac{\log(n/x_k)^{m_k}}{m_k!} \, \mathrm{d}x_k + \frac{2n^{a_k}}{a_k^{m_k}}\right]. \tag{5.23}$$

Here, we use that the integrand is maximised at $x^* = n\exp\{-m_t/a_t\}$, that $(x^*)^{a_k} \log(n/(x^*)^{m_k})/m_k! = n^{a_k} m_k^{m_k}/((\mathrm{e}a_k)^{m_k} m_k!)$ and that $x^x/(\mathrm{e}^x\Gamma(x+1)) \leq 1$ for any $x > 0$. In case (2) the summand on the left-hand side is decreasing in j_k, so that we arrive at an upper bound without the additional error term $n^{a_k}/a_k^{m_k}$. Using a substitution $y_k := \log(n/x_k)$, we obtain

$$\begin{aligned} &\prod_{t=1}^{k-1} j_t^{a_t} \frac{\log(n/j_t)^{m_t}}{m_t!} \left[\frac{n^{1+a_k}}{(1+a_k)^{m_k+1}} \int_0^{\log(n/j_{k-1})} \frac{(1+a_k)^{m_k+1}}{m_k! \, \mathrm{e}^{(1+a_k)y_k}} y_k^{m_k} \, \mathrm{d}y_k + 4\frac{n^{a_k}}{a_k^{m_k}}\right] \\ &= \prod_{t=1}^{k-1} j_t^{a_t} \frac{\log(n/j_t)^{m_t}}{m_t!} \left[\frac{n^{1+a_k}}{(1+a_k)^{m_k+1}} \mathbb{P}_W(Y_k < \log(n/j_{k-1})) + 4\frac{n^{a_k}}{a_k^{m_k}}\right], \end{aligned} \tag{5.24}$$

where, conditionally on W_k, Y_k is a $\Gamma(m_k+1, 1+a_k)$ random variable. As mentioned above, in the second case the second term in the square brackets can be omitted.

The aim is to continue this approach for the summation over the remaining indices $j_{k-1},\ldots,$ j_1. We deal with the two terms we have here in different ways. Let us start with the second term. We now use the exact same approach, but almost surely bound, for $2\leq t\leq k-1$, $\mathbb{P}_W(Y_t\leq \log(n/j_{t-1}))\leq 1$, and $\mathbb{P}_W(Y_1\leq \varepsilon\log n)\leq 1$, where, conditionally on W_t, Y_t is a $\Gamma(m_t+1,1+a_t)$ random variable for each $t\in[k-1]$. Hence, for the second term we obtain

$$\begin{aligned}\sum_{j_1=\lceil n^{1-\varepsilon}\rceil}^{n}\cdots\sum_{j_{k-1}=j_{k-2}+1}^{n}&\frac{n^{a_k}}{a_k^{m_k}}\prod_{t=1}^{k-1}j_t^{a_t}\frac{\log(n/j_t)^{m_t}}{m_t!}\\ &\leq \frac{n^{a_k}}{a_k^{m_k}}\prod_{t=1}^{k-1}\Big(\frac{n^{1+a_t}}{(1+a_t)^{m_t+1}}+4\frac{n^{a_t}}{a_t^{m_t}}\Big)\\ &=n^{k-1+\sum_{t=1}^k a_t}\prod_{t=1}^{k}\frac{1}{a_t^{m_t}}\prod_{t=1}^{k-1}\Big(\frac{a_t^{m_t}}{(1+a_t)^{m_t+1}}+\frac{4}{n}\Big).\end{aligned}\tag{5.25}$$

Using this in (5.22) thus yields a term

$$\frac{1}{n}\prod_{t=1}^{k-1}\Big(\frac{a_t^{m_t}}{(1+a_t)^{m_t+1}}+\frac{4}{n}\Big)(1+\zeta_n)^{\sum_{t=1}^k m_t}\big(1+\mathcal{O}(1/n)\big).$$

Again, if for any $t\in[k-1]$, m_t satisfies case (2) the term $2/n$ can be omitted for that specific value of t in the product. If m_k satisfies case (2), then this entire term can be omitted.

To continue the summation of the first term on the right-hand side of (5.24), we again use Corollary 8.2 (now with $g(x)=\mathbb{P}_W(Y_k<\log(n/x))$), to obtain

$$\begin{aligned}\sum_{j_{k-1}=j_{k-2}+1}^{n}&\prod_{t=1}^{k-1}j_t^{a_t}\frac{\log(n/j_t)^{m_t}}{m_t!}\frac{n^{1+a_k}}{(1+a_k)^{m_k+1}}\mathbb{P}_W(Y_k\leq\log(n/j_{k-1}))\\ &\leq\frac{n^{1+a_k}}{(1+a_k)^{m_k+1}}\prod_{t=1}^{k-2}j_t^{a_t}\frac{\log(n/j_t)^{m_t}}{m_t!}\Bigg[4\frac{n^{a_{k-1}}}{a_{k-1}^{m_{k-1}}}\\ &\quad+\int_{j_{k-2}}^{n}x_{k-1}^{a_{k-1}}\frac{\log(n/x_{k-1})^{m_{k-1}}}{m_{k-1}!}\mathbb{P}_W(Y_k<\log(n/x_{k-1}))\,\mathrm{d}x_{k-1}\Bigg].\end{aligned}$$

Using a substitution $y_{k-1}:=\log(n/x_{k-1})$ yields

$$\begin{aligned}&\frac{n^{1+a_k}}{(1+a_k)^{m_k+1}}\prod_{t=1}^{k-2}j_t^{a_t}\frac{\log(n/j_t)^{m_t}}{m_t!}\Bigg[4\frac{n^{a_{k-1}}}{a_{k-1}^{m_{k-1}}}\\ &+\frac{n^{1+a_{k-1}}}{(1+a_{k-1})^{m_{k-1}+1}}\int_0^{\log(n/j_{k-2})}\int_0^{y_{k-1}}\prod_{t=k-1}^{k}\frac{(1+a_t)^{m_t+1}}{m_t!}y_t^{m_t}\mathrm{e}^{-(1+a_t)y_t}\,\mathrm{d}y_k\mathrm{d}y_{k-1}\Bigg]\\ &=2\frac{n^{1+a_k}}{(1+a_k)^{m_k+1}}\frac{n^{a_{k-1}}}{a_{k-1}^{m_{k-1}}}\prod_{t=1}^{k-2}j_t^{a_t}\frac{\log(n/j_t)^{m_t}}{m_t!}\\ &\qquad+\prod_{t=k-1}^{k}\frac{n^{1+a_t}}{(1+a_t)^{m_t+1}}\mathbb{P}_W(Y_k<Y_{k-1}<\log(n/j_{k-2}))\prod_{t=1}^{k-2}j_t^{a_t}\frac{\log(n/j_t)^{m_t}}{m_t!}.\end{aligned}$$

Using the same approach as in (5.25) for the first term on the right-hand side yields the upper bound

$$n^{k-1+\sum_{t=1}^k a_t}\prod_{t=1}^{k}\frac{1}{a_t}\prod_{\substack{t\in[k]\\ t\neq k-1}}\Big(\frac{a_t^{m_t}}{(1+a_t)^{m_t+1}}+\frac{4}{n}\Big).$$

Using this in (5.22) yields the term

$$\frac{4}{n}\prod_{\substack{t\in[k]\\t\neq k-1}}\Big(\frac{a_t^{m_t}}{(1+a_t)^{m_t+1}}+\frac{4}{n}\Big)(1+\zeta_n)^{\sum_{t=1}^k m_t}\big(1+\mathcal{O}(1/n)\big).$$

It thus follows that we can continue this approach summing over $j_{k-2},\dots,j_1$ and use this in (5.22) to obtain

$$\prod_{t=1}^k \frac{a_t^{m_t}}{(1+a_t)^{m_t+1}}\mathbb{P}_W(Y_k<\dots<Y_1<\varepsilon\log n)\,(1+\zeta_n)^{\sum_{t=1}^k m_t}$$
$$+\,\mathcal{O}\Big(\sum_{\ell=1}^k\frac{1}{n}\prod_{\substack{t\in[k]\\t\neq\ell}}\Big(\frac{a_t^{m_t}}{(1+a_t)^{m_t+1}}+\frac{4}{n}\Big)\Big),$$

where we note that we can omit the terms $(1+\zeta_n)^{\sum_{t=1}^k m_t}$ with the big $\mathcal{O}$ notation as they are $1+o(1)$ by the specific choice of ζ_n and the bound on $m_1,\dots,m_k$. Finally, using his in (5.20) and then in (5.19) yields the upper bound

$$\begin{aligned}&\mathbb{E}\Big[\prod_{t=1}^k \frac{a_t^{m_t}}{(1+a_t)^{m_t+1}}\mathbb{P}_W(Y_k<\dots<Y_1<\varepsilon\log n)\Big]\Big(\frac{1+\zeta_n}{1-2\zeta_n}\Big)^{\sum_{t=1}^k m_t}\\ &\times\Big(1+\mathcal{O}\big(n^{-(1-\varepsilon)}\big)\Big)+\mathcal{O}\Big(\sum_{\ell=1}^k\frac{1}{n}\prod_{\substack{t\in[k]\\t\neq\ell}}\Big(\mathbb{E}\Big[\frac{a_t^{m_t}}{(1+a_t)^{m_t+1}}\Big]+\frac{4}{n}\Big)\Big),\end{aligned}\tag{5.26}$$

where the term on the second line contains all the error terms created throughout and the expected value can be included within the product by the independence of the a_t.

As mentioned below (5.22), any different order of the indices $j_1,\dots,j_k$ results in a permutation of $m_1,\dots,m_k$ and $a_1,\dots,a_k$. So, if we consider summing over $n^{1-\varepsilon}\leq j_1\neq\dots\neq j_k$ rather than the ordered indices $n^{1-\varepsilon}\leq j_1<\dots<j_k$, we obtain the above term for all permutations of the a_t,m_t and Y_t. That is, if we let P_k be the set of permutations on $[k]$, we obtain

$$\begin{aligned}&\sum_{\pi\in P_k}\mathbb{E}\Big[\prod_{t=1}^k \frac{a_t^{m_t}}{(1+a_t)^{m_t+1}}\mathbb{P}_W\big(Y_{\pi(k)}<\dots<Y_{\pi(1)}<\varepsilon\log n\big)\Big]\Big(\frac{1+\zeta_n}{1-2\zeta_n}\Big)^{\sum_{t=1}^k m_t}\\ &\qquad\times\Big(1+\mathcal{O}\big(n^{-(1-\varepsilon)}\big)\Big)+\mathcal{O}\Big(\sum_{\ell=1}^k\frac{1}{n}\prod_{\substack{t\in[k]\\t\neq\ell}}\Big(\mathbb{E}\Big[\frac{a_{\pi(t)}^{m_{\pi(t)}}}{(1+a_{\pi(t)})^{m_{\pi(t)}+1}}\Big]+\frac{4}{n}\Big)\Big)\\ &=\prod_{t=1}^k\mathbb{E}\Big[\frac{a_t^{m_t}}{(1+a_t)^{m_t+1}}\mathbb{P}_W(Y_t<\varepsilon\log n)\Big]\Big(\frac{1+\zeta_n}{1-2\zeta_n}\Big)^{\sum_{t=1}^k m_t}\Big(1+\mathcal{O}\big(n^{1-\varepsilon}\big)\Big)\\ &\qquad+\mathcal{O}\Big(\sum_{\pi\in P_k}\sum_{\ell=1}^k\frac{1}{n}\prod_{\substack{t\in[k]\\t\neq\ell}}\Big(\mathbb{E}\Big[\frac{a_{\pi(t)}^{m_{\pi(t)}}}{(1+a_{\pi(t)})^{m_{\pi(t)}+1}}\Big]+\frac{4}{n}\Big)\Big),\end{aligned}$$

where the last step follows from the conditional independence of the Y_t, the independence of the a_t, and the fact that $Y_t\neq Y_s$ almost surely for $t\neq s$. We can now simply bound the conditional probability from above by one almost surely.

Since $m_t<c\log n$ for all $t\in[k]$, the fraction on the right of the expected value in the last step is $1+o(n^{-\delta(1-\varepsilon)(1-\xi)})$ for any $\xi>0$. Furthermore, within the expected values, we can write

$$\begin{aligned}\frac{a_t^{m_t}}{(1+a_t)^{m_t+1}}&=\frac{\mathbb{E}\left[W\right]}{\mathbb{E}\left[W\right]+W}\Big(\frac{W}{\mathbb{E}\left[W\right]+W}\Big)^{m_t}(1+\zeta_n)\Big(1-\frac{\zeta_n\mathbb{E}\left[W\right]}{\mathbb{E}\left[W\right](1+\zeta_n)+W}\Big)^{m_t+1}\\&=\frac{\mathbb{E}\left[W\right]}{\mathbb{E}\left[W\right]+W}\Big(\frac{W}{\mathbb{E}\left[W\right]+W}\Big)^{m_t}(1+o(n^{-\delta(1-\varepsilon)(1-\xi)})),\end{aligned}\tag{5.27}$$

almost surely. In total, combining this with (5.16) yields the final upper bound

$$\begin{aligned}&\prod_{\ell=1}^k\mathbb{E}\left[\frac{\mathbb{E}\left[W\right]}{\mathbb{E}\left[W\right]+W}\Big(\frac{W}{\mathbb{E}\left[W\right]+W}\Big)^{m_\ell}\right](1+o(n^{-\beta_1}))+\mathcal{O}\Big(\Gamma\Big(\frac{1}{1-2\delta},c_\theta\lfloor n^{1-\varepsilon}\rfloor^{1-2\delta}\Big)\Big)\\&\quad+\mathcal{O}\Bigg(\sum_{\pi\in P_k}\sum_{\ell=1}^k\frac{1}{n}\prod_{\substack{t\in[k]\\t\neq\ell}}\Bigg(\mathbb{E}\left[\frac{\mathbb{E}\left[W\right]}{\mathbb{E}\left[W\right]+W}\Big(\frac{W}{\mathbb{E}\left[W\right]+W}\Big)^{m_{\pi(t)}}\right]+\frac{4}{n}\Bigg)\Bigg),\end{aligned}\tag{5.28}$$

for some $\beta_1>0$. We then finally observe that by Lemma 5.5 and since we assumed that $m_t=c_t\log n(1+o(1))$ with $c_t\in[0,1/(\theta-1)]$ for all $t\in[k]$,

$$\mathbb{E}\left[\frac{\mathbb{E}\left[W\right]}{\mathbb{E}\left[W\right]+W}\Big(\frac{W}{\mathbb{E}\left[W\right]+W}\Big)^{m_t}\right]\geq\frac{1}{(\theta+\xi)^{m_t}}\geq n^{-\log(\theta+\xi)(1+o(1))/(\theta-1)}\geq\frac{1}{n^{1-\gamma}},\tag{5.29}$$

for some small $\gamma>0$. The final step can be made for ξ,γ sufficiently small, as $\log(x)/(x-1)<1$ for all $x\in(1,2]$. This implies that, for any $\ell\in[k]$ and for some sufficiently small $\eta>0$,

$$\begin{aligned}&\frac{1}{n}\prod_{\substack{t\in[k]\\t\neq\ell}}\Bigg(\mathbb{E}\left[\frac{\mathbb{E}\left[W\right]}{\mathbb{E}\left[W\right]+W}\Big(\frac{W}{\mathbb{E}\left[W\right]+W}\Big)^{m_{\pi(t)}}\right]+\frac{4}{n}\Bigg)\\&\quad=\frac{1}{n}\prod_{\substack{t\in[k]\\t\neq\ell}}\mathbb{E}\left[\frac{\mathbb{E}\left[W\right]}{\mathbb{E}\left[W\right]+W}\Big(\frac{W}{\mathbb{E}\left[W\right]+W}\Big)^{m_{\pi(t)}}\right](1+o(1))\\&\quad=o\Bigg(n^{-\eta}\prod_{t=1}^k\mathbb{E}\left[\frac{\mathbb{E}\left[W\right]}{\mathbb{E}\left[W\right]+W}\Big(\frac{W}{\mathbb{E}\left[W\right]+W}\Big)^{m_t}\right]\Bigg).\end{aligned}$$

As a result, the second big $\mathcal{O}$ term in (5.28) can be incorporated in the $o(n^{-\beta_1})$ term when β_1 is taken sufficiently small and all m_t satisfy case (1). When all the m_t satisfy case (2), then the second big $\mathcal{O}$ term can be omitted entirely. Moreover, since the term in the first big $\mathcal{O}$ term is $o(n^{-\gamma})$ for any $\gamma>0$, this term can also be incorporated in the $o(n^{-\beta_1})$ term as well, as $m_\ell<c\log n$ for all $\ell\in[k]$. We thus obtain for both cases that

$$\mathbb{P}(\mathcal{Z}_n(v_\ell)=m_\ell,\ell\in[k])\leq\prod_{\ell=1}^k\mathbb{E}\left[\frac{\mathbb{E}\left[W\right]}{\mathbb{E}\left[W\right]+W}\Big(\frac{W}{\mathbb{E}\left[W\right]+W}\Big)^{m_\ell}\right](1+o(n^{-\beta_1})).\tag{5.30}$$

When the m_t are such that some of the c_t satisfy case (1) and some satisfy case (2), that is, $c_t\in[0,1/(\theta-1)]$ for some $t\in[k]$ and $c_t\in(1/(\theta-1),c)$ for the other indices t, then a combined approach can be used to yield (5.30).

Lower bound
We then focus on proving a similar lower bound. We define the event

$$\widetilde{E}_n:=\Big\{\sum_{\ell=k+1}^j W_\ell\in(\mathbb{E}\left[W\right](1-\zeta_n)j,\mathbb{E}\left[W\right](1+\zeta_n)j),\ \forall\, n^{1-\varepsilon}\leq j\leq n\Big\}.$$

With similar computations as in (5.14) it follows that $\mathbb{P}(\widetilde{E}_n)=(1-o(n^{-\gamma}))$ for any $\gamma>0$. We obtain a lower bound for the probability of the event $\{\mathcal{Z}_n(v_\ell)=m_\ell,\ell\in[k]\}$ by omitting the second term in (5.16). This yields

$$
\begin{aligned}
&\frac{1}{(n)_k}\sum_{n^{1-\varepsilon}\leq j_1\neq\ldots\neq j_k\leq n}\mathbb{E}\left[\mathbb{P}_W(\mathcal{Z}_n(j_\ell)=m_\ell,\ell\in[k])\right]\\
&\quad\geq ((n)_k)^{-1}\sum_{n^{1-\varepsilon}<j_1\neq\ldots\neq j_k\leq n}\ \sideset{}{^*}\sum_{\substack{j_\ell<i_{1,\ell}<\ldots<i_{m_\ell,\ell}\leq n,\\ \ell\in[k]}}\mathbb{E}\Bigg[\prod_{t=1}^k\prod_{s=1}^{m_t}\frac{W_{j_t}}{\sum_{\ell=1}^{i_{s,t}-1}W_\ell}\prod_{u=1}^k\prod_{\substack{s=j_u+1\\ s\neq i_{\ell,t},\ell\in[m_t],t\in[k]}}^{j_{u+1}}\Bigg(1-\frac{\sum_{\ell=1}^u W_{j_\ell}}{\sum_{\ell=1}^{s-1}W_\ell}\Bigg)\Bigg].
\end{aligned}
$$

We again start by only considering the ordered indices $n^{1-\varepsilon}<j_1<\ldots<j_k$ and also omit this sum for now for ease of writing. We also omit the constraint $s\neq i_{\ell,t},\ell\in[m_t],t\in[k]$ in the final product. As this introduces more terms smaller than one, we obtain a lower bound. Then, in the two denominators, we bound the vertex-weights $W_{j_1},\ldots,W_{j_k}$ from above and below by one and zero, respectively, to obtain a lower bound

$$
\begin{aligned}
((n)_k)^{-1}\sideset{}{^*}\sum_{\substack{j_\ell<i_{1,\ell}<\ldots<i_{m_\ell,\ell}\leq n,\\ \ell\in[k]}}\mathbb{E}\Bigg[&\prod_{t=1}^k\prod_{s=1}^{m_t}\frac{W_{j_t}}{\sum_{\ell=1}^{i_{s,t}-1}W_\ell\mathbb{1}_{\{\ell\neq j_t,t\in[k]\}}+k}\\
&\times\prod_{u=1}^k\prod_{s=j_u+1}^{j_{u+1}}\Bigg(1-\frac{\sum_{\ell=1}^u W_{j_\ell}}{\sum_{\ell=1}^{s-1}W_\ell\mathbb{1}_{\{\ell\neq j_t,t\in[k]\}}}\Bigg)\Bigg].
\end{aligned}
$$

As a result, we can now swap the labels of W_{j_t} and W_t for each $t\in[k]$, which again does not change the expected value, but it changes the value of the two denominators to $\sum_{\ell=k+1}^{i_{s,t}}W_\ell+k$ and $\sum_{\ell=k+1}^{i_{s,t}}W_\ell$, respectively. After this we introduce the indicator $\mathbb{1}_{\widetilde{E}_n}$ and use the bounds in $\widetilde{E}_n$ on these sums in the expected value to obtain a lower bound. Finally, we note that the (relabelled) weights $W_t,t\in[k]$, are independent of $\widetilde{E}_n$ so that we can take the indicator out of the expected value. Combining all of the above steps, we arrive at the lower bound

$$
\begin{aligned}
((n)_k)^{-1}\sideset{}{^*}\sum_{\substack{j_\ell<i_{1,\ell}<\ldots<i_{m_\ell,\ell}\leq n,\\ \ell\in[k]}}&\mathbb{E}\Bigg[\prod_{t=1}^k\Big(\frac{W_t}{\mathbb{E}[W]}\Big)^{m_t}\prod_{s=1}^{m_t}\frac{1}{i_{s,t}(1+2\zeta_n)}\\
&\times\prod_{u=1}^k\prod_{s=j_u+1}^{j_{u+1}}\Bigg(1-\frac{\sum_{\ell=1}^u W_\ell}{(s-1)\mathbb{E}[W](1-\zeta_n)}\Bigg)\Bigg]\mathbb{P}(\widetilde{E}_n).
\end{aligned}\tag{5.31}
$$

The $1+2\zeta_n$ in the fraction on the first line arises from the fact that, for n sufficiently large, $(i_{s,t}-1)(1+\zeta_n)+k\leq i_{s,t}(1+2\zeta_n)$. As stated above, $\mathbb{P}\Big(\widetilde{E}_n\Big)=1-o(n^{-\gamma})$ for any $\gamma>0$. Similar to the calculations in (5.18) and using $\log(1-x)\geq -x-x^2$ for x small, we obtain an almost sure lower bound for the final product for n sufficiently large of the form

$$
\begin{aligned}
\prod_{s=j_u+1}^{j_{u+1}}\Bigg(1-\frac{\sum_{\ell=1}^u W_\ell}{(s-1)\mathbb{E}[W](1-\zeta_n)}\Bigg)&\geq\exp\Bigg\{-\frac{1}{\mathbb{E}[W](1-\zeta_n)}\sum_{\ell=1}^u W_\ell\sum_{s=j_u+1}^{j_{u+1}}\frac{1}{s-1}\\
&\qquad-\Big(\frac{1}{\mathbb{E}[W](1-\zeta_n)}\sum_{\ell=1}^u W_\ell\Big)^2\sum_{s=j_u+1}^{j_{u+1}}\frac{1}{(s-1)^2}\Bigg\}\\
&\geq\Big(\frac{j_{u+1}}{j_u}\Big)^{-\sum_{\ell=1}^u W_\ell/(\mathbb{E}[W](1-\zeta_n))}\Big(1-\mathcal{O}\big(n^{-(1-\varepsilon)}\big)\Big).
\end{aligned}
$$

Using this in (5.31) yields the lower bound

$$
((n)_k)^{-1}\sideset{}{^*}\sum_{\substack{j_\ell<i_{1,\ell}<\ldots<i_{m_\ell,\ell}\leq n,\\ \ell\in[k]}}(1+2\zeta_n)^{-\sum_{t=1}^k m_t}\mathbb{E}\Bigg[\prod_{t=1}^k\Big(\frac{W_t}{\mathbb{E}[W]}\Big)^{m_t}\Big(\frac{j_t}{n}\Big)^{\widetilde{a}_t}\prod_{s=1}^{m_t}i_{s,t}^{-1}\Bigg]\Big(1-\mathcal{O}\big(n^{-(1-\varepsilon)}\big)\Big),
$$

where $\widetilde{a}_t := W_t/(\mathbb{E}[W](1-\zeta_n))$. We now reintroduce the sum over the indices $n^{1-\varepsilon} \le j_1 < \ldots < j_k \le n$ and bound the sum over the indices $i_{s,\ell}$ from below. We note that the expression in the expected value is decreasing in $i_{s,\ell}$ and we restrict the range of the indices to $j_\ell + \sum_{t=1}^k m_t < i_{1,\ell} < \ldots < i_{m_\ell,\ell} \le n, \ell \in [k]$, but no longer constrain the indices to be distinct (so that we can drop the $*$ in the sum). In the distinct sums and the suggested lower bound, the number of values the $i_{s,\ell}$ take on equal

$$\prod_{\ell=1}^k \binom{n-(j_\ell-1)-\sum_{t=1}^{\ell-1} m_t}{m_\ell} \quad \text{and} \quad \prod_{\ell=1}^k \binom{n-(j_\ell-1)-\sum_{t=1}^{k} m_t}{m_\ell},$$

respectively. It is straightforward to see that the former allows for more possibilities than the latter, as $\binom{b}{c} > \binom{a}{c}$ when $b > a \ge c$. As we omit the largest values of the expected value (since it decreases in $i_{s,\ell}$ and we omit the largest values of $i_{s,\ell}$), we thus arrive at the lower bound

$$\begin{aligned}\frac{1}{(n)_k} &\sum_{n^{1-\varepsilon}<j_1<\ldots<j_k\le n-\sum_{t=1}^k m_t} \;\sum_{\substack{j_\ell+\sum_{t=1}^k m_t<i_{1,\ell}<\ldots<i_{m_\ell,\ell}\le n,\\ \ell\in[k]}} (1+2\zeta_n)^{-\sum_{t=1}^k m_t} \\ &\mathbb{E}\Bigg[\prod_{t=1}^k (\widetilde{a}_t(1-\zeta_n))^{m_t} j_t^{\widetilde{a}_t} \prod_{s=1}^{m_t} i_{s,t}^{-1} n^{-\sum_{t=1}^k \widetilde{a}_t}\Bigg]\Big(1-\mathcal{O}\big(n^{-(1-\varepsilon)}\big)\Big),\end{aligned} \tag{5.32}$$

where we also restrict the range of indices in the upper bound of the outer sum, as otherwise there would be a contribution of zero from these values of $j_1,\ldots,j_k$. We now use similar techniques compared to the upper bound of the proof to switch from summation to integration. However, due to the altered bounds on the range of the indices over which we sum and the fact that we require lower bounds rather than upper bound, we face some more technicalities.

For now, we omit the expected value and focus on the terms

$$\sum_{n^{1-\varepsilon}<j_1<\ldots<j_k\le n-\sum_{t=1}^k m_t} \;\sum_{\substack{j_\ell+\sum_{t=1}^k m_t<i_{1,\ell}<\ldots<i_{m_\ell,\ell}\le n,\\ \ell\in[k]}} \prod_{t=1}^k j_t^{\widetilde{a}_t} \prod_{s=1}^{m_t} i_{s,t}^{-1}. \tag{5.33}$$

We start by restricting the upper bound on the outer sum to $n-2\sum_{t=1}^k m_t$. This will prove useful later. We then bound the sum over the indices $i_{s,t}$ from below by

$$\begin{aligned}&\sum_{\substack{j_\ell+\sum_{t=1}^k m_t<i_{1,\ell}<\ldots<i_{m_\ell,\ell}\le n\\ \ell\in[n]}} \prod_{t=1}^k\prod_{s=1}^{m_t} i_{s,t}^{-1} \\ &\ge \prod_{\ell=1}^k \int_{j_\ell+\sum_{t=1}^k m_t+1}^{n+1}\int_{x_{1,\ell}+1}^{n+1}\cdots\int_{x_{m_\ell-1,\ell}+1}^{n+1}\prod_{s=1}^{m_\ell} x_{s,\ell}^{-1}\,\mathrm{d}x_{m_\ell,\ell}\ldots\mathrm{d}x_{1,\ell} \\ &\ge \prod_{\ell=1}^k \int_{j_\ell+\sum_{t=1}^k m_t+1}^{n+1}\int_{x_{1,\ell}+1}^{n+1}\cdots\int_{x_{m_\ell-2,\ell}+1}^{n+1}\prod_{s=1}^{m_\ell-1} x_{s,\ell}^{-1}\log\Big(\frac{n+1}{x_{m_\ell-1,\ell}+1}\Big)\,\mathrm{d}x_{m_\ell-1,\ell}\ldots\mathrm{d}x_{1,\ell}.\end{aligned}$$

The integrand can be bounded from below by using $x_{m_\ell-1,\ell}^{-1} \ge (x_{m_\ell-1,\ell}+1)^{-1}$. We also restrict the upper integration bound of the innermost integral to n and use a variable substitution $y_{m_\ell-1,\ell} := x_{m_\ell-1,\ell}+1$ to obtain the lower bound

$$\prod_{\ell=1}^k \int_{j_\ell+\sum_{t=1}^k m_t+1}^{n+1}\int_{x_{1,\ell}+1}^{n+1}\cdots\int_{x_{m_\ell-3,\ell}+1}^{n+1}\frac{1}{2}\prod_{s=1}^{m_\ell-2} x_{s,\ell}^{-1}\log\Big(\frac{n+1}{x_{m_\ell-2,\ell}+2}\Big)^2\,\mathrm{d}x_{m_\ell-2,\ell}\ldots\mathrm{d}x_{1,\ell}.$$

Continuing this approach eventually leads to

$$\prod_{\ell=1}^{k}\frac{1}{m_\ell!}\log\Big(\frac{n+1}{j_\ell+\sum_{t=1}^k m_t+m_\ell}\Big)^{m_\ell}\geq\prod_{\ell=1}^{k}\frac{1}{m_\ell!}\log\Big(\frac{n}{j_\ell+2\sum_{t=1}^k m_t}\Big)^{m_\ell}.$$

Substituting this in (5.33) with the restriction on the outer sum discussed above yields

$$\sum_{n^{1-\varepsilon}<j_1<\ldots<j_k\leq n-2\sum_{t=1}^k m_t}\prod_{\ell=1}^{k}j_\ell^{\widetilde a_\ell}\frac{1}{m_\ell!}\log\Big(\frac{n}{j_\ell+2\sum_{t=1}^k m_t}\Big)^{m_\ell}.$$

To simplify the summation over $j_1,\ldots,j_k$, we write the summand as

$$\prod_{\ell=1}^{k}\Big(j_\ell+2\sum_{t=1}^k m_t\Big)^{\widetilde a_\ell}\frac{1}{m_\ell!}\log\Big(\frac{n}{j_\ell+2\sum_{t=1}^k m_t}\Big)^{m_\ell}\Big(1-\frac{2\sum_{t=1}^k m_t}{j_\ell+2\sum_{t=1}^k m_t}\Big)^{\widetilde a_\ell}.$$

Using that $m_t<c\log n, j_\ell\geq n^{1-\varepsilon}$ and $x^{\widetilde a_t}\geq x^{1/(\mathbb{E}[W](1-\zeta_n))}$ for $x\in(0,1)$, we obtain the lower bound

$$\prod_{\ell=1}^{k}\Big(j_\ell+2\sum_{t=1}^k m_t\Big)^{\widetilde a_\ell}\frac{1}{m_\ell!}\log\Big(\frac{n}{j_\ell+2\sum_{t=1}^k m_t}\Big)^{m_\ell}\Big(1-\mathcal{O}\Big(\frac{\log n}{n^{1-\varepsilon}}\Big)\Big).$$

We can then shift the bounds on the range of the sum to $n^{1-\varepsilon}+2\sum_{t=1}^k m_t$ and n to obtain the lower bound

$$\sum_{n^{1-\varepsilon}+2\sum_{t=1}^k m_t<j_1<\ldots<j_k\leq n}\prod_{\ell=1}^{k}j_\ell^{\widetilde a_\ell}\frac{1}{m_\ell!}\log(n/j_\ell)^{m_\ell}\Big(1-\mathcal{O}\Big(\frac{\log n}{n^{1-\varepsilon}}\Big)\Big).$$

We can now use a similar approach as for the upper bound in (5.23) through (5.26) by considering the cases (1) and (2). Assuming case (1) holds for all $m_1,\ldots,m_k$ and using Corollary 8.2, we obtain the lower bound

$$\begin{aligned}&\sum_{j_1=\lceil n^{1-\varepsilon}\rceil+2\sum_{t=1}^k m_t}^{n}\ \sum_{j_2=j_1+1}^{n}\cdots\sum_{j_k=k_{k-1}+1}^{n}\prod_{\ell=1}^{k}j_\ell^{\widetilde a_\ell}\frac{1}{m_\ell!}\log(n/j_\ell)^{m_\ell}\\&\geq\sum_{j_1=\lceil n^{1-\varepsilon}\rceil+2\sum_{t=1}^k m_t}^{n}\ \sum_{j_2=j_1+1}^{n}\cdots\sum_{j_{k-1}=j_{k-2}+1}^{n}\prod_{\ell=1}^{k-1}j_\ell^{\widetilde a_\ell}\frac{1}{m_\ell!}\log(n/j_\ell)^{m_\ell}\\&\quad\times\Big[\int_{j_{k-1}+1}^{n}\frac{x_k^{\widetilde a_k}}{m_k!}\log(n/j_\ell)^{m_k}\,\mathrm{d}x_k-4\frac{n^{\widetilde a_\ell}}{a_\ell^{m_\ell}}\Big],\end{aligned}$$

and we again have for case (2) that the error term in the square brackets can be omitted. Following the same approach as in the upper bound, (5.23) through (5.26), but subtracting the error term rather than adding it, we thus obtain the lower bound

$$\begin{aligned}&\sum_{n^{1-\varepsilon}<j_1<\ldots<j_k\leq n-2\sum_{t=1}^k m_t}\prod_{\ell=1}^{k}j_\ell^{\widetilde a_\ell}\frac{1}{m_\ell!}\log\Big(\frac{n}{j_\ell+2\sum_{t=1}^k m_t}\Big)^{m_\ell}\\&\geq n^{k+\sum_{t=1}^k\widetilde a_t}\prod_{t=1}^{k}\frac{\big(1-\mathcal{O}\big(n^{-(1-\varepsilon)}\log n\big)\big)}{(1+\widetilde a_t)^{m_t+1}}\mathbb{P}_W\Big(\widetilde Y_k<\ldots<\widetilde Y_1<\log\Big(\frac{n}{\lceil n^{1-\varepsilon}\rceil+2\sum_{t=1}^k m_t}\Big)\Big)\\&\quad+\mathcal{O}\Big(n^{k+\sum_{t=1}^k\widetilde a_t}\sum_{\ell=1}^{k}\frac{1}{n}\prod_{\substack{t\in[k]\\t\neq\ell}}\Big(\frac{1}{(1+\widetilde a_t)^{m_t+1}}-\frac{4}{n\widetilde a_t^{m_t}}\Big)\Big),\end{aligned}$$

where, conditionally on W_t, $\widetilde{Y}_t$ is a $\Gamma(m_t+1, 1+\widetilde{a}_t)$ random variable for each $t\in[k]$. Using this in (5.32) then finally yields the lower bound

$$\begin{aligned}&\mathbb{E}\left[\prod_{t=1}^k \frac{\widetilde{a}_t^{m_t}}{(1+\widetilde{a}_t)^{m_t+1}}\mathbb{P}_W\left(\widetilde{Y}_k<\ldots<\widetilde{Y}_1<\log\left(\frac{n}{\lceil n^{1-\varepsilon}\rceil+2\sum_{t=1}^k m_t}\right)\right)\right]\\&\times\Big(1+\mathcal{O}\big(n^{-\delta(1-\xi)(1-\varepsilon)}\big)\Big)+\mathcal{O}\bigg(\sum_{\ell=1}^k\frac{1}{n}\prod_{\substack{t\in[k]\\t\neq\ell}}\left(\mathbb{E}\left[\frac{\widetilde{a}_t^{m_t}}{(1+\widetilde{a}_t)^{m_t+1}}\right]-\frac{4}{n}\right)\bigg),\end{aligned}$$

where we use, as in the upper bound, that $((1-\zeta_n)/(1+2\zeta_n))^{\sum_{t=1}^k m_t}=1-o(n^{-\delta(1-\xi)(1-\varepsilon)})$. If we then consider the summation over indices $n^{1-\varepsilon}\le j_1\neq\ldots\neq j_k$ rather than $n^{1-\varepsilon}\le j_1<\ldots<j_k$ we obtain, as in the upper bound,

$$\begin{aligned}&\prod_{t=1}^k\mathbb{E}\left[\frac{\widetilde{a}_t^{m_t}}{(1+\widetilde{a}_t)^{m_t+1}}\mathbb{P}_W\left(\widetilde{Y}_t<\log\left(\frac{n}{\lceil n^{1-\varepsilon}\rceil+2\sum_{t=1}^k m_t}\right)\right)\right](1-o\big(n^{-\delta(1-\xi)(1-\varepsilon)}\big))\\&\quad-\mathcal{O}\bigg(\sum_{\pi\in P_k}\sum_{\ell=1}^k\frac{1}{n}\prod_{\substack{t\in[k]\\t\neq\ell}}\left(\mathbb{E}\left[\frac{\widetilde{a}_{\pi(t)}^{m_{\pi(t)}}}{(1+\widetilde{a}_{\pi(t)})^{m_{\pi(t)}+1}}\right]-\frac{4}{n}\right)\bigg).\end{aligned}$$

With a similar reasoning as in (5.27) and using that $m_t<c\log n$ for all $t\in[k]$, we can bound the expected value from below for large n as

$$\begin{aligned}&\prod_{t=1}^k\mathbb{E}\bigg[\frac{\mathbb{E}[W]}{\mathbb{E}[W]+W_t}\Big(\frac{W_t}{\mathbb{E}[W]+W_t}\Big)^{m_t}\mathbb{P}_W\Big(\widetilde{Y}_t<\varepsilon(1-\xi)\log n\Big)\bigg](1-o\big(n^{-\delta(1-\xi)(1-\varepsilon)}\big))\\&\quad+\mathcal{O}\bigg(\sum_{\pi\in P_k}\sum_{\ell=1}^k\frac{1}{n}\prod_{\substack{t\in[k]\\t\neq\ell}}\left(\mathbb{E}\left[\frac{\mathbb{E}[W]}{\mathbb{E}[W]+W}\Big(\frac{W}{\mathbb{E}[W]+W}\Big)^{m_{\pi(t)}}\right]-\frac{4}{n}\right)\bigg),\end{aligned}\tag{5.34}$$

for any $\xi\in(0,1)$. Unlike in the upper bound, we cannot trivially omit the conditional probability. Rather, it remains to show that it can be bounded from below by an indicator, at the cost of an additional error term. Since, conditionally on W_t, $\widetilde{Y}_t\sim\Gamma(m_t+1,1+\widetilde{a}_t)$, it follows that (again conditionally on W_t) $(1+\widetilde{a}_t)\widetilde{Y}_t\overset{d}{=}X_t$, where $X_t\sim\Gamma(m_t+1,1)$. We can thus write

$$\mathbb{P}_W\big(\widetilde{Y}_t\le\varepsilon(1-\xi)\log n\big)=\mathbb{P}_W(X_t\le\varepsilon(1-\xi)(1+\widetilde{a}_t)\log n)\ge\mathbb{P}(X_t\le\varepsilon(1-\xi)\log n)\,,$$

almost surely, where the final lower bound is obtained by bounding $\widetilde{a}_t$ from below by zero. As the event on the right-hand side no longer depends on the vertex-weights, we can also omit the conditional probability. In the case that $m_t=o(\log n)$ for all $t\in[k]$, by using the Chernoff inequality we then conclude that for n sufficiently large, almost surely,

$$\mathbb{P}(X_t\le\varepsilon(1-\xi)\log n)\ge 1-n^{-\varepsilon(1-\xi)/2}2^{m_t+1}=1-n^{-\varepsilon(1-\xi)(1+o(1))/2}.\tag{5.35}$$

Using this in (5.34), we thus arrive at the lower bound

$$\begin{aligned}&\prod_{t=1}^k\mathbb{E}\left[\frac{\mathbb{E}[W]}{\mathbb{E}[W]+W_t}\Big(\frac{W_t}{\mathbb{E}[W]+W_t}\Big)^{m_t}\right](1-o\big(n^{-\delta(1-\xi)(1-\varepsilon)\wedge\varepsilon(1-\xi)/4}\big))\\&\quad+\mathcal{O}\bigg(\sum_{\pi\in P_k}\sum_{\ell=1}^k\frac{1}{n}\prod_{\substack{t\in[k]\\t\neq\ell}}\left(\mathbb{E}\left[\frac{\mathbb{E}[W]}{\mathbb{E}[W]+W}\Big(\frac{W}{\mathbb{E}[W]+W}\Big)^{m_{\pi(t)}}\right]-\frac{4}{n}\right)\bigg).\end{aligned}\tag{5.36}$$

Then, via the same reasoning as in (5.29), the big $\mathcal{O}$ term can be included in the error term when the m_t are $o(\log n)$. We thus establish (5.1) when the m_t are $o(\log n)$ by combining (5.30), (5.36) and the above and by setting $\beta:=\beta_1\wedge(\varepsilon(1-\xi)/4)\wedge\delta(1-\varepsilon)(1-\xi)$.

It thus remains to prove (5.1) when (some of) the m_t grow faster. The upper bound in (5.30) suffices in this case as well. For the lower bound, (5.34) still holds and we can deal with all m_t such that $m_t = o(\log n)$ as in (5.35), so that we can assume without loss of generality that $m_t \geq \eta \log n$ for some $\eta \in (0, c)$ for all $t \in [k]$. Or, more specifically, we assume that $m_t \sim c_t \log n$ for some $c_t \in (0, c)$ for all $t \in [k]$ (so that taking $\eta < \min_{t\in[k]} c_t$ yields the same result). Then, we bound

$$\mathbb{P}_W(X_t \leq \varepsilon(1-\xi)(1+\widetilde{a}_t)\log n) \geq \mathbb{P}\Big(X_t \leq \frac{c_t\varepsilon}{1-\mu}\log n\Big)\mathbb{1}_{\{1+\widetilde{a}_t > c_t/((1-\mu)(1-\xi))\}}, \tag{5.37}$$

where $\mu \in (0,1)$ is a small constant and where we again can switch to the non-conditional probability measure $\mathbb{P}$ in the last step as there are no vertex-weights involved in the probability. Now, by choosing $\varepsilon \in (1-\mu, 1)$ we can obtain a bound on the rate of convergence to one by applying a standard large deviation bound. Let $(V_i)_{i\in\mathbb{N}}$ be i.i.d. exponential rate 1 random variables and let $I(a) := a - 1 - \log(a)$ be their rate function. Then, as we can think of X_t as the sum of $V_1, \ldots, V_{m_t+1}$,

$$\begin{aligned}\mathbb{P}\Big(X_t \geq \frac{c_t\varepsilon}{1-\mu}\log n\Big) &= \mathbb{P}\bigg(\sum_{i=1}^{m_t+1} V_i \geq (m_t+1)\frac{c_t\varepsilon\log n}{(1-\mu)(m_t+1)}\bigg)\\ &\leq \exp\Big\{-(m_t+1)I\Big(\frac{c_t\varepsilon\log n}{(1-\mu)(m_t+1)}\Big)\Big\}.\end{aligned}$$

In the first step, we express the upper bound within the probability in terms of the mean of the sum of random variables, which equals m_t+1. We then use the large deviations bound in the second step, which we can do as the argument of I is strictly greater than 1 when n is sufficiently large (as $m_t + 1 \sim c_t \log n$) and $\varepsilon \in (1-\mu, 1)$ is sufficiently close to one. Since $I(c_t\varepsilon\log n/((1-\mu)(m_t+1))) = (\varepsilon/(1-\mu) - 1 - \log(\varepsilon/(1-\mu)))(1+o(1))$, we thus arrive at

$$\mathbb{P}\Big(X_t \geq \frac{c_t\varepsilon}{1-\mu}\log n\Big) \leq \mathrm{e}^{-c_t\log n(\varepsilon/(1-\mu)-1-\log(\varepsilon/(1-\mu)))(1+o(1))} = n^{-c_{t,\mu,\varepsilon}(1+o(1))},$$

where $c_{t,\mu,\varepsilon} := c_t(\varepsilon/(1-\mu) - 1 - \log(\varepsilon/(1-\mu))) > 0$ as $\varepsilon > 1-\mu$. When combining this with (5.37) in (5.34) and recalling that $\widetilde{a}_t = W_t/(\mathbb{E}[W](1-\zeta_n)) \geq W_t/\mathbb{E}[W]$, we obtain the lower bound

$$\begin{aligned}&\frac{1}{(n)_k}\sum_{n^{1-\varepsilon}\leq j_1\neq\ldots\neq j_k\leq n}\mathbb{E}\left[\mathbb{P}_W(\mathcal{Z}_n(j_\ell) = m_\ell, \ell\in[k])\right]\\ &\geq \prod_{t=1}^k \mathbb{E}\Big[\frac{\mathbb{E}[W]}{\mathbb{E}[W]+W}\Big(\frac{W}{\mathbb{E}[W]+W}\Big)^{m_t}\mathbb{1}_{\{W>(c_t/((1-\mu)(1-\xi))-1)\mathbb{E}[W]\}}\Big](1+o(n^{-\beta_2}))\\ &\quad+\mathcal{O}\bigg(\sum_{\pi\in P_k}\sum_{\ell=1}^k\frac{1}{n}\prod_{\substack{t\in[k]\\ t\neq\ell}}\Big(\mathbb{E}\Big[\frac{\mathbb{E}[W]}{\mathbb{E}[W]+W}\Big(\frac{W}{\mathbb{E}[W]+W}\Big)^{m_{\pi(t)}}\Big]-\frac{4}{n}\Big)\bigg),\end{aligned}$$

for some $\beta_2 < \min_{t\in[k]} c_{t,\mu_\varepsilon} \wedge \delta(1-\varepsilon)(1-\xi) \wedge (\varepsilon(1-\xi)/4)$. We can then replace c_t in the indicator by $\widetilde{c} := \max_{t\in[k]} c_t$ to obtain a further lower bound. The indicator in the expected value then is non-zero with positive probability when $(\widetilde{c}/((1-\mu)(1-\xi))-1)\mathbb{E}[W] < 1$, or, equivalently, $\widetilde{c} < (1-\mu)(1-\xi)\theta/(\theta-1)$. We can therefore take any $c_1, \ldots, c_k < \theta/(\theta-1)$ and set μ and ξ small enough for this inequality to be satisfied.

We now argue that at the cost of an additional error term, we can omit the indicator in the expected value. We recall that we assumed that $\eta\log n \leq m_\ell \leq c\log n$ for all $\ell \in [k]$ for some $\eta \in (0, c)$. Recall $f_k(\theta, W)$ from (5.6) and note that $p_k = \mathbb{E}[f_k(\theta, W)]$. Let us also set $a := \widetilde{c}/((1-\mu)(1-\xi)) - 1)\mathbb{E}[W]$ and note that by the above, $a \in (0,1)$. Since $m_\ell > 1/\mathbb{E}[W]$ (this implies $f_{m_\ell}(\theta, x)$ is increasing in x), for n large,

$$\mathbb{E}\left[f_{m_\ell}(\theta, W)\right] = \mathbb{E}\left[f_{m_\ell}(\theta, W)(\mathbb{1}_{\{W>a\}} + \mathbb{1}_{\{W\leq a\}})\right] \leq \mathbb{E}\left[f_{m_\ell}(\theta, W)\mathbb{1}_{\{W>a\}}\right] + f_{m_\ell}(\theta, a).$$

We then bound

$$f_{m_\ell}(\theta,a)\leq\Big(\frac{a}{\theta-1+a}\Big)^{m_\ell}\leq(\theta+\xi)^{-m_\ell},$$

for some sufficiently small $\xi>0$, as $a\in(0,1)$ and the fraction is strictly increasing in a. Since it follows from the proof of Lemma 5.5 that $\mathbb{E}\left[f_k(\theta,W)\right](\theta+\xi)^k$ diverges exponentially fast as k tends to infinity when W is bounded by one, it then follows that

$$\mathbb{E}\left[f_{m_\ell}(\theta,W)\mathbb{1}_{\{W>a\}}\right]\geq\mathbb{E}\left[f_{m_\ell}(\theta,W)\right]-(\theta+\xi)^{-m_\ell}=\mathbb{E}\left[f_{m_\ell}(\theta,W)\right](1-o((1+\gamma)^{-m_\ell})),$$

for some small $\gamma>0$. As $m_\ell\geq\eta\log n$ for all $\ell\in[k]$ it then follows that $1-o((1+\gamma)^{-m_\ell})=1-o(n^{-\widetilde\beta_2})$ for some $\widetilde\beta_2>0$. We thus obtain the lower bound

$$\begin{aligned}&\prod_{t=1}^k\mathbb{E}\left[\frac{\mathbb{E}\left[W\right]}{\mathbb{E}\left[W\right]+W}\Big(\frac{W}{\mathbb{E}\left[W\right]+W}\Big)^{m_t}\right](1+o(n^{-\beta_2\wedge\widetilde\beta_2}))\\&\quad+\mathcal{O}\Big(\sum_{\pi\in P_k}\sum_{\ell=1}^k\frac1n\prod_{\substack{t\in[k]\\t\neq\ell}}\Big(\mathbb{E}\left[\frac{\mathbb{E}\left[W\right]}{\mathbb{E}\left[W\right]+W}\Big(\frac{W}{\mathbb{E}\left[W\right]+W}\Big)^{m_{\pi(t)}}\right]-\frac4n\Big)\Big).\end{aligned}$$

Finally, with a similar reasoning as in (5.29), we can either include the big $\mathcal{O}$ term in the error term $1+o(n^{-\beta_2\wedge\widetilde\beta_2})$ in case (1) or omit it completely in case (2). Again, a combination of m_t which satisfy either case (1) or case (2) can be dealt with by combining both approaches. The proof of (5.1) is then concluded by combining the upper bound in (5.30) and the lower bound above and setting $\beta:=\beta_1\wedge\beta_2\wedge\widetilde\beta_2\wedge(\varepsilon(1-\xi)/4)\wedge\delta(1-\varepsilon)(1-\xi)$. □

5.3. **Proof of Lemma 5.11.**

Proof of Lemma 5.11. We aim to bound

$$\frac{1}{(n)_k}\sum_{\mathbf{j}\in I_n(\varepsilon)}\mathbb{E}\left[\mathbb{P}_W(\mathcal{Z}_n(j_\ell)=m_\ell,\ell\in[k])\right],\tag{5.38}$$

where we recall that $I_n(\varepsilon):=\{\mathbf{j}=(j_1,\dots,j_k):1\leq j_1\neq\dots\neq j_k\leq n,\exists i\in[k]\ j_i<n^{1-\varepsilon}\}$. We first assume that $m_\ell=c_\ell\log n(1+o(1))$ for some $c_\ell\in[0,1/\log\theta)$ for all $\ell\in[k]$, where $c_\ell=0$ denotes that $m_\ell=o(\log n)$. We define

$$I_n(\varepsilon,i):=\{\mathbf{j}\in I_n(\varepsilon):|\{\ell\in[k]:j_\ell<n^{1-\varepsilon}\}|=i\},\qquad i\in[k],$$

that is, $I_n(\varepsilon,i)$ denotes the set of indices $\mathbf{j}=(j_1,\dots,j_k)$ such that exactly i of the indices are smaller than $n^{1-\varepsilon}$, and note that $I_n(\varepsilon)=I_n(\varepsilon,1)\cup\dots\cup I_n(\varepsilon,k)$. We then write

$$\frac{1}{(n)_k}\sum_{\mathbf{j}\in I_n(\varepsilon)}\mathbb{E}\left[\mathbb{P}(\mathcal{Z}_n(j_\ell)=m_\ell,\ell\in[k])\right]=\sum_{i=1}^k\frac{1}{(n)_k}\sum_{\mathbf{j}\in I_n(\varepsilon,i)}\mathbb{E}\left[\mathbb{P}_W(\mathcal{Z}_n(j_\ell)=m_\ell,\ell\in[k])\right],$$

and bound the probability on the right-hand side from above by omitting all events $\{\mathcal{Z}_n(j_\ell)=m_\ell\}$ whenever $j_\ell<n^{1-\varepsilon}$. This leaves us with

$$\sum_{i=1}^{k-1}\frac{1}{(n)_k}n^{i(1-\varepsilon)}\sum_{\substack{S\subseteq[k]\\|S|=k-i}}\sum_{\substack{n^{1-\varepsilon}\leq j_\ell\leq n\\\ell\in S}}^{*}\mathbb{E}\left[\mathbb{P}_W(\mathcal{Z}_n(j_\ell)=m_\ell,\ell\in S)\right]+\frac{n^{k(1-\varepsilon)}}{(n)_k},\tag{5.39}$$

where we recall that the $*$ on the second sum symbol denotes that we only consider distinct values of $j_\ell,\ell\in S$. We isolated the case $i=k$ here as in this case no indices are larger than $n^{1-\varepsilon}$ and we hence bound the probability from above by one, whereas $i=k$ would yield a contribution of zero in the triple sum. The inner sum can then be dealt with in the same manner as in the derivation of the upper bound in (5.30), to yield an upper bound

$$\sum_{i=1}^{k-1} n^{-i\varepsilon} \sum_{\substack{S\subseteq[k]\\|S|=k-i}} \prod_{\ell\in S} \mathbb{E}\left[\frac{\mathbb{E}[W]}{\mathbb{E}[W]+W}\Big(\frac{W}{\mathbb{E}[W]+W}\Big)^{m_\ell}\right](1+o(n^{-\beta}))+2n^{-k\varepsilon},$$

for some $\beta > 0$. It thus remains to show that for any $m_\ell = c_\ell \log n(1+o(1))$ with $c_\ell \in [0, 1/\log\theta)$ we can take ε sufficiently close to one and a small $\eta > 0$, such that

$$n^{-\varepsilon} = o\Big(\mathbb{E}\left[\frac{\mathbb{E}[W]}{\mathbb{E}[W]+W}\Big(\frac{W}{\mathbb{E}[W]+W}\Big)^{m_\ell}\right] n^{-\eta}\Big).$$

By Lemma 5.5, we have for any $\xi > 0$ and n sufficiently large, that

$$\mathbb{E}\left[\frac{\mathbb{E}[W]}{\mathbb{E}[W]+W}\Big(\frac{W}{\mathbb{E}[W]+W}\Big)^{m_\ell}\right] \geq (\theta+\xi)^{-m_\ell} = n^{-c_\ell \log(\theta+\xi)(1+o(1))},$$

and $n^{-\varepsilon} = o(n^{-\eta - c_\ell \log(\theta+\xi)(1+o(1))})$ when we choose η and ξ sufficiently small and ε sufficiently close to 1, since $c_\ell \log\theta < 1$ for any $\ell \in [k]$. As a result,

$$\frac{1}{(n)_k}\sum_{\mathbf{j}\in I_n(\varepsilon)} \mathbb{E}\left[\mathbb{P}_W(\mathcal{Z}_n(j_\ell)=m_\ell, \ell\in[k])\right] = o\Big(\prod_{\ell=1}^k \mathbb{E}\left[\frac{\mathbb{E}[W]}{\mathbb{E}[W]+W}\Big(\frac{W}{\mathbb{E}[W]+W}\Big)^{m_\ell}\right] n^{-\eta}\Big).$$

We now assume that $m_\ell = c_\ell \log n(1+o(1))$ with $c_\ell \in [1/\log\theta, \theta/(\theta-1))$ for all $\ell \in [k]$. In this case, the crude bound used above no longer suffices. Now, the aim is to use a similar approach as in the start of the proof of [13, Theorem 2.9, Bounded case] and combine this with the assumption that $W \geq w^* > 0$ almost surely for some $w^* \in (0,1)$. First, we consider the set of indices $I_n(\varepsilon, k)$. To make use of the negative quadrant dependence of the degrees $(\mathcal{Z}_n(i))_{i\in[n]}$ (see Remark 5.2 and [13, Lemma 7.1]), we create an upper bound by considering the event $\{\mathcal{Z}_n(j_\ell) \geq m_\ell, \ell \in [k]\}$. Then, using the tail distribution and the negative quadrant dependency of the degrees under the conditional probability measure $\mathbb{P}_W$ yields

$$\frac{1}{(n)_k}\sum_{\mathbf{j}\in I_n(\varepsilon,k)} \mathbb{P}(\mathcal{Z}_n(j_\ell)=m_\ell, \ell\in[k]) \leq \frac{1}{(n)_k}\sum_{1\leq j_1\neq\ldots\neq j_k<n^{1-\varepsilon}} \mathbb{E}\left[\prod_{\ell=1}^k \mathbb{P}_W(\mathcal{Z}_n(j_\ell)\geq m_\ell)\right].$$

We then also allow the indices $j_1, \ldots, j_k$ to take any value between 1 and $n^{1-\varepsilon}$, to obtain the upper bound

$$\frac{1}{(n)_k}\mathbb{E}\left[\prod_{\ell=1}^k \Big(\sum_{i<n^{1-\varepsilon}} \mathbb{P}_W(\mathcal{Z}_n(i)\geq m_\ell)\Big)\right].$$

As in the proof of [13, Theorem 2.9, Bounded case], we apply a Chernoff bound to the conditional probability measure $\mathbb{P}_W$ to obtain

$$\frac{1}{(n)_k}\mathbb{E}\left[\prod_{\ell=1}^k \Big(\sum_{i<n^{1-\varepsilon}} \mathbb{P}_W(\mathcal{Z}_n(i)\geq m_\ell)\Big)\right] \leq \frac{1}{(n)_k}\mathbb{E}\left[\prod_{\ell=1}^k \Big(\sum_{i<n^{1-\varepsilon}} \exp\{m_\ell(1-u_{i,\ell}+\log u_{i,\ell})\}\Big)\right],$$

where $u_{i,\ell} := W_i(H_n - H_i)/m_\ell$ and $H_n := \sum_{j=1}^{n-1} 1/S_j$. We then introduce the constants $\delta \in (0, 1/2), C > kc_\theta^{-1}\theta\log(\theta)/(\theta-1)$ (with $c_\theta := 1/(2\theta^2)$), the sequence $\zeta_n' := (C\log n)^{-\delta/(1-2\delta)}/\mathbb{E}[W]$, $n \in \mathbb{N}$, and introduce the event

$$E_n' := \Big\{\sum_{\ell=1}^j W_\ell \geq j\mathbb{E}[W](1-\zeta_n'),\ \forall\ (C\log n)^{1/(1-2\delta)} \leq j \leq n\Big\}.$$

The event E_n' is similar to the event E_n introduced in (5.13), but considers a larger range of indices j. The particular choice of the lower bound on the indices j follows from the fact that we want as much control over the partial sums of the vertex-weights as possible, but need to ensure that $\mathbb{P}((E_n')^c)$ decays sufficiently fast, which we can achieve via this choice.

We can use the event E_n' in the expected value to arrive at the upper bound

$$\frac{1}{(n)_k}\mathbb{E}\left[\prod_{\ell=1}^k\Big(\sum_{i<n^{1-\varepsilon}}\exp\{m_\ell(1-u_{i,\ell}+\log u_{i,\ell})\}\Big)\mathbb{1}_{E_n'}\right]+\mathbb{P}\big((E_n')^c\big). \tag{5.40}$$

We defer the proof that $\mathbb{P}((E_n')^c)$ decays sufficiently fast for now and focus on the first term. We bound $u_{i,\ell}$ from above by

$$u_{i,\ell}\leq \frac{H_n}{m_\ell}=\frac{1}{m_\ell}\Big[\sum_{j<(C\log n)^{1/(1-2\delta)}}\frac{1}{S_j}+\sum_{j=\lceil (C\log n)^{1/(1-2\delta)}\rceil}^{n}\frac{1}{S_j}\Big],$$

and using $W_i\geq w^*$ almost surely for all $i\in\mathbb{N}$ as well as the bound in the event E_n' then yields

$$u_{i,\ell}\leq \frac{1}{m_\ell}\Big[\frac{1-2\delta}{w^*}\log(C\log n)+\frac{1}{\mathbb{E}[W]}\log\Big(\frac{n}{\lceil (C\log n)^{1/(1-2\delta)}\rceil}\Big)\Big](1+o(1))=\frac{1+o(1)}{c_\ell\mathbb{E}[W]}.$$

Since $\mathbb{E}[W]=\theta-1$ and $c_\ell\geq 1/\log\theta$, it follows that $1/(c_\ell\mathbb{E}[W])\leq \log\theta/(\theta-1)<1$ for all $\theta\in(1,2]$. Since $x\mapsto 1-x+\log x$ is increasing for $x\in(0,1)$ we can thus use this upper bound in the first term of (5.40) to bound it from above by

$$\begin{aligned}
&\frac{1}{(n)_k}\prod_{\ell=1}^k\Big(\sum_{i<n^{1-\varepsilon}}\exp\Big\{c_\ell\log n\Big(1-\frac{1}{c_\ell\mathbb{E}[W]}+\log\Big(\frac{1}{c_\ell\mathbb{E}[W]}\Big)\Big)(1+o(1))\Big\}\Big)\\
&\quad\leq \frac{1}{(n)_k}\prod_{\ell=1}^k\exp\Big\{\log n\Big((1-\varepsilon)+c_\ell\Big(1-\frac{1}{c_\ell\mathbb{E}[W]}+\log\Big(\frac{1}{c_\ell\mathbb{E}[W]}\Big)\Big)\Big)(1+o(1))\Big\}\\
&\quad\leq \exp\Big\{\log n(1+o(1))\sum_{\ell=1}^k\Big(-\varepsilon+c_\ell\Big(1-\frac{1}{c_\ell\mathbb{E}[W]}+\log\Big(\frac{1}{c_\ell\mathbb{E}[W]}\Big)\Big)\Big)\Big\}.
\end{aligned}$$

We then require that

$$\begin{aligned}
&\exp\Big\{\log n(1+o(1))\sum_{\ell=1}^k\Big(-\varepsilon+c_\ell\Big(1-\frac{1}{c_\ell\mathbb{E}[W]}+\log\Big(\frac{1}{c_\ell\mathbb{E}[W]}\Big)\Big)\Big)\Big\}\\
&\qquad=o\Big(n^{-\eta}\prod_{\ell=1}^k p_{m_\ell}\Big),
\end{aligned}\tag{5.41}$$

for some $\eta>0$. As, by Lemma 5.5, $p_{m_\ell}\geq(\theta+\xi)^{-m_\ell}=\exp\{-\log n(1+o(1))c_\ell\log(\theta+\xi)\}$, it suffices to show that

$$\sum_{\ell=1}^k\Big(-\varepsilon+c_\ell\Big(1-\frac{1}{c_\ell\mathbb{E}[W]}+\log\Big(\frac{1}{c_\ell\mathbb{E}[W]}\Big)\Big)\Big)<-\sum_{\ell=1}^k c_\ell\log(\theta+\xi), \tag{5.42}$$

when ξ is sufficiently small and ε sufficiently close to one. We show that this strict inequality can be achieved for each term individually, by arguing that we can choose $\varepsilon\in(0,1)$ such that

$$\varepsilon>c_\ell\Big(1-\frac{1}{c_\ell(\theta-1)}+\log\Big(\frac{\theta+\xi}{c_\ell(\theta-1)}\Big)\Big),\qquad \ell\in[k],$$

where we note that we have written $\mathbb{E}[W]$ as $\theta-1$. The right-hand side is increasing in c_ℓ when $c_\ell\in[1/\log\theta,\theta/(\theta-1))$, so that all k inequalities are satisfied when we solve

$$\varepsilon>\widetilde{c}\Big(1-\frac{1}{\widetilde{c}(\theta-1)}+\log\Big(\frac{\theta+\xi}{\widetilde{c}(\theta-1)}\Big)\Big),$$

with $\widetilde{c} := \max_{\ell\in[k]} c_\ell$. We now show that the right-hand side strictly smaller than one when ξ is sufficiently small. We write

$$\begin{aligned}\widetilde{c}\Big(1-\frac{1}{\widetilde{c}(\theta-1)}+\log\Big(\frac{\theta+\xi}{\widetilde{c}(\theta-1)}\Big)\Big) &= \widetilde{c}\Big(1-\frac{1}{\widetilde{c}(\theta-1)}+\log\Big(\frac{\theta}{\widetilde{c}(\theta-1)}\Big)\Big)+\widetilde{c}\log\Big(1+\frac{\xi}{\theta}\Big)\\ &\le \widetilde{c}\Big(1-\frac{1}{\widetilde{c}(\theta-1)}+\log\Big(\frac{\theta}{\widetilde{c}(\theta-1)}\Big)\Big)+\frac{\xi}{\theta-1},\end{aligned}$$

where the final upper bound follows from the fact that $\log(1+x)\le x$ for $x>-1$ and $\widetilde{c}<\theta/(\theta-1)$. We denote the first term on the right-hand side by $\kappa=\kappa(\widetilde{c},\theta)$. As κ is increasing in $\widetilde{c}$ when $\widetilde{c}\in[1/\log\theta,\theta/(\theta-1))$ (which is the case when $c_\ell\in[1/\log\theta,\theta/(\theta-1))$ for all $\ell\in[k]$), we have $\kappa<1$, as $\widetilde{c}<\theta/(\theta-1)$. Thus, setting $\xi<(1-\kappa)(\theta-1)/2$ we achieve the desired result. Now, taking $\varepsilon\in(\kappa+\xi/(\theta-1),1)$, we arrive at (5.41) for some small $\eta>0$. It thus follows that

$$\frac{1}{(n)_k}\sum_{\mathbf{j}\in I_n(\varepsilon,k)}\mathbb{E}\left[\mathbb{P}_W(\mathcal{Z}_n(j_\ell)=m_\ell,\ell\in[k])\right]=o\Big(n^{-\eta}\prod_{\ell=1}^k p_{m_\ell}\Big), \tag{5.43}$$

for some small $\eta>0$.

We now consider the remaining sets $I_n(\varepsilon,1),\dots,I_n(\varepsilon,k-1)$ and aim to bound

$$\frac{1}{(n)_k}\sum_{i=1}^{k-1}\sum_{\mathbf{j}\in I_n(\varepsilon,i)}\mathbb{P}(\mathcal{Z}_n(j_\ell)\ge m_\ell,\ell\in[k])\,.$$

Again, using the negative quadrant dependence and introducing the events E_n' and E_n (recall E_n from (5.13)) yields the upper bound

$$\frac{1}{(n)_k}\sum_{i=1}^{k-1}\sum_{\mathbf{j}\in I_n(\varepsilon,i)}\mathbb{E}\left[\mathbb{1}_{E_n'\cap E_n}\prod_{\ell=1}^k\mathbb{P}_W(\mathcal{Z}_n(j_\ell)\ge m_\ell)\right]+\mathbb{P}\big((E_n')^c\big)+\mathbb{P}(E_n^c)\,.$$

The aim is to treat the probabilities of indices which are at most $n^{1-\varepsilon}$ in the same way as when dealing with the indices in $I_n(\varepsilon,k)$ to reach a bound as in (5.41), for which we use the event E_n'. For the indices which are larger than $n^{1-\varepsilon}$ such an upper bound will not suffice. Instead, we aim to bound $\mathbb{P}_W(\mathcal{Z}_n(j_\ell)\ge m_\ell)$ when $n^{1-\varepsilon}\le j_\ell\le n$ in a similar way as we bounded $\mathbb{P}_W(\mathcal{Z}_n(j_\ell)=m_\ell)$ from above in the proof of Lemma 5.10, for which we use E_n.

First, we split the summation over $I_n(\varepsilon,i)$ over all possible configurations of indices with are at most and at least $n^{1-\varepsilon}$, similar to (5.39). That is,

$$\begin{aligned}&\frac{1}{(n)_k}\sum_{i=1}^{k-1}\sum_{\mathbf{j}\in I_n(\varepsilon,i)}\mathbb{E}\left[\mathbb{1}_{E_n'}\mathbb{1}_{E_n}\prod_{\ell=1}^k\mathbb{P}_W(\mathcal{Z}_n(j_\ell)\ge m_\ell)\right]\\ &=\frac{1}{(n)_k}\sum_{i=1}^{k-1}\sum_{\substack{S\subseteq[k]\\|S|=i}}\sideset{}{^*}\sum_{\substack{1\le j_\ell<n^{1-\varepsilon}\\ \ell\in S}}\sideset{}{^*}\sum_{\substack{n^{1-\varepsilon}\le j_\ell\le n\\ \ell\in[k]\setminus S}}\mathbb{E}\Bigg[\mathbb{1}_{E_n'\cap E_n}\prod_{\ell\in S}\mathbb{P}_W(\mathcal{Z}_n(j_\ell)\ge m_\ell)\prod_{\ell\in[k]\setminus S}\mathbb{P}_W(\mathcal{Z}_n(j_\ell)\ge m_\ell)\Bigg].\end{aligned}$$

Using the event E_n', we can follow similar steps as above to bound the sum over the indices j_ℓ and the product of probabilities $\mathbb{P}_W(\mathcal{Z}_n(j_\ell)\ge m_\ell)$ for $\ell\in S$ from above by the deterministic upper bound

$$\exp\Big\{\log n(1+o(1))\sum_{\ell\in S}\Big(-\varepsilon+c_\ell\Big(1-\frac{1}{c_\ell\mathbb{E}[W]}+\log\Big(\frac{1}{c_\ell\mathbb{E}[W]}\Big)\Big)\Big)\Big\}=:n^{C(S)(1+o(1))},$$

which yields

$$\sum_{i=1}^{k-1}\frac{n^i}{(n)_k}\sum_{\substack{S\subseteq[k]\\|S|=i}}n^{C(S)(1+o(1))}\sideset{}{^*}\sum_{\substack{n^{1-\varepsilon}\le j_\ell\le n\\ \ell\in[k]\setminus S}}\mathbb{E}\left[\mathbb{1}_{E_n}\prod_{\ell\in[k]\setminus S}\mathbb{P}_W(\mathcal{Z}_n(j_\ell)\ge m_\ell)\right]. \tag{5.44}$$

We now proceed to bound each individual probability $\mathbb{P}_W(\mathcal{Z}_n(j_\ell) \geq m_\ell)$ when $\ell \in [k]\backslash S$. This follows a similar approach to the upper bound of $\mathbb{P}_W(\mathcal{Z}_n(j_\ell) = m_\ell)$ in the proof of Lemma 5.10, with a couple of modifications. Introducing indices $j_\ell < i_1 < \ldots < i_{m_\ell} \leq n$, which denote the steps at which vertex j_ℓ increases its degree, we can write

$$\mathbb{P}_W(\mathcal{Z}_n(j_\ell) \geq m_\ell) = \sum_{j_\ell < i_1 < \ldots < i_{m_\ell} \leq n} \prod_{t=1}^{m_\ell} \frac{W_{j_\ell}}{\sum_{r=1}^{i_t - 1} W_r} \prod_{\substack{s=j+1 \\ s \neq i_t, t \in [m_\ell]}}^{i_{m_\ell}-1} \Big(1 - \frac{W_{j_\ell}}{\sum_{r=1}^{s-1} W_r}\Big).$$

The second product, in comparison to dealing with the event $\{\mathcal{Z}_n(j_\ell) = m_\ell\}$, goes up to $i_{m_\ell} - 1$ instead of n. This is due to the fact that we now only need to control the connections vertex j does and does not make up to its m_ℓ^{th} connection. Using the same idea as in (5.17) and using the event E_n, we obtain the upper bound

$$\begin{aligned}
&\sum_{j_\ell < i_1 < \ldots < i_{m_\ell} \leq n} \prod_{t=1}^{m_\ell} \frac{W_{j_\ell}}{(i_t - 1)\mathbb{E}[W](1-\zeta_n) - 1} \prod_{s=j+1}^{i_{m_\ell}-1} \Big(1 - \frac{W_{j_\ell}}{s\mathbb{E}[W](1+\zeta_n)}\Big) \\
&\qquad \leq \sum_{j_\ell < i_1 < \ldots < i_{m_\ell} \leq n} \prod_{t=1}^{m_\ell} \frac{W_{j_\ell}}{i_t \mathbb{E}[W](1-2\zeta_n)} \prod_{s=j+1}^{i_{m_\ell}-1} \Big(1 - \frac{W_{j_\ell}}{s\mathbb{E}[W](1+\zeta_n)}\Big).
\end{aligned}$$

The last step follows from the fact that $(i_t - 1)(1-\zeta_n)\mathbb{E}[W] - 1 \geq i_t(1-2\zeta_n)\mathbb{E}[W]$ for n sufficiently large. Using this in the expected value of (5.44) yields

$$\sideset{}{^*}\sum_{\substack{n^{1-\varepsilon} \leq j_\ell \leq n \\ \ell \in [k]\backslash S}} \mathbb{E}\Bigg[\prod_{\ell \in [k]\backslash S} \Bigg(\sum_{j_\ell < i_1 < \ldots < i_{m_\ell} \leq n} \prod_{t=1}^{m_\ell} \frac{W_{j_\ell}}{i_t \mathbb{E}[W](1-2\zeta_n)} \prod_{s=j+1}^{i_{m_\ell}-1} \Big(1 - \frac{W_{j_\ell}}{s\mathbb{E}[W](1+\zeta_n)}\Big)\Bigg)\Bigg].$$

We can now relabel the vertex-weights W_{j_ℓ} by W_ℓ, $\ell \in [k]\backslash S$. This does not change the expected value and is possible since the indices $j_\ell, \ell \in [k]\backslash S$ are distinct. Directly after this, we omit the requirement that the indices j_ℓ are distinct, which is now of no consequence as the weights have been relabelled already. We hence arrive at the upper bound

$$\prod_{\ell \in [k]\backslash S} \mathbb{E}\Bigg[\sum_{n^{1-\varepsilon} \leq j_\ell \leq n} \sum_{j_\ell < i_1 < \ldots < i_{m_\ell} \leq n} \prod_{t=1}^{m_\ell} \frac{W_\ell}{i_t \mathbb{E}[W](1-2\zeta_n)} \prod_{s=j+1}^{i_{m_\ell}-1} \Big(1 - \frac{W_\ell}{s\mathbb{E}[W](1+\zeta_n)}\Big)\Bigg], \tag{5.45}$$

where the product can be taken out of the expected value due to the independence of the vertex-weights $W_1, \ldots W_\ell$. As a result, we can deal with each of expected values individually. Following the same approach as in (5.18) and setting $a_\ell := W_\ell/(\mathbb{E}[W](1+\zeta_n))$, we obtain the upper bound

$$\begin{aligned}
&\mathbb{E}\Bigg[\sum_{n^{1-\varepsilon} \leq j_\ell \leq n} \sum_{j_\ell < i_1 < \ldots < i_{m_\ell} \leq n} \prod_{t=1}^{m_\ell} \frac{W_\ell}{i_t \mathbb{E}[W](1-2\zeta_n)} \Big(\frac{i_{m_\ell}}{j_\ell}\Big)^{-a_\ell}\Bigg]\Big(1 + \mathcal{O}\big(n^{-(1-\varepsilon)}\big)\Big) \\
&= \mathbb{E}\Bigg[a_\ell^{m_\ell} \sum_{n^{1-\varepsilon} \leq j_\ell \leq n} \sum_{j_\ell < i_1 < \ldots < i_{m_\ell} \leq n} \prod_{t=1}^{m_\ell} i_t^{-1} \Big(\frac{i_{m_\ell}}{j_\ell}\Big)^{-a_\ell}\Bigg]\Big(\frac{1+\zeta_n}{1-2\zeta_n}\Big)^{m_\ell}\Big(1 + \mathcal{O}\big(n^{-(1-\varepsilon)}\big)\Big).
\end{aligned} \tag{5.46}$$

We then observe that the summand of the inner sum over the indices $i_1, \ldots, i_{m_\ell}$ is decreasing, so that we can bound it from above almost surely by the multiple integrals

$$\int_{j_\ell}^n \int_{x_1}^n \cdots \int_{x_{m_\ell - 1}}^n \prod_{t=1}^{m_\ell - 1} x_t^{-1} x_{m_\ell}^{-(1+a_\ell)} \,\mathrm{d}x_{m_\ell} \ldots \mathrm{d}x_1 =: I_{m_\ell}.$$

Calculating the value of the innermost integral yields the recursion

$$I_{m_\ell} = \frac{I_{m_\ell - 1}}{a_\ell} - \frac{1}{n^{a_\ell} a_\ell} \int_{j_\ell}^n \int_{x_1}^n \cdots \int_{x_{m_\ell - 2}}^n \prod_{t=1}^{m_\ell - 1} x_t^{-1} \,\mathrm{d}x_{m_\ell - 1} \ldots \mathrm{d}x_1 = \frac{I_{m_\ell - 1}}{a_\ell} - \frac{\log(n/j_\ell)^{m_\ell - 1}}{n^{a_\ell} a_\ell (m_\ell - 1)!},$$

where the last step follows from (5.21). By continuing the recursion we find that

$$I_{m_\ell} = \frac{I_1}{a_\ell^{m_\ell-1}} - n^{-a_\ell}\sum_{s=1}^{m_\ell-1} a_\ell^{m_\ell-s}\frac{\log(n/j_\ell)^{m_\ell-s}}{(m_\ell-s)!} = a_\ell^{-m_\ell} j_\ell^{-a_\ell}\bigg(1-\Big(\frac{n}{j_\ell}\Big)^{-a_\ell}\sum_{s=0}^{m_\ell-1} a_\ell^s\frac{\log(n/j_\ell)^s}{s!}\bigg).$$

Multiplying this with the $a_\ell^{m_\ell} j^{a_\ell}$ in the expected value in (5.46), we arrive at

$$1-\sum_{s=0}^{m_\ell-1}(n/j_\ell)^{-a_\ell}\frac{(a_\ell\log(n/j_\ell))^s}{s!} = \mathbb{P}_W(P(a_\ell)\geq m_\ell)\,,$$

where, conditionally on W_ℓ, $P(a_\ell)\sim \mathrm{Poi}(a_\ell\log(n/j_\ell))$. We now use the following duality between Poisson and gamma random variables. Let $X\sim\Gamma(m_\ell,1)$ be a gamma random variable. We can also interpret X as a sum of m_ℓ rate one exponential random variables. Then, conditionally on W_ℓ, the event $\{P(a_\ell)\geq m_\ell\}$ can be thought of as the event that in a rate one Poisson process at least m_ℓ particles have arrived before time $a_\ell\log(n/j_\ell)$. This is equivalent to the sum of the first m_ℓ inter-arrival times (which are rate one exponentially distributed) being at most $a_\ell\log(n/j_\ell)$. As we mentioned, this sum of m_ℓ rate one exponential random variables is, in law, identical to X, so

$$\mathbb{P}_W(P(a_\ell)\geq m_\ell) = \mathbb{P}_W(X\leq a_\ell\log(n/j_\ell)) = \mathbb{P}_W(X/a_\ell\leq \log(n/j_\ell)) = \mathbb{P}_W(Y\leq\log(n/j_\ell))\,,$$

where, conditionally on W_ℓ, $Y\sim\Gamma(m_\ell,a_\ell)$. Then, by the choice of ζ_n, $((1+\zeta_n)/(1-2\zeta_n))^{m_\ell} = 1+\mathcal{O}\big(n^{-\delta(1-\varepsilon)}\log n\big)$. Using both these results in (5.46), we arrive at

$$\mathbb{E}\bigg[\sum_{n^{1-\varepsilon}\leq j_\ell\leq n}\mathbb{P}_W(Y\leq\log(n/j_\ell))\bigg]\Big(1+\mathcal{O}\big(n^{-\delta(1-\varepsilon)}\log n\big)\Big). \tag{5.47}$$

As the conditional probability is decreasing in j_ℓ, we can bound the sum from above by an integral almost surely to obtain

$$\begin{aligned}\int_{\lfloor n^{1-\varepsilon}\rfloor}^n \mathbb{P}_W(Y\leq\log(n/x))\,\mathrm{d}x &= \int_{\lfloor n^{1-\varepsilon}\rfloor}^n\int_0^{\log(n/x)}\frac{a_\ell^{m_\ell}}{m_\ell!}y^{m_\ell-1}\mathrm{e}^{-a_\ell y}\,\mathrm{d}y\mathrm{d}x\\ &= \int_0^{\log(n/\lfloor n^{1-\varepsilon}\rfloor)}\int_{\lfloor n^{1-\varepsilon}\rfloor}^{n\mathrm{e}^{-y}}\frac{a_\ell^{m_\ell}}{m_\ell!}y^{m_\ell-1}\mathrm{e}^{-a_\ell y}\,\mathrm{d}x\mathrm{d}y\\ &= n\int_0^{\log(n/(\lfloor n^{1-\varepsilon}\rfloor}\frac{a_\ell^{m_\ell}}{m_\ell!}y^{m_\ell-1}\mathrm{e}^{-(1+a_\ell)y}\,\mathrm{d}y\\ &= n\Big(\frac{a_\ell}{1+a_\ell}\Big)^{m_\ell}\mathbb{P}_W\big(Y'\leq\log\big(n/\lfloor n^{1-\varepsilon}\rfloor\big)\big)\,.\end{aligned}$$

Here, we switch the integration over x and y in the second step and let Y', conditionally on W_ℓ, be a $\Gamma(m_\ell,1+a_\ell)$ random variable. We can then bound the conditional probability from above by one almost surely. Combining this almost sure upper bound with (5.47) in (5.45), we arrive at

$$n^{k-|S|}\prod_{\ell\in[k]\setminus S}\mathbb{E}\Big[\Big(\frac{a_\ell}{1+a_\ell}\Big)^{m_\ell}\Big]\Big(1+\mathcal{O}\big(n^{-\delta(1-\varepsilon)}\log n\big)\Big). \tag{5.48}$$

Finally, with the same steps as in (5.27), we obtain

$$n^{k-|S|}\prod_{\ell\in[k]\setminus S}\mathbb{E}\Big[\Big(\frac{W}{\mathbb{E}[W]+W}\Big)^{m_\ell}\Big]\big(1+o\big(n^{-\delta(1-\varepsilon)(1-\xi)}\big)\big),$$

for any $\xi > 0$. We then use this bound in (5.44) to find, for some positive constant K, the upper bound

$$K\sum_{i=1}^{k-1}\sum_{\substack{S\subseteq[k]\\|S|=i}} n^{C(S)(1+o(1))}\prod_{\ell\in[k]\setminus S}\mathbb{E}\left[\Big(\frac{W}{\mathbb{E}[W]+W}\Big)^{m_\ell}\right] = K\sum_{i=1}^{k-1}\sum_{\substack{S\subseteq[k]\\|S|=i}} n^{C(S)(1+o(1))}\prod_{\ell\in[k]\setminus S} p_{\geq m_\ell}.$$

By Remark 5.4, the tail probability $p_{\geq m_\ell} = \mathcal{O}(p_{m_\ell})$ and by (5.42) we have $n^{C(S)(1+o(1))} = o\big(n^{-\eta(S)}\prod_{\ell\in S} p_{m_\ell}\big)$ for some $\eta(S) > 0$. Combined, this yields

$$\begin{aligned}\frac{1}{(n)_k}\sum_{i=1}^{k-1}\sum_{\mathbf{j}\in I_n(\varepsilon,i)}\mathbb{E}\left[\mathbb{1}_{E_n'}\mathbb{1}_{E_n}\prod_{\ell=1}^{k}\mathbb{P}_W(\mathcal{Z}_n(j_\ell)\geq m_\ell)\right] &\leq K\sum_{i=1}^{k-1}\sum_{\substack{S\subseteq[k]\\|S|=i}} n^{C(S)(1+o(1))}\prod_{\ell\in[k]\setminus S} p_{\geq m_\ell}\\ &= o\big(n^{-\widetilde{\eta}}\prod_{\ell=1}^{k} p_{m_\ell}\big),\end{aligned}$$

with

$$\widetilde{\eta} := \min_{\substack{S\subseteq[k]\\1\leq|S|\leq k-1}}\Big(C(S) - \sum_{\ell\in S}\log(\theta+\xi)c_\ell\Big),$$

which is strictly positive when ξ is sufficiently small and ε is set sufficiently close to one, similar to what is discussed above. Combining this with the fact that $\mathbb{P}((E_n')^c)$ and $\mathbb{P}(E_n^c)$ are $o\big(n^{-\eta}\prod_{\ell=1}^k p_{m_\ell}\big)$ uniformly in $m_1,\dots,m_k < c\log n$ for some $\eta > 0$ (we prove this for the former probability at the end, and for the latter probability this follows from (5.15)), and the result in (5.43), we finally conclude that

$$\begin{aligned}\frac{1}{(n)_k}\sum_{\mathbf{j}\in I_n(\varepsilon)}\mathbb{E}\left[\mathbb{P}_W(\mathcal{Z}_n(j_\ell)=m_\ell,\ell\in[k])\right] &= \frac{1}{(n)_k}\sum_{\mathbf{j}\in I_n(\varepsilon,k)}\mathbb{E}\left[\mathbb{P}_W(\mathcal{Z}_n(j_\ell)=m_\ell,\ell\in[k])\right]\\ &\quad+\frac{1}{(n)_k}\sum_{i=1}^{k-1}\sum_{\mathbf{j}\in I_n(\varepsilon,i)}\mathbb{E}\left[\mathbb{P}_W(\mathcal{Z}_n(j_\ell)=m_\ell,\ell\in[k])\right]\\ &= o\Big(n^{-\eta}\prod_{\ell=1}^{k} p_{m_\ell}\Big),\end{aligned}$$

for some $\eta > 0$ in the case that $m_\ell = c_\ell\log n(1+o(1))$ with $c_\ell\in[1/\log\theta,\theta/(\theta-1))$ for all $\ell\in[k]$ as well.

When the m_ℓ do not all behave the same, that is, for some $\ell\in[k]$ $c_\ell\in[0,1/\log\theta)$ and for some $c_\ell\in[1/\log\theta,\theta/(\theta-1))$, we can use a combination of the approaches outlined for either of the cases.

It remains to prove that $\mathbb{P}((E_n')^c)$ decays sufficiently fast. By a union bound and using the same approach as in (5.14) and (5.15), we find that for some positive constant $C_{\theta,\delta}$,

$$\mathbb{P}\big((E_n')^c\big) \leq \sum_{j=\lceil (C\log n)^{1/(1-2\delta)}\rceil}^{\infty}\exp\big\{-c_\theta j^{1-2\delta}\big\} \leq C_{\theta,\delta}\Gamma\Big(\frac{1}{1-2\delta}, c_\theta\lfloor (C\log n)^{1/(1-2\delta)}\rfloor^{1-2\delta}\Big).$$

Using that $\Gamma(s,x) = x^{s-1}\mathrm{e}^{-x}(1+o(1))$ for a fixed $s\in\mathbb{R}$ and as x tends to infinity, we obtain

$$\begin{aligned}\mathbb{P}\big((E_n')^c\big) &\leq C_{\theta,\delta}\Big(c_\theta C\log n\Big)^{2\delta/(1-2\delta)}\exp\Big\{-c_\theta\lfloor (C\log n)^{1/(1-2\delta)}\rfloor^{1-2\delta}\Big\}(1+o(1))\\ &= \widetilde{C}_{\theta,\delta}(\log n)^{2\delta/(1-2\delta)}\exp\{-c_\theta C\log n\}(1+o(1))\\ &= n^{-c_\theta C(1+o(1))}.\end{aligned}\tag{5.49}$$

As mentioned when introducing the event E_n' in (5.40), the choice of C yields $\mathbb{P}((E_n')^c) \leq n^{-(k\theta\log(\theta)/(\theta-1)+\eta)}$ for n large and η sufficiently small, so that $\mathbb{P}((E_n')^c) = o(\prod_{\ell=1}^k p_{m_\ell} n^{-\eta})$ for any choice of $m_\ell < (\theta/(\theta-1))\log n(1+o(1)), \ell \in [k]$, which concludes the proof. □

5.4. **Proof of Proposition 5.1.** We finally prove Proposition 5.1, using Lemmas 5.10 and 5.11.

Proof of Proposition 5.1. As discussed before, (5.1) directly follows from (5.12) combined with Lemmas 5.10 and 5.11. Using (5.1), we then prove (5.2). For ease of writing, we recall that

$$p_k := \mathbb{E}\left[\frac{\theta-1}{\theta-1+W}\Big(\frac{W}{\theta-1+W}\Big)^k\right], \qquad p_{\geq k} := \mathbb{E}\left[\Big(\frac{W}{\theta-1+W}\Big)^k\right].$$

We start by assuming that $m_\ell = c_\ell \log n(1+o(1))$ with $c_\ell \in (0,c)$ for each $\ell \in [k]$. We discuss how to adjust the proof when $m_\ell = o(\log n)$ for some or all $\ell \in [k]$ at the end.

For each $\ell \in [k]$ take an $\eta_\ell \in (0, c-c_\ell)$ so that $\lceil (1+\eta_\ell)m_\ell\rceil < c\log n$. Then, we use the upper bound

$$\begin{aligned}\mathbb{P}(\mathcal{Z}_n(v_\ell) \geq m_\ell, \ell \in [k]) \leq{}& \sum_{j_1=m_1}^{\lfloor (1+\eta_1)m_1\rfloor}\cdots\sum_{j_k=m_k}^{\lfloor (1+\eta_k)m_k)\rfloor} \mathbb{P}(\mathcal{Z}_n(v_\ell) = j_\ell, \ell \in [k])\\ &+\sum_{i=1}^k \mathbb{P}(\mathcal{Z}_n(v_i) \geq \lceil (1+\eta_i)m_i\rceil, \mathcal{Z}_n(v_\ell) \geq m_\ell, \ell \neq i).\end{aligned} \tag{5.50}$$

We first discuss the first term on the right-hand side. As (5.1) holds uniformly in $m_1, \dots, m_k < c\log n$, we find

$$\begin{aligned}\sum_{j_1=m_1}^{\lfloor (1+\eta_1)m_1\rfloor}\cdots\sum_{j_k=m_k}^{\lfloor (1+\eta_k)m_k)\rfloor} \mathbb{P}(\mathcal{Z}_n(v_\ell) = j_\ell, \ell \in [k]) &= \sum_{j_1=m_1}^{\lfloor (1+\eta_1)m_1\rfloor}\cdots\sum_{j_k=m_k}^{\lfloor (1+\eta_k)m_k)\rfloor} \prod_{\ell=1}^k p_{j_\ell}\big(1+o\big(n^{-\beta}\big)\big)\\ &= \prod_{\ell=1}^k \big(p_{\geq m_\ell} - p_{\geq \lfloor (1+\eta_\ell)m_\ell\rfloor}\big)\big(1+o\big(n^{-\beta}\big)\big) \\ &\leq \prod_{\ell=1}^k p_{m_\ell}\big(1+o\big(n^{-\beta}\big)\big).\end{aligned} \tag{5.51}$$

To finish the upper bound, it remains to show the the term on the second line of (5.50) can be incorporated in the $o\big(n^{-\beta}\big)$ term, and it suffices to show this can be done for each term in the sum, independent of the value of i. Using the negative quadrant dependence of the degrees under the conditional probability measure $\mathbb{P}_W$ (see Remark 5.2 and [13, Lemma 7.1]), we find

$$\begin{aligned}&\mathbb{P}(\mathcal{Z}_n(v_i) \geq \lceil (1+\eta_i)m_i\rceil, \mathcal{Z}_n(v_\ell) \geq m_\ell, \ell \in [k]\backslash\{i\})\\ &\quad= \frac{1}{(n)_k}\sum_{1\leq j_1\neq\ldots\neq j_k\leq n} \mathbb{E}\Big[\mathbb{P}_W(\mathcal{Z}_n(j_i) \geq \lceil (1+\eta_i)m_i\rceil)\prod_{\ell\in[k]\backslash\{i\}} \mathbb{P}_W(\mathcal{Z}_n(j_\ell) \geq m_\ell)\Big]\\ &\quad= \frac{1}{(n)_k}\sum_{\mathbf{j}\in I_n(\varepsilon)} \mathbb{E}\Big[\mathbb{P}_W(\mathcal{Z}_n(j_i) \geq \lceil (1+\eta_i)m_i\rceil)\prod_{\ell\in[k]\backslash\{i\}} \mathbb{P}_W(\mathcal{Z}_n(j_\ell) \geq m_\ell)\Big]\\ &\qquad+ \frac{1}{(n)_k}\sum_{n^{1-\varepsilon}\leq j_1\neq\ldots\neq j_k\leq n} \mathbb{E}\Big[\mathbb{P}_W(\mathcal{Z}_n(j_i) \geq \lceil (1+\eta_i)m_i\rceil)\prod_{\ell\in[k]\backslash\{i\}} \mathbb{P}_W(\mathcal{Z}_n(j_\ell) \geq m_\ell)\Big].\end{aligned}$$

The first term in the last step can be included in the little o term in (5.51) (even when considering m_i rather than $\lceil (1+\eta_i)m_i\rceil$ in the probability), as follows from computations similar to the ones in (5.38) through (5.49), combined with Remark 5.4 (which states that $p_{\geq k} = \mathcal{O}(p_k)$). It remains to show that the same holds for the second term in the last step.

Again, we use an argument similar to the steps performed in (5.44) through (5.48) to arrive at

$$\begin{aligned}\frac{1}{(n)_k}&\sum_{n^{1-\varepsilon}\le j_1\neq\ldots\neq j_k\le n}\mathbb{E}\Big[\mathbb{P}_W(\mathcal{Z}_n(j_i)\ge\lceil(1+\eta_i)m_i\rceil)\prod_{\ell\in[k]\setminus\{i\}}\mathbb{P}_W(\mathcal{Z}_n(j_\ell)\ge m_\ell)\Big]\\&\le Kp_{\ge\lceil(1+\eta_i)m_i\rceil}\prod_{\ell\in[k]\setminus\{i\}}p_{\ge m_\ell},\end{aligned}$$

for some positive constant K. By Lemma 5.5 we have the inequalities

$$p_{\ge\lceil(1+\eta_i)m_i\rceil}\le\theta^{-\lceil(1+\eta_i)m_i\rceil}\le\theta^{-m_i}\theta^{-\eta_i m_i},\qquad p_{\ge m_i}\ge(\theta+\xi)^{-m_i},$$

for any $\xi>0$. As a result, taking $\xi\in(0,\theta(\theta^{\eta_i}-1))$ and setting $\phi_i:=1-(1+\xi/\theta)\theta^{-\eta_i}>0$, we obtain

$$p_{\ge\lceil(1+\eta_i)m_i\rceil}\le(\theta+\xi)^{-m_i}\big((1+\xi/\theta)\theta^{-\eta_i}\big)^{m_i}\le p_{\ge m_i}(1-\phi_i)^{m_i}. \tag{5.52}$$

As $m_i=c_i\log n(1+o(1))$, it follows that $(1-\phi_i)^{m_i}=n^{-c_i\log(1/(1-\phi_i))(1+o(1))}$, so that

$$\frac{1}{(n)_k}\sum_{n^{1-\varepsilon}\le j_1\neq\ldots\neq j_k\le n}\mathbb{E}\Big[\mathbb{P}_W(\mathcal{Z}_n(j_i)\ge\lceil(1+\eta_i)m_i\rceil)\prod_{\ell\in[k]\setminus\{i\}}\mathbb{P}_W(\mathcal{Z}_n(j_\ell)\ge m_\ell)\Big]$$

can be incorporated in the little o term in (5.51) for each $i\in[k]$ when we take a $\beta'<\beta\wedge\min_{i\in[k]}c_i\log(1/(1-\phi_i))$. This yields

$$\mathbb{P}(\mathcal{Z}_n(v_\ell)\ge m_\ell,\ell\in[k])\le\prod_{\ell=1}^k\mathbb{E}\left[\Big(\frac{W}{\theta-1+W}\Big)^{m_\ell}\right]\big(1+o(n^{-\beta'})\big). \tag{5.53}$$

For a lower bound, we can omit the second line of (5.50) and use (5.51) to immediately obtain

$$\begin{aligned}\mathbb{P}(\mathcal{Z}_n(v_\ell)\ge m_\ell,\ell\in[k])&\ge\sum_{j_1=m_1}^{\lfloor(1+\eta_1)m_1\rfloor}\cdots\sum_{j_k=m_k}^{\lfloor(1+\eta_k)m_k)\rfloor}\mathbb{P}(\mathcal{Z}_n(v_\ell)=i_\ell,\ell\in[k])\\&=\prod_{\ell=1}^k\big(p_{\ge m_\ell}-p_{\ge\lfloor(1+\eta_\ell)m_\ell\rfloor}\big)\big(1+o(n^{-\beta})\big).\end{aligned}$$

Again using (5.52) yields $p_{\ge m_\ell}-p_{\ge\lfloor(1+\eta_\ell)m_\ell\rfloor}=p_{\ge m_\ell}\big(1+o(n^{-\beta'})\big)$ when we set $\beta'<\beta\wedge\min_{i\in[k]}c_i\log(1/(1-\phi_i))$. Combined with (5.53) this yields (5.2) and concludes the proof. □

6. Proofs of the main theorems

With the tools developed in Section 5, in particular Propositions 5.1 and 5.6 and Lemma 5.8, we now prove the main results formulated in Section 2.

First, we prove the main result for high degree vertices when the vertex-weight distribution has an atom at one, as in the **(Atom)** case.

Proof of Theorem 2.5. The proof follows the same argument as [1, Theorem 1.2]. For an integer subsequence $(n_\ell)_{\ell\in\mathbb{N}}$ such that $\varepsilon_{n_\ell}\to\varepsilon$ as $\ell\to\infty$, it suffices to prove that for any $i<i'\in\mathbb{Z}$,

$$(X_i^{(n_\ell)},X_{i+1}^{(n_\ell)},\ldots,X_{i'-1}^{(n_\ell)},X_{\ge i'}^{(n_\ell)})\xrightarrow{d}(\mathcal{P}^\varepsilon(i),\mathcal{P}^\varepsilon(i+1),\ldots,\mathcal{P}^\varepsilon(i'-1),\mathcal{P}^\varepsilon([i',\infty)))\qquad\text{as }\ell\to\infty$$

holds. We obtain this via the convergence of the factorial moments of $X_i^{(n_\ell)}, \ldots, X_{i'-1}^{(n_\ell)}, X_{\geq i'}^{(n_\ell)}$. Recall r_k in (2.8). By Proposition 5.6, for any non-negative integers $a_i, \ldots, a_{i'}$,

$$\begin{aligned}\mathbb{E}\left[\left(X_{\geq i'}^{(n_\ell)}\right)_{a_{i'}} \prod_{k=i}^{i'-1}\left(X_k^{(n_\ell)}\right)_{a_k}\right] &= \left(q_0\theta^{-i'+\varepsilon_{n_\ell}}\right)^{a_{i'}} \prod_{k=i}^{i'-1}\left(q_0(1-\theta^{-1})\theta^{-k+\varepsilon_{n_\ell}}\right)^{a_k} \\ &\quad \times \left(1+\mathcal{O}\big(r_{\lfloor \log_\theta n_\ell\rfloor + i} \vee n_\ell^{-\beta}\big)\right) \\ &\to \left(q_0\theta^{-i'+\varepsilon}\right)^{a_{i'}} \prod_{k=i}^{i'-1}\left(q_0(1-\theta^{-1})\theta^{-k+\varepsilon}\right)^{a_k},\end{aligned}$$

as $\ell \to \infty$. By using the properties of Poisson processes, it follows that the limit equals

$$\mathbb{E}\left[\left(\mathcal{P}^\varepsilon[i', \infty)\right)_{a_{i'}} \prod_{k=i}^{i'-1}\left(\mathcal{P}^\varepsilon(k)\right)_{a_k}\right],$$

due to the particular form of the intensity measure of the Poisson process $\mathcal{P}$ (which is used in the definition of the Poisson process $\mathcal{P}^\varepsilon$). The result then follows from [12, Theorem 6.10]. □

For the results for the (**Weibull**) and (**Gumbel**) cases, as outlined in Theorems 2.6 and 2.7, respectively, we combine the asymptotic behaviour of $p_{\geq k}$ in Theorem 5.3 with Proposition 5.1 and Lemma 5.8.

Proof of Theorem 2.6. To establish the convergence in probability, it follows from Lemma 5.8 that we need only consider $n\mathbb{P}(\mathcal{Z}_n(v_1) \geq k_n)$ for some adequate integer-valued k_n such that $k_n < c\log n$ for some $c \in (0, \theta/(\theta-1))$ and where v_1 is a vertex selected from $[n]$ uniformly at random. By Proposition 5.1, this quantity equals $np_{\geq k_n}(1+o(1))$. Then, we use Theorem 5.3 and Remark 5.4 to obtain that, when W satisfies the (**Weibull**) case in Assumption 2.3, this quantity is at most

$$n\overline{C}\,\overline{L}(k_n)k_n^{-(\alpha-1)}\theta^{-k_n},$$

where $\overline{C} > 1$ is a constant. Now fix an arbitrary $\eta > 0$ and set $k_n := \lfloor \log_\theta n - (\alpha-1)(1-\eta)\log_\theta\log_\theta n\rfloor$. This yields

$$\begin{aligned} &n\overline{C}\,\overline{L}(\log_\theta n(1+o(1)))(\log_\theta n)^{-(\alpha-1)}\theta^{-\lfloor \log_\theta n-(\alpha-1)(1-\eta)\log_\theta\log_\theta n\rfloor}(1+o(1)) \\ &\quad\leq \overline{C}_2\overline{L}(\log_\theta n)(\log_\theta n)^{-(\alpha-1)}(\log_\theta n)^{(\alpha-1)(1-\eta)} \\ &\quad= \overline{C}_2\overline{L}(\log_\theta n)(\log_\theta n)^{-(\alpha-1)\eta}.\end{aligned} \tag{6.1}$$

Here, $\overline{C}_2 > 0$ is a suitable constant and we use that $k_n = \log_\theta n(1+o(1))$ in the first step. Furthermore, we use [2, Theorem 1.5.2], which states that for a slowly-varying function $\overline{L}$, $\overline{L}(\lambda x)/\overline{L}(x)$ converges to one uniformly in λ on any interval $[a,b]$ such that $0 < a \leq b < \infty$. As a result, $\overline{L}(\log_\theta n(1+o(1))) \leq c\overline{L}(\log_\theta n)$ for some constant $c > 1$ and n sufficiently large. Finally, we use [2, Proposition 1.3.6 (v)] to obtain that for any $\eta > 0$, the final line of (6.1) tends to zero with n. This shows that for any $\eta > 0$, with high probability,

$$\max_{j\in[n]} \frac{\mathcal{Z}_n(j) - \log_\theta n}{\log_\theta\log_\theta n} \leq -(\alpha-1)(1-\eta)$$

holds, due to the first result in Lemma 5.8. A similar approach, when setting $k_n := \lfloor \log_\theta n - (\alpha-1)(1+\eta)\log_\theta\log_\theta n\rfloor$, yields

$$n\mathbb{P}(\mathcal{Z}_n(v_1) \geq k_n) \to \infty,$$

so that for any $\eta > 0$, with high probability,

$$\max_{j\in[n]} \frac{\mathcal{Z}_n(j) - \log_\theta n}{\log_\theta\log_\theta n} \geq -(\alpha-1)(1+\eta)$$

holds. Together, these two bounds prove the desired result. □

Proof of Theorem 2.7. The proof of this theorem follows a similar approach to the proof of Theorem 2.6. That is, we again apply the results from Theorem 5.3 together with the fact that

$$n\mathbb{P}(\mathcal{Z}_n(v_1) \geq k_n) = np_{\geq k_n}(1+o(1)),$$

for some adequate integer-valued k_n such that $k_n < c\log n$ for some $c \in (0, \theta/(\theta-1))$, as follows from Proposition 5.1 and Lemma 5.8. In the **(Gumbel)**-**(RV)** sub-case, we know that

$$p_{\geq k_n} = \exp\Big\{-\frac{\tau^\gamma}{1-\gamma}\Big(\frac{(1-\theta^{-1})k_n}{c_1}\Big)^{1-\gamma}(1+o(1))\Big\}\theta^{-k_n}, \tag{6.2}$$

where we recall that $\gamma = 1/(\tau+1)$. To prove the desired results, we first set $k_n = \lfloor \log_\theta n - (1+\eta)C_{\theta,\tau,c_1}(\log_\theta n)^{1-\gamma}\rfloor$ for any $\eta > 0$, where we recall C_{θ,τ,c_1} from (2.5). Using this in (6.2) then yields

$$np_{\geq k_n} = \frac{n}{\theta^{k_n}}\mathrm{e}^{-\log(\theta)C_{\theta,\tau,c_1}k_n^{1-\gamma}(1+o(1))} \geq \mathrm{e}^{\eta\log(\theta)C_{\theta,\tau,c_1}(\log_\theta n)^{1-\gamma}(1+o(1))},$$

where we use that $k_n^{1-\gamma} = (\log_\theta n)^{1-\gamma}(1+o(1))$ in the last step. Hence, $n\mathbb{P}(\mathcal{Z}_n(v_1) \geq k_n)$ diverges. We thus conclude from Lemma 5.8 that

$$\max_{j\in[n]}\frac{\mathcal{Z}_n(j) - \log_\theta n}{(\log_\theta n)^{1-\gamma}} \geq -(1+\eta)C_{\theta,\tau,c_1}$$

holds with high probability. A similar approach, setting $k_n := \lceil \log_\theta n - (1-\eta)C_{\theta,\tau,c_1}(\log_\theta n)^{1-\gamma}\rceil$ and combining this with the first result of Lemma 5.8 yields

$$\max_{j\in[n]}\frac{\mathcal{Z}_n(j) - \log_\theta n}{(\log_\theta n)^{1-\gamma}} \leq -(1-\eta)C_{\theta,\tau,c_1}$$

holds with high probability. Together, these two bounds prove (2.5). To prove (2.6) we apply the same methodology but use the asymptotic expression of p_k (and $p_{\geq k}$ by adjusting constants), as in (5.5). We recall the constants C_1, C_2, C_3 from (2.7) and set $k_n =: \lceil \log_\theta n - C_1(\log_\theta\log_\theta n)^\tau + C_2(\log_\theta\log_\theta n)^{\tau-1}\log_\theta\log_\theta\log_\theta n + (C_3+\eta)(\log_\theta\log_\theta n)^{\tau-1}\rceil$, for any $\eta > 0$. Then, (5.5) yields

$$np_{\geq k_n} = \frac{n}{\theta^{k_n}}\exp\Big\{-\Big(\frac{\log k_n}{c_1}\Big)^\tau\Big(1 + \frac{\tau(\tau-1)\log\log k_n}{\log k_n} - \frac{K_{\tau,c_1,\theta}}{\log k_n}(1+o(1))\Big)\Big\}.$$

Using Taylor expansions, we obtain

$$\begin{aligned}
-\Big(\frac{\log k_n}{c_1}\Big)^\tau &= -\Big(\frac{\log\log_\theta n}{c_1}\Big)^\tau + o(1) = -\log(\theta)C_1(\log_\theta\log_\theta n)^\tau + o(1),\\
\frac{\tau(\tau-1)}{c_1^\tau}(\log k_n)^{\tau-1}\log\log k_n &= \frac{\tau(\tau-1)}{c_1^\tau}(\log\log_\theta n)^{\tau-1}\log\log\log_\theta n + o(1)\\
&= \log(\theta)C_2(\log_\theta\log_\theta n)^{\tau-1}\log_\theta\log_\theta\log_\theta n\\
&\quad + (\log_\theta(\log\theta))(\log\theta)^\tau\frac{\tau(\tau-1)}{c_1^\tau}(\log_\theta\log_\theta n)^{\tau-1} + o(1),\\
-\frac{K_{\tau,c_1,\theta}}{c_1^\tau}(\log k_n)^{\tau-1} &= -\frac{K_{\tau,c_1,\theta}}{c_1^\tau}(\log\log_\theta n)^{\tau-1} + o(1)\\
&= -(\log\theta)^{\tau-1}\frac{K_{\tau,c_1,\theta}}{c_1^\tau}(\log\log_\theta n)^{\tau-1} + o(1),
\end{aligned}$$

where we recall $K_{\tau,c_1,\theta}$ from (5.5) in Theorem 5.3, so that

$$\begin{aligned}
&n\exp\Big\{-\Big(\frac{\log k_n}{c_1}\Big)^\tau + \frac{\tau(\tau-1)}{c_1^\tau}(\log k_n)^{\tau-1}\log\log k_n - \frac{K_{\tau,c_1,\theta}}{c_1^\tau}(\log k_n)^{\tau-1}(1+o(1))\Big\}\theta^{-k_n}\\
&= n\exp\big\{-\log(\theta)C_1(\log_\theta\log_\theta n)^\tau + \log(\theta)C_2(\log_\theta\log_\theta n)^{\tau-1}\log_\theta\log_\theta\log_\theta n\\
&\qquad + \log(\theta)C_3(\log_\theta\log_\theta n)^{\tau-1}(1+o(1))\big\}\theta^{-k_n}\\
&\leq \exp\big\{-(\eta - o(1))(\log_\theta\log_\theta n)^{\tau-1}\big\},
\end{aligned}$$

where we use that $k_n \geq \log_\theta n - C_1(\log_\theta \log_\theta n)^\tau + C_2(\log_\theta \log_\theta n)^{\tau-1} \log_\theta \log_\theta \log_\theta n + (C_3 + \eta)(\log_\theta \log_\theta n)^{\tau-1}$ in the last step. As the right-hand side tends to zero with n, Lemma 5.8 yields for any fixed $\eta > 0$, with high probability,

$$\max_{i\in[n]} \frac{\mathcal{Z}_n(i) - \big(\log_\theta n - C_1(\log_\theta \log_\theta n)^\tau + C_2(\log_\theta \log_\theta n)^{\tau-1} \log_\theta \log_\theta \log_\theta n\big)}{(\log_\theta \log_\theta n)^{\tau-1}} \leq C_3 + \eta.$$

With a similar approach, setting

$$k_n = \lfloor \log_\theta n - C_1(\log_\theta \log_\theta n)^\tau + C_2(\log_\theta \log_\theta n)^{\tau-1} \log_\theta \log_\theta \log_\theta n + (C_3 - \eta)(\log_\theta \log_\theta n)^{\tau-1} \rfloor,$$

we can obtain that for any fixed $\eta > 0$, with high probability,

$$\max_{i\in[n]} \frac{\mathcal{Z}_n(i) - \big(\log_\theta n - C_1(\log_\theta \log_\theta n)^\tau + C_2(\log_\theta \log_\theta n)^{\tau-1} \log_\theta \log_\theta \log_\theta n\big)}{(\log_\theta \log_\theta n)^{\tau-1}} \geq C_3 - \eta.$$

Together these two bounds yield (2.6), which concludes the proof. $\square$

Proof of Theorem 2.8. The proof follows the same approach as the proof of [1, Theorem 1.3]. We thus need to consider two cases: $i = \mathcal{O}(1)$ and $i \to \infty$ such that $i + \log_\theta n < (\theta/(\theta-1)) \log n$ and $\liminf_{n\to\infty} i > -\infty$. For the former case, as $\exp\{-q_0\theta^{-i+\varepsilon_n}\} = \mathcal{O}(1)$, it suffices to prove

$$\mathbb{P}\bigg(\max_{j\in[n]} \mathcal{Z}_n(j) \geq \lfloor \log_\theta n \rfloor + i\bigg) - (1 - \exp\{-q_0\theta^{-i+\varepsilon_n}\}) \to 0 \qquad \text{as } n \to \infty.$$

By the definition of $X^{(n)}_{\geq i}$ in (2.3), this is equivalent to

$$\mathbb{P}\Big(X^{(n)}_{\geq i} = 0\Big) - \exp\{-q_0\theta^{-i+\varepsilon_n}\} \to 0 \qquad \text{as } n \to \infty. \tag{6.3}$$

This follows from Theorem 2.5 and the subsubsequence principle. That is, if we assume the convergence in (6.3) does not hold, then there exists a subsequence $(n_\ell)_{\ell\in\mathbb{N}}$ and a $\delta > 0$ such that

$$\mathbb{P}\Big(X^{(n_\ell)}_{\geq i} = 0\Big) - \exp\{-q_0\theta^{-i+\varepsilon_{n_\ell}}\} > \delta \;\; \forall\, \ell \in \mathbb{N}. \tag{6.4}$$

However, as ε_{n_ℓ} is bounded, there exists a subsubsequence $\varepsilon_{n_{\ell_k}}$ such that $\varepsilon_{n_{\ell_k}} \to \varepsilon$ for some $\varepsilon \in (0, 1]$. Then, by Theorem 2.5, the statement in (6.4) is false, from which the result follows

In the latter case, we need only consider $\mathbb{E}\left[X^{(n)}_{\geq i}\right]$ and $\mathbb{E}\left[(X^{(n)}_{\geq i})_2\right]$, as

$$\mathbb{E}\left[X^{(n)}_{\geq i}\right] - \frac{1}{2}\mathbb{E}\left[(X^{(n)}_{\geq i})_2\right] \leq \mathbb{P}\Big(X^{(n)}_{\geq i} > 0\Big) \leq \mathbb{E}\left[X^{(n)}_{\geq i}\right], \tag{6.5}$$

again see [1, Theorem 1.3] and its proof for more details. By Proposition 5.6, we have that

$$\mathbb{E}\left[X^{(n)}_{\geq i}\right] = q_0\theta^{-i+\varepsilon_n}(1 + o(1)), \qquad \mathbb{E}\left[(X^{(n)}_{\geq i})_2\right] = \big(q_0\theta^{-i+\varepsilon_n}\big)^2(1 + o(1)).$$

Hence, as $i \to \infty$ and ε_n is bounded,

$$\begin{aligned} \mathbb{E}\left[X^{(n)}_{\geq i}\right] &= \big(1 - \exp\left\{q_0\theta^{-i+\varepsilon_n}\right\}\big)(1 + o(1)), \\ \mathbb{E}\left[X^{(n)}_{\geq i}\right] - \frac{1}{2}\mathbb{E}\left[(X^{(n)}_{\geq i})_2\right] &= \big(1 - \exp\left\{q_0\theta^{-i+\varepsilon_n}\right\}\big)(1 + o(1)). \end{aligned}$$

Combining this with (6.5) yields the desired result. $\square$

Proof of Theorem 2.9. The proof follows the same argument as the proof of [1, Theorem 1.4], which is based on [3, Theorem 1.24]. Let $1 \leq a \leq b$ be integers. Then, by Proposition 5.6 and since $i = o(\log n)$,

$$\mathbb{E}\left[\Big(X^{(n)}_i\Big)_a\right] - \Big(q_0(1 - \theta^{-1})\theta^{-i+\varepsilon_n}\Big)^a = \mathcal{O}\big(\theta^{-ia}\big(r_{\lfloor \log_\theta n + i \rfloor} \vee n^{-\beta}\big)\big).$$

It then remains to show that the right-hand side is in fact $o(\theta^{ib})$. We note that $i = o(\log r_{\log_\theta n} \wedge \log n)$, so that we can write the right-hand side as

$$\mathcal{O}\big((r_{\log_\theta n})^{1-ia\log\theta/\log r_{\log_\theta n}} \vee n^{-\beta-ia\log\theta/\log n}\big) = \mathcal{O}\big((r_{\log_\theta n})^{1-o(1)} \vee n^{-\beta-o(1)}\big) = o(\theta^{ib}),$$

by the constraints on i, from which the result follows. $\square$

7. Technical details of examples

In this section we discuss some technical details of the examples discussed in Section 4. In particular, for each example we provide a precise asymptotic expression of p_k and $p_{\geq k}$ as well as a key element that leads to the results in Section 4. That is, for each of the examples we state and prove the analogue of Proposition 5.6. The three theorems presented in each of the examples in Section 4 mimic three of the theorems presented in Section 2. That is, Theorems 4.2 and 4.6 are the analogue of Theorems 2.5, Theorems 4.3 and 4.7 are the analogue of Theorem 2.8 and Theorems 4.4 and 4.8 are the analogue of Theorem 2.9. As a result, their proofs are very similar to the proofs of Theorems 2.5, 2.8 and 2.9, so we omit proving the theorems stated in Section 4.

7.1. **Example 4.1, beta distribution bounded away from zero.** Let the distribution of W satisfy (4.1) for some $\alpha, \beta > 0, w^* \in (0,1)$. We prove a result on (the tail of) the limiting degree distribution and provide additional building blocks required to prove the results in Example 4.1.

Lemma 7.1. *Let the distribution of W satisfy (4.1) for some $\alpha, \beta > 0, w^* \in (0,1)$ and recall $p_k, p_{\geq k}$ from (2.1). Then,*

$$\begin{aligned} p_k &= Z_{w^*}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)}(1-\theta^{-1})^{1-\beta}k^{-\beta}\theta^{-k}\big(1+\mathcal{O}(1/k)\big),\\ p_{\geq k} &= Z_{w^*}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)}(1-\theta^{-1})^{-\beta}k^{-\beta}\theta^{-k}\big(1+\mathcal{O}(1/k)\big). \end{aligned} \tag{7.1}$$

Note that this Lemma improves on the bounds in (5.3) by providing a precise multiplicative constant, rather than two slowly-varying functions that are (possibly) of different order.

Proof. By the distribution of W, we immediately obtain that

$$p_k = \int_{w^*}^1 (\theta-1)x^k(\theta-1+x)^{-(k+1)}Z_{w^*}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}\,\mathrm{d}x,$$

where Γ is the gamma function. Let us denote the integrand by $f(x,k)$. We first consider the asymptotic behaviour of the integral when its lower bound on the integration variable x is zero, and show that it is equal to the right-hand side of the first equation in (7.1). Then, we show that

$$\int_0^{w^*} f(x,k)\,\mathrm{d}x = o\bigg(\int_{w^*}^1 f(x,k)\,\mathrm{d}x\bigg),$$

which combined yields the desired result. So, we start with

$$\begin{aligned} &\int_0^1 (\theta-1)x^k(\theta-1+x)^{-(k+1)}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}\,\mathrm{d}x\\ &\quad= (\theta-1)^{-k}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_0^1 x^{k+\alpha-1}(1-x)^{\beta-1}(1+x/(\theta-1))^{-(k+1)}\,\mathrm{d}x, \end{aligned} \tag{7.2}$$

where we omit Z_{w^*} for ease of writing. We now use Euler's integral representation of the hypergeometric function. That is, for $a,b,c,z \in \mathbb{C}$ such that $\mathrm{Re}(c) > \mathrm{Re}(b) > 0$ and z is not a real number greater than one,

$$\int_0^1 x^{b-1}(1-x)^{c-b-1}(1-zx)^{-a}\,\mathrm{d}x = \frac{\Gamma(c-b)\Gamma(b)}{\Gamma(c)}{}_2F_1(a,b,c,z),$$

where ${}_2F_1$ is the hypergeometric function. Applying this in (7.2), we thus obtain

$$(\theta-1)^{-k}\frac{\Gamma(\alpha+\beta)\Gamma(k+\alpha)}{\Gamma(\alpha)\Gamma(k+\alpha+\beta)}{}_2F_1(k+1,k+\alpha,k+\alpha+\beta,-1/(\theta-1)).$$

We then use one of the Euler transformations of the hypergeometric function,

$${}_2F_1(a,b,c,z)=(1-z)^{c-a-b}{}_2F_1(c-a,c-b,c,z),$$

to arrive at

$$\theta^{-k}\frac{\Gamma(\alpha+\beta)\Gamma(k+\alpha)}{\Gamma(\alpha)\Gamma(k+\alpha+\beta)}\Big(\frac{\theta}{\theta-1}\Big)^{\beta-1}{}_2F_1(\alpha+\beta-1,\beta,k+\alpha+\beta,-1/(\theta-1)). \tag{7.3}$$

We directly find a particular case in which we can find the value of the hypergeometric function explicitly, namely when $\alpha+\beta=1$. When $\alpha+\beta=1$, we find that the hypergeometric function on the right-hand side of (7.3) equals one as the first argument equals zero, independent of the other arguments, so that we arrive at

$$\frac{(1-\theta^{-1})^{1-\beta}\Gamma(k+\alpha)}{\Gamma(\alpha)\Gamma(k+1)}\theta^{-k}=\frac{(1-\theta^{-1})^{1-\beta}}{\Gamma(\alpha)}k^{-\beta}\theta^{-k}\big(1+\mathcal{O}(1/k)\big),$$

since $\Gamma(x+a)/\Gamma(x)=x^a\big(1+\mathcal{O}(1/x)\big)$ as $x\to\infty$ and $\alpha=1-\beta$ in this particular case. When $\alpha+\beta\neq 1$, we can obtain a similar expression. First, we use one of Pfaff's transformations for the hypergeometric function,

$${}_2F_1(a,b,c,z)=(1-z)^{-b}{}_2F_1(b,c-a,c,z/(z-1)).$$

Then, applying this to the right-hand side of (7.3), so that $z/(z-1)=1/\theta\in(-1,1)$, we can express the hypergeometric function as a power series. Namely, for z such that $|z|<1$,

$${}_2F_1(a,b,c,z)=\sum_{j=0}^{\infty}\frac{a^{(j)}b^{(j)}}{c^{(j)}}\frac{z^j}{\Gamma(j)},$$

where $a^{(j)}:=\prod_{\ell=1}^{j}(a+(\ell-1))$ (and similarly for $b^{(j)},c^{(j)}$). Thus, combining the Pfaff transformation and the power series representation yields

$${}_2F_1(\alpha+\beta-1,\beta,k+\alpha+\beta,-1/(\theta-1))=\Big(\frac{\theta}{\theta-1}\Big)^{-\beta}\sum_{j=0}^{\infty}\frac{\beta^{(j)}(k+1)^{(j)}}{(k+\alpha+\beta)^{(j)}}\frac{\theta^{-j}}{j!}. \tag{7.4}$$

From the $\alpha+\beta=1$ case, we immediately distil that

$$\sum_{j=0}^{\infty}\frac{\beta^{(j)}}{j!}\theta^{-j}=\Big(\frac{\theta}{\theta-1}\Big)^{\beta}. \tag{7.5}$$

The aim is to show that for k large, the series in (7.4) is close to $(\theta/(\theta-1))^{\beta}$ for any choice of $\alpha,\beta>0$, so that the entire term in (7.4) is close to one. We consider two cases, namely $\alpha+\beta<1$ and $\alpha+\beta>1$. Let us start with the latter. We immediately obtain the upper bound $(k+\alpha+\beta)^{(j)}>(k+1)^{(j)}$, so that using (7.5) yields that the right-hand side of (7.4) is at most one. For a lower bound, we note that

$$\frac{(k+1)^{(j)}}{(k+\alpha+\beta)^{(j)}}=\prod_{\ell=1}^{j}\Big(1-\frac{\alpha+\beta-1}{k+\alpha+\beta+(\ell-1)}\Big)\geq\Big(1-\frac{\alpha+\beta-1}{k+\alpha+\beta}\Big)^{j},$$

as the fraction in the second step in decreasing in ℓ, since $\alpha+\beta-1>0$. We thus obtain the lower bound

$$\Big(\frac{\theta}{\theta-1}\Big)^{-\beta}\sum_{j=0}^{\infty}\frac{\beta^{(j)}(k+1)^{(j)}}{(k+\alpha+\beta)^{(j)}}\frac{\theta^{-j}}{j!}\geq\Big(\frac{\theta}{\theta-1}\Big)^{-\beta}\sum_{j=0}^{\infty}\frac{\beta^{(j)}}{j!}\Big(\Big(1-\frac{\alpha+\beta-1}{k+\alpha+\beta}\Big)\frac{1}{\theta}\Big)^{j},$$

which, as in (7.5), equals

$$\Big(\frac{\theta-1}{\theta-1+\frac{\alpha+\beta-1}{k+\alpha+\beta}}\Big)^{\beta}=\Big(1-\frac{\alpha+\beta-1}{(\theta-1)(k+\alpha+\beta)+(\alpha+\beta-1)}\Big)^{\beta}=1-\mathcal{O}(1/k).$$

A similar approach can be used for $\alpha+\beta<1$, where one would have an elementary lower bound and an upper bound that is $1+\mathcal{O}(1/k)$. In total, combining the above in (7.4) and then in (7.3) yields

$$\int_0^1 f(x,k)\,\mathrm{d}x = Z_{w^*}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)}(1-\theta^{-1})^{1-\beta}k^{-\beta}\theta^{-k}\big(1+\mathcal{O}(1/k)\big).$$

We then consider

$$\int_0^{w^*}(\theta-1)x^k(\theta-1+x)^{-(k+1)}Z_{w^*}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}\,\mathrm{d}x.$$

We bound the term $(1-x)^{\beta-1}$ from above by $(1-w^*)^{(\beta-1)\wedge 0}$. Since $x^{k+\alpha-1}(1+x/(\theta-1))^{-(k+1)}$ is increasing for $x\in(0,1)$ when k is sufficiently large, we obtain the upper bound

$$\begin{aligned}(\theta-1)^{-k}Z_{w^*}&\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}(1-w^*)^{(\beta-1)\wedge 0}(w^*)^{k+\alpha}\Big(1+\frac{w^*}{\theta-1}\Big)^{-(k+1)}\\ &= Z_{w^*}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}(1-w^*)^{(\beta-1)\wedge 0}(w^*)^{\alpha}\Big(1+\frac{w^*}{\theta-1}\Big)^{-1}\Big(\frac{w^*}{\theta-1+w^*}\Big)^{k}.\end{aligned}$$

Since $w^*/(\theta-1+w^*)$ is increasing in w^*, it is strictly smaller than $1/\theta$. Hence,

$$\int_0^{w^*}(\theta-1)x^k(\theta-1+x)^{-(k+1)}Z_{w^*}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}\,\mathrm{d}x = o\big(\theta^{-k}k^{-(1+\beta)}\big),$$

independent of the value of β, so that

$$\begin{aligned}p_k &= \int_0^1(\theta-1)x^k(\theta-1+x)^{-(k+1)}Z_{w^*}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}\,\mathrm{d}x\\ &\quad -\int_0^{w^*}(\theta-1)x^k(\theta-1+x)^{-(k+1)}Z_{w^*}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}\,\mathrm{d}x\\ &= Z_{w^*}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)}(1-\theta^{-1})^{1-\beta}k^{-\beta}\theta^{-k}\big(1+\mathcal{O}(1/k)\big)-o\big(\theta^{-k}k^{-(1+\beta)}\big)\\ &= Z_{w^*}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)}(1-\theta^{-1})^{1-\beta}k^{-\beta}\theta^{-k}\big(1+\mathcal{O}(1/k)\big).\end{aligned}$$

which proves the first line of (7.1).

An equivalent computation can be performed for

$$\int_0^1 x^k(\theta-1+x)^{-k}Z_{w^*}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}\,\mathrm{d}x, \tag{7.6}$$

to obtain that it equals

$$\begin{aligned}\theta^{-k}Z_{w^*}&\frac{\Gamma(\alpha+\beta)\Gamma(k+\alpha)}{\Gamma(\alpha)\Gamma(k+\alpha+\beta)}\Big(\frac{\theta}{\theta-1}\Big)^{\beta}{}_2F_1(\alpha+\beta,\beta,k+\alpha+\beta,-1/(\theta-1))\\ &=\theta^{-k}Z_{w^*}\frac{\Gamma(\alpha+\beta)\Gamma(k+\alpha)}{\Gamma(\alpha)\Gamma(k+\alpha+\beta)}{}_2F_1(\beta,k,k+\alpha+\beta,1/\theta)\\ &=\theta^{-k}Z_{w^*}\frac{\Gamma(\alpha+\beta)\Gamma(k+\alpha)}{\Gamma(\alpha)\Gamma(k+\alpha+\beta)}\sum_{j=0}^{\infty}\frac{\beta^{(j)}k^{(j)}}{(k+\alpha+\beta)^{(j)}}\frac{\theta^{-j}}{j!}.\end{aligned}$$

In this case an equivalent approach for bounding the sum on the right-hand side works for all $\alpha,\beta>0$. Hence, for (7.6) we obtain the asymptotic expression

$$Z_{w^*}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)}(1-\theta^{-1})^{-\beta}k^{-\beta}\theta^{-k}\big(1+\mathcal{O}(1/k)\big).$$

Via the same we can show that the integral from zero to w^* is asymptotically negligible, which finishes the proof. □

Recall that in this example we set

$$\begin{aligned} X_i^{(n)} &:= |\{j \in [n] : \mathcal{Z}_n(j) = \lfloor \log_\theta n - \beta \log_\theta \log_\theta n \rfloor + i\}|, \\ X_{\geq i}^{(n)} &:= |\{j \in [n] : \mathcal{Z}_n(j) \geq \lfloor \log_\theta n - \beta \log_\theta \log_\theta n \rfloor + i\}|, \\ \varepsilon_n &:= (\log_\theta n - \beta \log_\theta \log_\theta n) - \lfloor \log_\theta n - \beta \log_\theta \log_\theta n \rfloor. \end{aligned}$$

We then state the analogue of Proposition 5.6.

Proposition 7.2. *Consider the WRT model as in Definition 2.1 with vertex-weights $(W_i)_{i\in[n]}$ whose distribution satisfies (4.1) for some $\alpha, \beta > 0, w^* \in (0,1)$. For a fixed $K \in \mathbb{N}, c \in (0, \theta/(\theta-1))$, the following holds. For any $i, i' = i(n), i'(n)$ in $\mathbb{Z}$ such that $0 < \log_\theta n + i < \log_\theta n + i' < c \log n$ and $i, i' \sim \delta \log_\theta n$, for some $\delta \in (-1, c\log\theta - 1) \cup \{0\}$ ($\delta = 0$ denotes $i, i' = o(\log n)$) and for $a_i, \ldots, a_{i'} \in \mathbb{N}_0$ satisfying $a_i + \cdots + a_{i'} = K$,*

$$\begin{aligned} &\mathbb{E}\Big[\Big(X_{\geq i'}^{(n_\ell)}\Big)_{a_{i'}} \prod_{k=i}^{i'-1} \Big(X_k^{(n_\ell)}\Big)_{a_k}\Big] \\ &= \Big(Z_{w^*} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)} \frac{(1-\theta^{-1})^{1-\beta}}{(\theta-1)(1+\delta)^\beta} \theta^{-i'+1+\varepsilon_n}\Big)^{a_{i'}} \prod_{k=i}^{i'-1} \Big(Z_{w^*} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)} \frac{(1-\theta^{-1})^{1-\beta}}{(1+\delta)^\beta} \theta^{-k+\varepsilon_n}\Big)^{a_k} \\ &\quad \times \Big(1 + \mathcal{O}\Big(\frac{\log\log n}{\log n} \vee \frac{|i - \delta \log_\theta n| \vee |i' - \delta \log_\theta n|}{\log n}\Big)\Big). \end{aligned}$$

Proof. Set $K' := K - a_{i'}$ and for each $i \leq k \leq i'$ and for each u such that $\sum_{\ell=i}^{k-1} a_\ell < u \leq \sum_{\ell=i}^{k} a_\ell$, let $m_u = \lfloor \log_\theta n - \log_\theta \log_\theta n \rfloor + k$. Also, let $(v_u)_{u\in[K]}$ be K vertices selected uniformly at random without replacement from $[n]$. Then, as the $X_{\geq k}^{(n)}$ and $X_k^{(n)}$ can be expressed as sums of indicators, following the same steps as in the proof of Proposition 5.6,

$$\mathbb{E}\left[\Big(X_{\geq i'}^{(n)}\Big)_{a_{i'}} \prod_{k=i}^{i'-1} \Big(X_k^{(n)}\Big)_{a_k}\right] = (n)_K \sum_{\ell=0}^{K'} \sum_{\substack{S \subseteq [K'] \\ |S| = \ell}} (-1)^\ell \mathbb{P}\big(\deg(v_u) \geq m_u + \mathbb{1}_{\{u\in S\}}, u \in [K]\big).$$

By Proposition 5.1,

$$\mathbb{P}\big(\deg(v_\ell) \geq m_u + \mathbb{1}_{\{u\in S\}}, u \in [K]\big) = \prod_{u=1}^{K} \mathbb{E}\left[\Big(\frac{W}{\mathbb{E}[W] + W}\Big)^{m_u + \mathbb{1}_{\{u\in S\}}}\right](1 + o(n^{-\beta})),$$

for some $\beta > 0$. By Lemma 7.1, when $|S| = \ell$,

$$\begin{aligned} &\prod_{u=1}^{K} \mathbb{E}\left[\Big(\frac{W}{\mathbb{E}[W] + W}\Big)^{m_u + \mathbb{1}_{\{u\in S\}}}\right] \\ &\quad = \Big(Z_{w^*} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)} (1-\theta^{-1})^{-\beta}\Big)^K \theta^{-\sum_{u=1}^K m_u - \ell} \prod_{u=1}^{K} (m_u + \mathbb{1}_{\{u\in S\}})^{-\beta} (1 + \mathcal{O}(1/\log n)). \end{aligned}$$

Here, we are able to obtain the error term $1 - \mathcal{O}(1/\log n)$ due to the fact that $\log_\theta n + i > \eta \log n$ for some $\eta \in (0, 1+\delta)$ when n is large. Moreover, as $i, i' \sim \delta \log_\theta n$ and thus $m_u \sim (1+\delta)\log_\theta n$ for each $u \in [K]$,

$$\prod_{u=1}^{K} (m_u + \mathbb{1}_{\{u\in S\}})^{-\beta} = ((1+\delta)\log_\theta n)^{-\beta K} \Big(1 + \mathcal{O}\Big(\frac{\log\log n}{\log n} \vee \frac{|i - \delta\log_\theta n| \vee |i' - \delta \log_\theta n|}{\log n}\Big)\Big),$$

uniformly in S (and ℓ). We thus arrive at

$$\begin{aligned}
&(n)_K\sum_{\ell=0}^{K'}\sum_{\substack{S\subseteq[K']\\|S|=\ell}}(-1)^\ell\Big(Z_{w^*}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)}(1-\theta^{-1})^{-\beta}(1+\delta)^{-\beta}(\log_\theta n)^{-\beta}\Big)^K\theta^{-\sum_{u=1}^K m_u-\ell}\\
&\qquad\times\Big(1+\mathcal{O}\Big(\frac{\log\log n}{\log n}\vee\frac{|i-\delta\log_\theta n|\vee|i'-\delta\log_\theta n|}{\log n}\Big)\Big)\\
&=\Big(Z_{w^*}\frac{\Gamma(\alpha+\beta)(1-\theta^{-1})^{1-\beta}}{\Gamma(\alpha)(\theta-1)(1+\delta)^\beta}\theta^{-i'+1+\varepsilon_n}\Big)^{a_{i'}}\prod_{k=i}^{i'-1}\Big(Z_{w^*}\frac{\Gamma(\alpha+\beta)(1-\theta^{-1})^{1-\beta}}{\Gamma(\alpha)(1+\delta)^\beta}\theta^{-k+\varepsilon_n}\Big)^{a_k}\\
&\qquad\times\Big(1+\mathcal{O}\Big(\frac{\log\log n}{\log n}\vee\frac{|i-\delta\log_\theta n|\vee|i'-\delta\log_\theta n|}{\log n}\Big)\Big),
\end{aligned}$$

where the last step follows from a similar argument as in the proof of Proposition 5.6. $\square$

With Proposition 7.2 at hand, the proofs of Theorems 4.2, 4.3 and 4.4 follow the same approach as the proofs of Theorems 2.5, 2.8 and 2.9, respectively.

7.2. **Example 4.5, fraction of 'gamma-like' random variables.** Let the distribution of W satisfy (4.2) for some $\alpha,\beta>0, w^*\in(0,1)$. We prove a result on (the tail of) the limiting degree distribution and provide additional building blocks required to prove the results in Example 4.5.

Lemma 7.3. *Let the distribution of W satisfy (4.2) and recall $p_k, p_{\geq k}, C$ from (2.1) and (4.3), respectively. Then,*

$$\begin{aligned}
p_k&=Z_{w^*}(1-\theta^{-1})Ck^{b/2+1/4}\mathrm{e}^{-2\sqrt{c_1^{-1}(1-\theta^{-1})k}}\theta^{-k}\big(1+\mathcal{O}(1/\sqrt{k})\big),\\
p_{\geq k}&=Z_{w^*}Ck^{b/2+1/4}\mathrm{e}^{-2\sqrt{c_1^{-1}(1-\theta^{-1})k}}\theta^{-k}\big(1+\mathcal{O}(1/\sqrt{k})\big).
\end{aligned}$$

Note that this Lemma improves on the bounds in (5.4) by providing a polynomial correction term and a precise multiplicative constant.

Proof. We start by proving the equality for $p_{\geq k}$ and then show the similar result for p_k. By (4.2), we obtain the following expression for $p_{\geq k}$.

$$\begin{aligned}
p_{\geq k}=&\int_{w^*}^1 x^k(\theta-1+x)^{-k}Z_{w^*}c_1^{-1}(1-x)^{-(2+b)}\mathrm{e}^{-c_1^{-1}x/(1-x)}\,\mathrm{d}x\\
&-\int_{w^*}^1 x^k(\theta-1+x)^{-k}Z_{w^*}b(1-x)^{-(1+b)}\mathrm{e}^{-c_1^{-1}x/(1-x)}\,\mathrm{d}x.
\end{aligned}\tag{7.7}$$

The second integral is of a similar form as the first, with a different constant in front and a different polynomial exponent. We hence only provide an explicit analysis of the first integral. As in the proof of Lemma 7.1, we start by considering the integral

$$\int_0^1 x^k(\theta-1+x)^{-k}c_1^{-1}(1-x)^{-(2+b)}\mathrm{e}^{-c_1^{-1}x/(1-x)}\,\mathrm{d}x,\tag{7.8}$$

where we omit the constant Z_{w^*} for ease of writing, and then show that the integral from zero up to w^* is asymptotically negligible compared to the integral from zero to one. Using a variable transform $u=x/(1-x)$, we find that (7.8) equals

$$\theta^{-k}c_1^{-1}\int_0^\infty u^k(1+u)^{b-k}\Big(1-\frac{1}{\theta(1+u)}\Big)^{-k}\mathrm{e}^{-c_1^{-1}u}\,\mathrm{d}u.$$

We now define X_u to be a negative binomial random variable with parameters k and $p_u:=(\theta(1+u))^{-1}$, for any $u>0$. As the sum over the probability mass function of X_u is one irrespectively of the value of u, we obtain that the above equals

$$\begin{aligned}
&\theta^{-k}c_1^{-1}\int_0^\infty \sum_{j=0}^\infty \binom{j+k-1}{j} p_u^j(1-p_u)^k u^k(1+u)^{b-k}\Big(1-\frac{1}{\theta(1+u)}\Big)^{-k}\mathrm{e}^{-c_1^{-1}u}\,\mathrm{d}u\\
&\quad=\theta^{-k}c_1^{-1}\int_0^\infty \sum_{j=0}^\infty \binom{j+k-1}{j}\theta^{-j}u^k(1+u)^{b-(j+k)}\mathrm{e}^{-c_1^{-1}u}\,\mathrm{d}u\\
&\quad=\theta^{-k}c_1^{-1}\sum_{j=0}^\infty \binom{j+k-1}{j}\theta^{-j}\Gamma(k+1)\,U(k+1,2+b-j,c_1^{-1}),
\end{aligned}$$

where $U(a,b,z)$ is the confluent hypergeometric function of the second kind, defined as

$$U(a,b,z):=\frac{1}{\Gamma(a)}\int_0^\infty x^{a-1}(1+x)^{b-a-1}\mathrm{e}^{-zx}\,\mathrm{d}x,$$

whenever $\mathrm{Re}(a)>0$. Using the Kummer transform $U(a,b,z)=z^{1-b}U(1+a-b,2-b,z)$ yields

$$\theta^{-k}c_1^{-1}\sum_{j=0}^\infty \binom{j+k-1}{j}\theta^{-j}\Gamma(k+1)\,c_1^{b-(j-1)}U(j+k-b,j-b,c_1^{-1}).$$

Again using the definition of the confluent hypergeometric function of the second kind, we obtain

$$\begin{aligned}
&c_1^b\theta^{-k}\sum_{j=0}^\infty \frac{\Gamma(j+k)\Gamma(k+1)}{\Gamma(k)\Gamma(j+1)\Gamma(j+k-b)}(c_1\theta)^{-j}\int_0^\infty u^{j+k-b-1}(1+u)^{-(k+1)}\mathrm{e}^{-c_1^{-1}u}\,\mathrm{d}u\\
&\quad=c_1^bk\theta^{-k}\frac{\Gamma(k)}{\Gamma(k-b)}\sum_{j=0}^\infty \frac{\Gamma(j+k)\Gamma(k-b)}{\Gamma(j+k-b)\Gamma(k)}\frac{1}{j!}(c_1\theta)^{-j}\int_0^\infty u^{j+k-b-1}(1+u)^{-(k+1)}\mathrm{e}^{-c_1^{-1}u}\,\mathrm{d}u\\
&\quad=c_1^bk\theta^{-k}\frac{\Gamma(k)}{\Gamma(k-b)}\sum_{j=0}^\infty \frac{(k)^{(j)}}{(k-b)^{(j)}}\frac{1}{j!}(c_1\theta)^{-j}\int_0^\infty u^{j+k-b-1}(1+u)^{-(k+1)}\mathrm{e}^{-c_1^{-1}u}\,\mathrm{d}u,
\end{aligned}$$

where $(x)^{(j)}:=x(x+1)\cdots(x+(j-1))$, $x\in\mathbb{R}, j\in\mathbb{N}_0$. When can then bound

$$\frac{(k)^{(j)}}{(k-b)^{(j)}}\geq\begin{cases}1 & \text{if } k>b\geq 0,\\ \Big(\frac{k}{k-b}\Big)^j & \text{if } b<0.\end{cases}\quad\text{and}\quad \frac{(k)^{(j)}}{(k-b)^{(j)}}\leq\begin{cases}1 & \text{if } b<0,\\ \Big(\frac{k}{k-b}\Big)^j & \text{if } k>b\geq 0.\end{cases}\tag{7.9}$$

As the bounds are symmetric, we can assume that $b\geq 0$ without loss of generality; the other case follows similarly. We deal with the lower bound first. This yields

$$\begin{aligned}
&c_1^bk\theta^{-k}\frac{\Gamma(k)}{\Gamma(k-b)}\sum_{j=0}^\infty \frac{1}{j!}(c_1\theta)^{-j}\int_0^\infty u^{j+k-b-1}(1+u)^{-(k+1)}\mathrm{e}^{-c_1^{-1}u}\,\mathrm{d}u\\
&\quad=c_1^bk\theta^{-k}\frac{\Gamma(k)}{\Gamma(k-b)}\int_0^\infty u^{k-b-1}(1+u)^{-(k+1)}\mathrm{e}^{-c_1^{-1}(1-\theta^{-1})u}\,\mathrm{d}u\\
&\quad=c_1^bk\theta^{-k}\frac{\Gamma(k)}{\Gamma(k-b)}\Gamma(k-b)U(k-b,-b,c_1^{-1}(1-\theta^{-1})).
\end{aligned}\tag{7.10}$$

It follows from [18, (3.12) and (3.15)] that, when $a>d/2$ is large and d,z are bounded,

$$\Gamma(a)U(a,d,z^2)=2\mathrm{e}^{z^2/2}\Big(\frac{2z}{u}\Big)^{1-d}K_{1-d}(uz)\big(1+\mathcal{O}(1/u)\big),$$

where $u=2\sqrt{a-d/2}$ and $K_{1-d}(uz)$ is a modified Bessel function. Combined with the asymptotic expression for the modified Bessel function as in [14, (10.40.2)], we obtain

$$\Gamma(a)U(a,d,z^2)=2\sqrt{\frac{\pi}{2uz}}\mathrm{e}^{z^2/2-uz}\Big(\frac{2z}{u}\Big)^{1-d}\big(1+\mathcal{O}(1/u)\big).\tag{7.11}$$

In this particular case, it yields

$$\begin{aligned}&\Gamma(k-b)U(k-b,-b,c_1^{-1}(1-\theta^{-1}))\\&\qquad=\mathrm{e}^{c_1^{-1}(1-\theta^{-1})/2}\sqrt{\pi}(c_1^{-1}(1-\theta^{-1}))^{1/4+b/2}k^{-b/2-3/4}\mathrm{e}^{-2\sqrt{c_1^{-1}(1-\theta^{-1})k}}\big(1+\mathcal{O}\big(1/\sqrt{k}\big)\big).\end{aligned}$$

Using this, as well as $\Gamma(k)/\Gamma(k-b)=k^b(1+\mathcal{O}(1/k))$, in (7.10), we arrive at

$$\mathrm{e}^{c_1^{-1}(1-\theta^{-1})/2}\sqrt{\pi}c_1^{-1/4+b/2}((1-\theta^{-1})k)^{1/4+b/2}\mathrm{e}^{-2\sqrt{c_1^{-1}(1-\theta^{-1})k}}\theta^{-k}\big(1+\mathcal{O}\big(1/\sqrt{k}\big)\big).$$

We then tend to the upper bound in (7.9) for $b\geq 0$, which yields

$$\begin{aligned}&c_1^bk\theta^{-k}\frac{\Gamma(k)}{\Gamma(k-b)}\sum_{j=0}^\infty\frac{1}{j!}\Big(\frac{(c_1\theta)^{-1}k}{k-b}\Big)^j\int_0^\infty u^{j+k-b-1}(1+u)^{-(k+1)}\mathrm{e}^{-c_1^{-1}u}\,\mathrm{d}u\\&\quad=c_1^bk\theta^{-k}\frac{\Gamma(k)}{\Gamma(k-b)}\int_0^\infty u^{k-b-1}(1+u)^{-(k+1)}\mathrm{e}^{-(c_1^{-1}(1-\theta^{-1})-(c_1\theta)^{-1}b/(k-b))u}\,\mathrm{d}u\\&\quad=c_1^bk\theta^{-k}\frac{\Gamma(k)}{\Gamma(k-b)}\Gamma(k-b)U(k-b,-b,c_1^{-1}(1-\theta^{-1})-(c_1\theta)^{-1}b/(k-b)).\end{aligned}$$

From the asymptotic results in (7.11) we find that

$$U\Big(k-b,-b,\frac{1}{c_1}\big(1-\theta^{-1}\big)-\frac{1}{c_1\theta}\frac{b}{k-b}\Big)=U\Big(k-b,-b,\frac{1}{c_1}\big(1-\theta^{-1}\big)\Big)\big(1+\mathcal{O}(1/\sqrt{k})\big),$$

so that the lower and upper bound match up to error terms (of the same order). By (7.11), we thus arrive at

$$\begin{aligned}&\int_0^1 x^k(\theta-1+x)^{-k}Z_{w^*}c_1^{-1}(1-x)^{-(2+b)}\mathrm{e}^{-c_1^{-1}x/(1-x)}\,\mathrm{d}x\\&\quad=Z_{w^*}\mathrm{e}^{c_1^{-1}(1-\theta^{-1})/2}\sqrt{\pi}c_1^{-1/4+b/2}((1-\theta^{-1})k)^{1/4+b/2}\mathrm{e}^{-2\sqrt{c_1^{-1}(1-\theta^{-1})k}}\theta^{-k}\big(1+\mathcal{O}\big(1/\sqrt{k}\big)\big)\\&\quad=Z_{w^*}Ck^{1/4+b/2}\mathrm{e}^{-2\sqrt{c_1^{-1}(1-\theta^{-1})k}}\theta^{-k}\big(1+\mathcal{O}\big(1/\sqrt{k}\big)\big).\end{aligned}$$

Then, we bound

$$\begin{aligned}&\int_0^{w^*}x^k(\theta-1+x)^{-k}Z_{w^*}c_1^{-1}(1-x)^{-(2+b)}\mathrm{e}^{-c_1^{-1}x/(1-x)}\,\mathrm{d}x\\&\qquad\leq\frac{Z_{w^*}}{c_1}(1-w^*)^{-((2+b)\vee 0)}\int_0^{w^*}x^k(\theta-1+x)^{-k}\,\mathrm{d}x\\&\qquad\leq\frac{Z_{w^*}}{c_1}(1-w^*)^{-((2+b)\vee 0)}w^*\Big(\frac{w^*}{\theta-1+w^*}\Big)^k,\end{aligned}$$

where we use that $x\mapsto(x/(\theta-1+x))^k$ is increasing in x. This also implies that $w^*/(\theta-1+w^*)<1/\theta$, so that

$$\int_0^{w^*}x^k(\theta-1+x)^{-k}Z_{w^*}c_1^{-1}(1-x)^{-(2+b)}\mathrm{e}^{-c_1^{-1}x/(1-x)}\,\mathrm{d}x=o\big(\theta^{-k}\mathrm{e}^{-2\sqrt{c_1^{-1}(1-\theta^{-1})}}k^{1/4+b/2}\big),$$

independent of the values of c_1,b and θ. As a result,

$$\begin{aligned}&\int_{w^*}^1 x^k(\theta-1+x)^{-k}Z_{w^*}c_1^{-1}(1-x)^{-(2+b)}\mathrm{e}^{-c_1^{-1}x/(1-x)}\,\mathrm{d}x\\&\quad=Z_{w^*}Ck^{1/4+b/2}\mathrm{e}^{-2\sqrt{c_1^{-1}(1-\theta^{-1})k}}\theta^{-k}\big(1+\mathcal{O}\big(1/\sqrt{k}\big)\big).\end{aligned}\tag{7.12}$$

Finally, when considering the second integral in (7.7), we observe its integrand is similar to that of the first integral but with a different constant in front and with a constant $\widetilde{b}=b-1$ in the polynomial exponent. We can thus follow the exact same steps as for the first integral in (7.7) to conclude that it can be included in the $\mathcal{O}(1/\sqrt{k})$ term in (7.12). In total,

$$p_{\geq k}=Z_{w^*}Ck^{1/4+b/2}\mathrm{e}^{-2\sqrt{c_1^{-1}(1-\theta^{-1})k}}\theta^{-k}\big(1+\mathcal{O}\big(1/\sqrt{k}\big)\big),$$

as required.

We now show the result for p_k, which uses the above steps with several minor adjustments. First,

$$\begin{aligned} p_k = Z_{w^*}(\theta-1)\int_{w^*}^1 x^k(\theta-1+x)^{-(k+1)}c_1^{-1}(1-x)^{-(2+b)}\mathrm{e}^{-c_1^{-1}x/(1-x)}\,\mathrm{d}x \\ - Z_{w^*}(\theta-1)\int_{w^*}^1 x^k(\theta-1+x)^{-(k+1)}b(1-x)^{-(b+1)}\mathrm{e}^{-c_1^{-1}x/(1-x)}\,\mathrm{d}x. \end{aligned} \tag{7.13}$$

As for the proof of the asymptotic expression of $p_{\geq k}$, we consider the first integral only as the second one is of lower order. Moreover, we again consider the first integral with a lower bound of zero for the integration variable x. So, omitting Z_{w^*} for now, we have

$$\begin{aligned} &(\theta-1)\int_0^1 x^k(\theta-1+x)^{-(k+1)}c_1^{-1}(1-x)^{-(2+b)}\mathrm{e}^{c_1^{-1}x/(1-x)}\,\mathrm{d}x \\ &\quad= (1-\theta^{-1})c_1^{-1}\theta^{-k}\int_0^\infty u^k(1+u)^{b-k}\Big(1-\frac{1}{\theta(1+u)}\Big)^{-(k+1)}\mathrm{e}^{-c_1^{-1}u}\,\mathrm{d}u \\ &\quad= (1-\theta^{-1})c_1^{-1}\theta^{-k}\sum_{j=0}^\infty\binom{j+k}{j}\theta^{-j}\int_0^\infty u^k(1+u)^{b-(j+k)}\mathrm{e}^{-c_1^{-1}u}\,\mathrm{d}u \\ &\quad= (1-\theta^{-1})c_1^{-1}\theta^{-k}\sum_{j=0}^\infty\binom{j+k}{j}\theta^{-j}\Gamma(k+1)U(k+1,2+b-j,c_1^{-1}) \\ &\quad= (1-\theta^{-1})c_1^{b}\theta^{-k}\sum_{j=0}^\infty\binom{j+k}{j}(c_1\theta)^{-j}\Gamma(k+1)U(k+j-b,j-b,c_1^{-1}) \\ &\quad= (1-\theta^{-1})c_1^{b}\theta^{-k}\frac{\Gamma(k+1)}{\Gamma(k-b)}\sum_{j=0}^\infty\frac{(k+1)^{(j)}}{(k-b)^{(j)}}\frac{(c_1\theta)^{-j}}{j!}\int_0^\infty u^{k+j-b-1}(1+u)^{-(k+1)}\mathrm{e}^{-c_1^{-1}u}\,\mathrm{d}u. \end{aligned}$$

Similar to (7.9), we bound

$$\frac{(k-1)^{(j)}}{(k-b)^{(j)}}\geq\begin{cases}1 & \text{if } k>b\geq-1,\\ \left(\frac{k+1}{k-b}\right)^j & \text{if } b<-1.\end{cases}\qquad\text{and}\qquad \frac{(k)^{(j)}}{(k-b)^{(j)}}\leq\begin{cases}1 & \text{if } b<-1,\\ \left(\frac{k+1}{k-b}\right)^j & \text{if } k>b\geq-1.\end{cases}$$

Again, we assume without loss of generality that $b\geq -1$. Moreover, we only concern ourselves with the lower bound on $(k+1)^{(j)}/(k-b)^{(j)}$ when $b\geq-1$, since we obtain a matching upper bound with the required error term when using the upper bound on $(k+1)^{(j)}/(k-b)^{(j)}$ when $b\geq-1$, as in the proof for $p_{\geq k}$. Thus, we obtain the lower bound

$$\begin{aligned} &(1-\theta^{-1})c_1^b\theta^{-k}\frac{\Gamma(k+1)}{\Gamma(k-b)}\sum_{j=0}^\infty\frac{(c_1\theta)^{-j}}{j!}\int_0^\infty u^{k+j-b-1}(1+u)^{-(k+1)}\mathrm{e}^{-c_1^{-1}u}\,\mathrm{d}u \\ &\qquad= (1-\theta^{-1})c_1^b k\theta^{-k}\frac{\Gamma(k)}{\Gamma(k-b)}\int_0^\infty u^{k+j-b-1}(1+u)^{-(k+1)}\mathrm{e}^{-c_1^{-1}(1-\theta^{-1})u}\,\mathrm{d}u, \end{aligned}$$

which, up to the constant $(1-\theta^{-1})$, is the exact same expression as in (7.10). As discussed above, using the upper bound on $(k+1)^{(j)}/(k-b)^{(j)}$ yields a matching upper bound (up to error terms). Then, the same approach as in the proof of $p_{\geq k}$ can be used to show that

$$\int_0^{w^*}x^k(\theta-1+x)^{-(k+1)}c_1^{-1}(1-x)^{b-2}\mathrm{e}^{c_1^{-1}x/(1-x)}\,\mathrm{d}x = o\big(\theta^{-k}\mathrm{e}^{2\sqrt{c_1^{-1}(1-\theta^{-1})k}}k^{-1/4+b/2}\big),$$

so that the integral from w^* to 1 is asymptotically equivalent to the integral from 0 to 1. Following the same steps as above for the second integral in (7.13), we find it can be included in the error term as well. Hence, the result follows. □

Recall that in this example we set

$$\begin{aligned}X_i^{(n)} &:= \big|\big\{j\in[n] : \mathcal{Z}_n(j) = \big\lfloor \log_\theta n - C_{\theta,1,c_1}\sqrt{\log_\theta n} + (b/2+1/4)\log_\theta\log_\theta n\big\rfloor + i\big\}\big|,\\ X_{\geq i}^{(n)} &:= \big|\big\{j\in[n] : \mathcal{Z}_n(j) \geq \big\lfloor \log_\theta n - C_{\theta,1,c_1}\sqrt{\log_\theta n} + (b/2+1/4)\log_\theta\log_\theta n\big\rfloor + i\big\}\big|,\\ \varepsilon_n &:= \big(\log_\theta n - C_{\theta,1,c_1}\sqrt{\log_\theta n} + (b/2+1/4)\log_\theta\log_\theta n\big)\\ &\qquad - \big\lfloor \log_\theta n - C_{\theta,1,c_1}\sqrt{\log_\theta n} + (b/2+1/4)\log_\theta\log_\theta n\big\rfloor.\end{aligned}$$

We then state the analogue of Proposition 5.6.

Proposition 7.4. *Consider the WRT model as in Definition 2.1 with vertex-weights $(W_i)_{i\in[n]}$ whose distribution satisfies (4.2) for some $b\in\mathbb{R}, c_1>0, w^*\in(0,1)$. For a fixed $K\in\mathbb{N}, c\in(1,\theta/(\theta-1))$ the following holds. For any $i,i'=i(n),i'(n)$ in $\mathbb{Z}$ such that $0<\log_\theta n+i<\log_\theta n+i'<c\log_\theta n$ and $i,i'\sim\delta\sqrt{\log_\theta n}$ for some $\delta\in\mathbb{R}$ ($\delta=0$ denotes $i,i'=o(\sqrt{\log_\theta n})$) and for $a_i,\dots,a_{i'}\in\mathbb{N}_0$ satisfying $a_i+\dots+a_{i'}=K$,*

$$\begin{aligned}&\mathbb{E}\Big[\Big(X_{\geq i'}^{(n_\ell)}\Big)_{a_{i'}}\prod_{k=i}^{i'-1}\Big(X_k^{(n_\ell)}\Big)_{a_k}\Big]\\ &= \Big(\frac{\widetilde{Z}}{\theta-1}\theta^{-i'+1+\varepsilon_n+C_{\theta,1,c_1}(C_{\theta,1,c_1}-\delta)/2}\Big)^{a_{i'}}\prod_{k=i}^{i'-1}\Big(\widetilde{Z}\theta^{-k+\varepsilon_n+C_{\theta,1,c_1}(C_{\theta,1,c_1}-\delta)/2}\Big)^{a_k}\\ &\quad\times\Big(1+\mathcal{O}\Big(\frac{\log_\theta\log_\theta n}{\sqrt{\log_\theta n}}\vee\frac{|i-\sqrt{\log_\theta n}|\vee|i'-\sqrt{\log_\theta n}|}{\sqrt{\log_\theta n}}\Big)\Big),\end{aligned}$$

with $\widetilde{Z}:=Z_{w^*}C$.

Proof. Set $K':=K-a_{i'}$ and for each $i\leq k\leq i'$ and for each u such that $\sum_{\ell=i}^{k-1}a_\ell<u\leq\sum_{\ell=i}^{k}a_\ell$, let $m_u=\big\lfloor\log_\theta n-C_{\theta,1,c_1}\sqrt{\log_\theta n}+(b/2+1/4)\log_\theta\log_\theta n\big\rfloor+k$. Also, let $(v_u)_{u\in[K]}$ be K vertices selected uniformly at random without replacement from $[n]$. Then, as the $X_{\geq k}^{(n)}$ and $X_k^{(n)}$ can be expressed as sums of indicators, using the same steps as in the proof of Proposition 5.6,

$$\mathbb{E}\left[\Big(X_{\geq i'}^{(n)}\Big)_{a_{i'}}\prod_{k=i}^{i'-1}\Big(X_k^{(n)}\Big)_{a_k}\right]=(n)_K\sum_{\ell=0}^{K'}\sum_{\substack{S\subseteq[K']\\|S|=\ell}}(-1)^\ell\mathbb{P}\big(\deg(v_u)\geq m_u+\mathbb{1}_{\{u\in S\}},u\in[K]\big).$$

By Proposition 5.1,

$$\mathbb{P}\big(\deg(v_\ell)\geq m_u+\mathbb{1}_{\{u\in S\}},u\in[K]\big)=\prod_{u=1}^K\mathbb{E}\left[\Big(\frac{W}{\mathbb{E}[W]+W}\Big)^{m_u+\mathbb{1}_{\{u\in S\}}}\right](1+o(n^{-\beta})),$$

for some $\beta>0$. By Lemma 7.3 (and recalling the constant C in (4.3)), when $|S|=\ell$,

$$\begin{aligned}&\prod_{u=1}^K\mathbb{E}\left[\Big(\frac{W}{\mathbb{E}[W]+W}\Big)^{m_u+\mathbb{1}_{\{u\in S\}}}\right]\\ &=\widetilde{Z}^K\theta^{-\sum_{u=1}^K m_u-\ell}\exp\Big\{-2\sum_{u=1}^K\sqrt{\frac{1-\theta^{-1}}{c_1}\big(m_u+\mathbb{1}_{\{u\in S\}}\big)}\Big\}\prod_{u=1}^K(m_u+\mathbb{1}_{\{u\in S\}})^{b/2+1/4}\\ &\quad\times(1+\mathcal{O}(1/\sqrt{\log n})).\end{aligned}$$

Here, we are able to obtain the error term $1+\mathcal{O}(1/\sqrt{\log n})$ due to the fact that $\log_\theta n+i>\eta\log n$ for some $\eta>0$ when n is large. We note that $C_{\theta,1,c_1}\log\theta=2\sqrt{c_1^{-1}(1-\theta^{-1})}$. As

$i, i' \sim \delta\sqrt{\log_\theta n}$,

$$\prod_{u=1}^K (m_u + \mathbb{1}_{\{u\in S\}})^{b/2+1/4} = (\log_\theta n)^{K(b/2+1/4)}\big(1+\mathcal{O}\big(1/\sqrt{\log_\theta n}\big)\big),$$

uniformly in S (and ℓ). Moreover, again uniform in S and ℓ,

$$\begin{aligned}\exp\Big\{ &- C_{\theta,1,c_1}\log\theta\sum_{u=1}^K\sqrt{m_u+\mathbb{1}_{\{u\in S\}}}\Big\}\\ &= \exp\Big\{-\Big(C_{\theta,1,c_1}\log\theta\sqrt{\log_\theta n}-\frac{C_{\theta,1,c_1}-\delta}{2}\Big)\Big\}^K\\ &\quad\times\Big(1+\mathcal{O}\Big(\frac{\log_\theta\log_\theta n}{\sqrt{\log_\theta n}}\vee\frac{|i-\sqrt{\log_\theta n}|\vee|i'-\sqrt{\log_\theta n}|}{\sqrt{\log_\theta n}}\Big)\Big).\end{aligned}$$

This last step follows from the fact that the first-order term of m_u is $\log_\theta n$ and its second-order term is $-(C_{\theta,1,c_1}-\delta)\sqrt{\log_\theta n}$. Finally, its lower-order terms are $\log_\theta\log_\theta n + (|i-\sqrt{\log_\theta n}|\vee|i'-\sqrt{\log_\theta n}|)$. Then using a Taylor expansion for the square root yields the result. Combining all of the above, we thus arrive at

$$\begin{aligned}(n)_K&\sum_{\ell=0}^{K'}\sum_{\substack{S\subseteq[K']\\|S|=\ell}}(-1)^\ell\Big(\widetilde{Z}(\log_\theta n)^{b/2+1/4}\exp\Big\{-\Big(C_{\theta,1,c_1}\log\theta\Big(\sqrt{\log_\theta n}-\frac{C_{\theta,1,c_1}-\delta}{2}\Big)\Big)\Big\}\Big)^K\\ &\times\theta^{-\sum_{u=1}^K m_u-\ell}\Big(1+\mathcal{O}\Big(\frac{\log_\theta\log_\theta n}{\sqrt{\log_\theta n}}\vee\frac{|i-\sqrt{\log_\theta n}|\vee|i'-\sqrt{\log_\theta n}|}{\sqrt{\log_\theta n}}\Big)\Big)\\ =&\Big(\frac{\widetilde{Z}}{\theta-1}\theta^{-i'+1+\varepsilon_n+C_{\theta,1,c_1}(C_{\theta,1,c_1}-\delta)/2}\Big)^{a_{i'}}\prod_{k=i}^{i'-1}\Big(\widetilde{Z}\theta^{-k+\varepsilon_n+C_{\theta,1,c_1}(C_{\theta,1,c_1}-\delta)/2}\Big)^{a_k}\\ &\times\Big(1+\mathcal{O}\Big(\frac{\log_\theta\log_\theta n}{\sqrt{\log_\theta n}}\vee\frac{|i-\sqrt{\log_\theta n}|\vee|i'-\sqrt{\log_\theta n}|}{\sqrt{\log_\theta n}}\Big)\Big),\end{aligned}$$

where the last step follows from a similar argument as in the proof of Proposition 5.6. □

With Proposition 7.4 at hand, the proofs of Theorems 4.6, 4.7 and 4.8 follow the same approach as the proofs of Theorems 2.5, 2.8 and 2.9.

References

[1] L. Addario-Berry and L. Eslava. High degrees in random recursive trees. *Random Structures Algorithms*, 52(4):560–575, 2018.

[2] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular variation*, volume 27 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1989.

[3] B. Bollobás. *Random graphs*. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], London, 1985.

[4] K. A. Borovkov and V. A. Vatutin. On the asymptotic behaviour of random recursive trees in random environments. *Adv. in Appl. Probab.*, 38(4):1047–1070, 2006.

[5] L. Devroye and J. Lu. The strong convergence of maximal degrees in uniform random recursive trees and dags. *Random Structures Algorithms*, 7(1):1–14, 1995.

[6] L. Eslava. Depth of vertices with high degree in random recursive trees. *arXiv preprint arXiv:1611.07466*, 2016.

[7] N. Fountoulakis and T. Iyer. Condensation phenomena in preferential attachment trees with neighbourhood influence. *arXiv preprint arXiv:2101.02734*, 2021.

[8] N. Fountoulakis, T. Iyer, C. Mailler, and H. Sulzbach. Dynamical models for random simplicial complexes. *arXiv preprint arXiv:1910.12715*, 2019.

[9] W. Goh and E. Schmutz. Limit distribution for the maximum degree of a random recursive tree. *J. Comput. Appl. Math.*, 142(1):61–82, 2002. Probabilistic methods in combinatorics and combinatorial optimization.

[10] E. Hiesmayr and U. Işlak. Asymptotic results on Hoppe trees and their variations. *J. Appl. Probab.*, 57(2):441–457, 2020.
[11] T. Iyer. Degree distributions in recursive trees with fitnesses. *Preprint, arXiv:2005.02197*, 2020.
[12] S. Janson, T. Łuczak, and A. Rucinski. *Random graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000.
[13] B. Lodewijks and M. Ortgiese. The maximal degree in random recursive graphs with random weights. *arXiv preprint arXiv:2007.05438*, 2020.
[14] D. W. Lozier. NIST Digital Library of Mathematical Functions. *Ann. Math. Artif. Intell.*, 38(1-3):105–119, 2003. Mathematical knowledge management.
[15] C. Mailler and G. Uribe Bravo. Random walks with preferential relocations and fading memory: a study through random recursive trees. *J. Stat. Mech. Theory Exp.*, (9):093206, 49, 2019.
[16] M. Pain and D. Sénizergues. Correction terms for the height of weighted recursive trees. *arXiv preprint arXiv:2101.01156*, 2021.
[17] D. Sénizergues. Geometry of weighted recursive and affine preferential attachment trees. *Preprint, arXiv:1904.07115*, 2019.
[18] N. M. Temme. Remarks on Slater's asymptotic expansions of Kummer functions for large values of the $a-$parameter. *arXiv preprint arXiv:1306.5328*, 2013.

8. Appendix

Lemma 8.1. *Fix $\ell, n \in \mathbb{N}_0$ such that $\ell < n$. Suppose $f : \mathbb{R} \to \mathbb{R}$ is increasing on $[\ell, n+1]$, and $g : \mathbb{R} \to \mathbb{R}$ decreasing on $[\ell-1, n]$ and both f and g are positive and integrable. Then,*

$$\int_{\ell-1}^{n} f(x)g(x)\,\mathrm{d}x - 2f(n)g(\ell-1) \leq \sum_{k=\ell}^{n} f(k)g(k) \leq \int_{\ell-1}^{n} f(x)g(x)\,\mathrm{d}x + 2f(n+1)g(\ell-1).$$

Proof. We only prove the upper bound, the lower bound follows from an analogous approach. By definition for $k \in \{\ell, \dots, n\}$ and $x \in [k-1, k]$,

$$f(k) \leq f(x+1) \quad \text{and} \quad g(k) \leq g(x).$$

Integrating both sides gives

$$\begin{aligned} f(k)g(k) &\leq \int_{k-1}^{k} f(x+1)g(x)\,\mathrm{d}x \\ &\leq \int_{k-1}^{k} f(x)g(x)\,\mathrm{d}x + g(\ell-1)\int_{k-1}^{k} f(x+1) - f(x)\,\mathrm{d}x \\ &\leq \int_{k-1}^{k} f(x)g(x)\,\mathrm{d}x + g(\ell-1)(f(k+1) - f(k-1)). \end{aligned}$$

Hence, summing from ℓ to n gives

$$\begin{aligned} \sum_{k=\ell}^{n} f(k)g(k) &\leq \int_{\ell-1}^{n} f(x)g(x)\,\mathrm{d}x + g(\ell-1)\sum_{k=\ell}^{n}(f(k+1) - f(k-1)) \\ &\leq \int_{\ell-1}^{n} f(x)g(x)\,\mathrm{d}x + 2f(n+1)g(\ell-1), \end{aligned}$$

as required. □

Corollary 8.2. *Fix $\ell, n \in \mathbb{N}$ such that $\ell < n$. Suppose $f : \mathbb{R} \to \mathbb{R}$ is a positive integrable function, increasing on $[\ell, x^*]$ and decreasing on $[x^*, n]$, where x^* is not necessarily an integer. Suppose $g : \mathbb{R} \to [0,1]$ is a positive, integrable function, decreasing on $[\ell, n+1]$ and bounded by 1. Then,*

$$\int_{\ell}^{n} f(x)g(x)\,\mathrm{d}x - 4f(x^*) \leq \sum_{k=\ell+1}^{n} f(k)g(k) \leq \int_{\ell}^{n} f(x)g(x)\,\mathrm{d}x + 4f(x^*).$$

Proof. We only prove the upper bound, the lower bound follows from an analogous approach. Note that f is increasing on $[\ell, \lfloor x^* \rfloor]$, so that by Lemma 8.1

$$\sum_{k=\ell+1}^{\lfloor x^* \rfloor -1} f(k)g(k) \leq \int_{\ell}^{\lfloor x^* \rfloor -1} f(x)g(x)\,\mathrm{d}x + 2f(x^*),$$

where we used that g is decreasing and bounded by 1.

Also, note that by the fact that f and g are both decreasing on $[x^*, n]$,

$$\sum_{k=\lceil x^* \rceil +1}^{n} f(k)g(k) \leq \int_{\lceil x^* \rceil}^{n} f(x)g(x)\,\mathrm{d}x.$$

It remains to bound $f(\lfloor x^* \rfloor)g(\lfloor x^* \rfloor) + f(\lceil x^* \rceil)g(\lceil x^* \rceil)$. We use that f is maximised at x^* and that g is bounded by one to obtain the upper bound $2f(x^*)$. Combining all of the above and including in the integrals in the upper bounds the range $[\lfloor x^* \rfloor - 1, \lceil x^* \rceil]$, we thus obtain

$$\sum_{k=\ell+1}^{n} f(k)g(k) \leq \int_{\ell}^{n} f(x)g(x)\,\mathrm{d}x + 4f(x^*),$$

as required. □

Lemma 8.3. *Consider the sequences $(s_k, r_k)_{k\in\mathbb{N}}$ in (2.8). These sequences have the following properties:*

(i) s_k is increasing,
(ii) r_k is decreasing and $\lim_{k\to\infty} r_k = 0$.

Proof. (i) Assume that $s_{k+1} < s_k$ for some $k \in \mathbb{N}$ and take $x \in (s_{k+1}, s_k)$. By the definition of s_{k+1}, s_k and the choice of x,

$$\mathbb{P}(W \in (x,1)) \leq \mathrm{e}^{-(1-\theta^{-1})(1-x)(k+1)} < \mathrm{e}^{-(1-\theta^{-1})(1-x)k} < \mathbb{P}(W \in (x,1))\,,$$

which leads to a contradiction.

(ii) Assume that $s_k < s_{k+1}$ (otherwise the claim is immediately clear). Note that since the function $\mathbb{P}(W \in (x,1))$ is càdlàg, we have for any $x < s_k$,

$$\mathbb{P}(W \in (s_k,1)) \leq r_k \leq \lim_{y\uparrow s_k} \mathbb{P}(W \in (y,1)) \leq \mathbb{P}(W \in (x,1))\,.$$

Hence, we have that for any $x \in (s_k, s_{k+1})$.

$$r_k \geq \mathbb{P}(W \in (s_k,1)) \geq \mathbb{P}(W \in (x,1)) \geq r_{k+1}.$$

For the second part, since s_k is increasing by (i), we have that $s_k \to s^* \in (0,1]$. Suppose that $s^* \in (0,1)$. Then, for k sufficiently large, we have $s_k \leq (1+s^*)/2$ and so $r_k \leq \mathrm{e}^{-(1-\theta^{-1})(1-s^*)k/2}$ and so r_k converges to 0.

Therefore, we can assume that $s_k \uparrow 1$. Let k_0 be such the smallest k such that $s_k < s_{k+1}$. Such a k_0 exists, otherwise $s^* < 1$ since each $s_k < 1$. Then, for $k \geq k_0$, let ℓ_k be the largest integer such that $s_{\ell_k} < s_k$. The assumption that $s_k \uparrow 1$ also excludes that s_{ℓ_k} is eventually constant and so $\ell_k \to \infty$. In particular, we can argue as in (8) to see that

$$r_k \leq \mathbb{P}(W \in (s_{\ell_k}, 1))\,.$$

Moreover, as $s_{\ell_k} \to 1$ as $k \to \infty$, we deduce that $r_k \to 0$. □

[†]Universidad Nacional Autonoma Mexico, Instituto de investigaciones en matematicas y en sistemas, CDMX, 04510, Mexico

Email address: laura@sigma.iimas.unam.mx

[⋆]Department of Mathematical Sciences, University of Bath, Claverton Down, Bath, BA2 7AY, United Kingdom.

Email address: b.lodewijks@bath.ac.uk,m.ortgiese@bath.ac.uk

4.2 Conclusion

In this chapter we investigated the properties of the maximum degree of the weighted recursive tree when the weights are almost surely bounded. This extends the results provided in the previous chapter, where only a first-order result for bounded weights was provided. Using a different approach, extending the results related to the degree distribution presented in the previous chapter, we were able to extend the analysis of the maximum degree to finer asymptotics.

The main approach was to obtain precise asymptotics for number of vertices with degree exceeding k when k is allowed to grow with n. Combining this with asymptotics of the limiting degree distribution enabled us to establish a more precise understanding of the behaviour of the maximum degree.

We focussed on three main classes of vertex-weight distributions: (i) distributions with an atom at one (recall that we can without loss of generality assume that bounded weights have their essential supremum at one), (ii) distributions in the Weibull domain of attraction and (iii) distributions in the Gumbel domain of attraction.

In case (i) we observe behaviour similar to that observed for the random recursive tree, in the sense that the maximum degree is random to second order and that the limit can be described only along certain subsequences, as presented in Theorem 2.5. Moreover, this result leads to a precise understanding of the tail distribution of the maximum degree, as well as the number of 'near-maximum' degrees, which is asymptotically normal. These results are presented in Theorems 2.8 and 2.9, respectively.

In the second case we obtain a deterministic and negative second order correction term, as presented in Theorem 2.5. Compared to (i) this negative second order correction term can be explained by the fact that a much smaller proportion of vertices has a vertex-weight close to one, whereas a positive proportion of all vertices has a vertex-weight of exactly one in case (i). This implies that these vertices are able to acquire edges at a faster rate compared to case (ii). Hence, the maximum degree has a negative second order correction term. The same follows for case (iii), though the precise correction term depends on the assumptions on the distribution, as presented in Theorem 2.6, and for particular distributions even finer asymptotics can be obtained.

We also distinguished two examples of vertex-weight distributions that fall in the Weibull and Gumbel maximum domain of attraction for which we were able to extend the results presented in Section 2. In Theorems 4.1 through 4.4 we discussed the properties of the maximum degree and near-maximum degrees when the vertex-weights follow a beta distribution bounded away from zero, and in Theorem 4.6 through 4.8 we discussed the case when vertex-weights follow a particular distribution that can be interpreted as a fraction of 'gamma-like' random variables. In both cases, we obtained more precise correction terms up to random order, in which case the behaviour was similar to case (i). This begs the question whether such behaviour is universal for *all* vertex-weight distributions with finite support.

Chapter 5

Conclusion and open problems

In this thesis we have studied two types of models of evolving random graphs in random environment: preferential attachment models with additive fitness and weighted recursive graphs. The focus of the research has been to develop tools to analyse their degree distribution and high degrees and to distinguish their properties under influence of the random environment.

In Chapter 2 we studied preferential attachment models with additive fitness and presented a phase transition for the behaviour of the degree distribution and maximum degree. We used novel approaches to extend the results known so far related to the degree distribution and maximum degree. Our analysis provides an explanation as to why, when and how the properties of this model in random environment, the degree distribution and maximum degree specifically, differ significantly from its counterpart without random environment, and shows the random environment allows for richer behaviour to be observed.

Chapters 3 and 4 focussed on the degree distribution and high degrees in weighted recursive graphs and weighted recursive trees (the tree case of the model), respectively. Again, we were able to outline different conditions on the random environment for which different behaviour can be observed for the degree distribution and high degrees. In Chapter 3 we investigated the properties of weighted recursive graphs for a large range of vertex-weight distributions, and were able to specify the size of the maximum degree to first order, and in particular cases to second order as well. The location of the maximum degree has been analysed for distributions with unbounded support. Chapter 4 provided a more in-depth look into the particular case of distributions with bounded support. Here, we were able to obtain more precise and refined asymptotics for high degrees by developing improved results related to the degree distribution, which we then translated to high degrees. Again, these chapters clearly show how the presence of a random environment allows for a wider range of behaviour to be observed compared to the random recursive tree, the counterpart of the weighted recursive tree without random environment.

Beyond the influence of the random environment on the properties of the models studied, it also provides a much more natural and realistic model of real-world networks, where the connectivity of vertices is allowed to develop differently over time. The presence of fitness or weights allows for more heterogeneity in real-world networks to be reflected in the models and therefore, on a conceptual level, is desirable over linear of affine preferential attachment models and the random recursive tree and directed acyclic graph models. On top of that, the additive nature of the fitness in the preferential attachment models in Chapter 2 allows the models to remain tractable, whereas multiplicative preferential attachment models can be much more difficult to analyse to the same extent.

As is only natural, in search of answers to the questions described above during the PhD research carried out by the author, many more questions came to light that we were not able to address or answer yet. Here, we would like to discuss some of these open problems which form interesting possibilities for further research.

Preferential attachment models with additive fitness
The results presented in Theorem 2.7 in Chapter 2 describe several regimes in which significantly different behaviour can be observed for the preferential attachment models with additive fitness studied in that chapter. In particular, Equation (2.10) provides the almost sure limit of the rescaled maximum degree in the weak disorder regime, as the supremum $\sup_{i\geq 1}\xi_i$ of the individual limits of the rescaled degrees $(\xi_i)_{i\in\mathbb{N}}$ described in Equation (2.8). It would be interesting to see whether more information and properties can be extracted from these limits. As an example, Sénizergues is able to obtain explicit distributional identities for these limits for some particular choices of deterministic fitness sequences in [128]: whenever the fitness values are of the form $(\mathcal{F}_1,\mathcal{F}_2,\mathcal{F}_3,\ldots)=(a,b,b,\ldots)$ with $a>-1, b>0$, the limits ξ_i satisfy the recursion $\xi_i=\beta_i\xi_{i+1}$, where ξ_i is the almost sure limit of $\mathcal{Z}_n(i)/n^{1/\theta_m}$, $i\in\mathbb{N}$, and β_i is a $\mathrm{Beta}(\sum_{j=1}^i\mathcal{F}_j+j,\mathcal{F}_{j+1})$ random variable which is independent of ξ_{i+1}. This recursion is also known as the Mittag-Lefler Markov chain family introduced by Goldschmidt and Haas [64] and also studied by James [81] in the context of preferential attachment models. We observe that in the precise model description in [128] vertex two always connects to vertex one, independently of the fitness of vertex one, so that $\mathcal{Z}_n(1)\geq 1$ for all $n\geq 2$ and hence $\mathcal{Z}_n(2)+\mathcal{F}_1\geq 0$, so that despite a (possibly) negative fitness $\mathcal{F}_1$ the connection probabilities are still well-defined.

The other example provided by Sénizergues entails a periodic fitness sequence

$$(\mathcal{F}_1,\mathcal{F}_2,\ldots,\mathcal{F}_{\ell+1},\mathcal{F}_{\ell+2},\ldots,\mathcal{F}_{2\ell+1},\ldots)=(a,b_1,\ldots,b_\ell,b_1,\ldots,b_\ell,\ldots),$$

for any $a>-1,\ell\in\mathbb{N},b_1,\ldots,b_\ell\in\mathbb{N}$ (note that $\ell=1$ recovers the previous example for which $b\in\mathbb{N}$). In this case, the limits ξ_i can be described using an Intertwined Product of Generalised Gamma Processes, defined in [128].

In the case of random fitness values, we wonder if such precise distributional identities can be obtained for the limiting random variables ξ_i as well. Instead of the martingale techniques used in Chapter 2 to obtain the almost sure convergence $\mathcal{Z}_n(i)/n^{1/\theta_m}\xrightarrow{a.s.}\xi_i, i\in\mathbb{N}$ (Equation 2.8), other approaches using, for example, Pólya urns might be required, as this is what Sénizergues uses for the two examples described above.

The same holds for the limit of the rescaled maximum degree and the rescaled location of the maximum degree in the extreme disorder regime, in Equation (2.12) of Theorem 2.7. Though an explicit representation of the limit in terms of a function of a Poisson point process is presented, we do not know anything about its distribution or other properties. Even less is known about the limit of the rescaled location.

Next to identifying the limit, fluctuations around the limit of the maximum degree are also interesting to investigate. Móri shows in [112, Theorem 4.1] that the maximum degree exhibits a central limit theorem with a mixed normal distribution as the limit. That is, if we set $\mu:=\sup_{i\geq 1}\xi_i$ and $m=1$,

$$n^{1/(2\theta_1)}\big(\max_{i\in[n]}\mathcal{Z}_n(i)/n^{1/\theta_1}-\mu\big)\xrightarrow{d}\mathcal{N}\big(0,\sqrt{\mu}\big), \tag{5.0.1}$$

where $\mathcal{N}(0, \sqrt{\mu})$ is defined as $\sqrt{\mu}Z$ with Z a standard normal random variable independent of μ. Note that this results holds when $\mathcal{F}_i = \beta > 0$ almost surely for all $i \in \mathbb{N}$ only, i.e. for the affine preferential attachment model. A more general result is obtained by Sénizergues in [128, Proposition 32] for preferential attachment trees in with fitness sequences in the weak disorder regime. The martingale techniques used to analyse the weak disorder regime in Chapter 2 is based on the martingale techniques developed by Móri. The proof of (5.0.1) in [112] are based on the Doob-Meyer decomposition of the submartingale $\max_{i\in[n]} M_n^1(i)$ into a convergent martingale and a predictable increasing process, where we recall the martingales $M_n^1(i)$ in Chapter 2, Lemma 6.3, and note that a maximum of martingales is a submartingale. On the other hand, Sénizergues applies Pólya urn techniques to obtain the more general result. It is not clear whether such techniques can be adapted to obtain these results for the case in which $m > 1$, which does not yield a tree but a multi-graph.

Finally, in the phase transition described for the preferential attachment models with fitness in Chapter 2, the cases when $\alpha = 1 + \theta_m$ and $\alpha = 2$, no results are known for the behaviour of the degrees of fixed vertices and the maximum degree, and when $\alpha = 2$ no results are known for the degree distribution as well. When $\alpha = 1 + \theta_m$ we expect behaviour similar to what is described for the weak or strong disorder regime to be observed (as in Equations (2.8) and (2.10) or (2.8) and (2.11), respectively), and when $\alpha = 2$ we expect behaviour similar to what is described for the strong or extreme disorder regime to be observed (as in Equations (2.8) and (2.11) or (2.9) and (2.12), respectively), though possibly with a different rescaling due to the effect of the slowly-varying function of the fitness distribution, similarly to *(ii)* in Theorem 2.6. Which of the regimes describes the behaviour best, depends on the particular form of the slowly-varying function ℓ. Moreover, when $\alpha = 2$, we expect the limiting degree distribution $p(k)$ to behave as in case *(ii)* of Theorem 2.6 or for a result similar to what is described in case *(iii)* to be observed. That is, $p(k) = \Theta(\ell(k)k^{-2})$ or the proportion of leaves in the graph converges to one in mean, respectively.

The difficulty in establishing the behaviour of the degree distribution, degree evolution of fixed vertices and the maximum degree in the cases $\alpha = 1 + \theta_m$ and $\alpha = 2$ lies in the fact that (the proofs of) most preliminary results developed in Chapter 2 required to prove Theorems 2.4, 2.6 and 2.7 break down when either $\alpha = 1 + \theta_m$ or $\alpha = 2$. For example, Proposition 5.2 and Lemma 5.4 require that $\alpha < 2$, the results in Proposition 6.1, (6.1) and (6.2), require $\alpha < 1+\theta_m$ and $\alpha < 2$, respectively, the results in Proposition 6.2, (6.3) and (6.4), require $\alpha < 1+\theta_m$ and $\alpha < 2$, respectively, and Lemma 6.6 requires $\alpha > 1+\theta_m$. Also, the limiting random variables of the rescaled location of the maximum degree and the rescaled maximum degree are ill-defined when $\alpha = 1+\theta_m$. More careful and refined analysis is required to unveil this behaviour.

Beyond the properties of the degree distribution and maximum degree, one could wonder how the random environment influences other properties of these preferential attachment models, such as typical distances. Though we have a good understanding of typical distances in this model when the degree distribution has infinite variance and finite mean or when the fitness distribution has an infinite mean, the mathematics we developed was not complete yet and details need to be fleshed out. In the former case, when $\tau := (\alpha \wedge (1 + \theta_m)) \in (2, 3)$ we conjecture to observe double logarithmic typical distances with a constant that depends on the power-law exponent of the degree distribution, much like what is observed in affine preferential attachment models (see e.g. [73] for an overview). To be more precise, if $d_G(\cdot, \cdot)$ denotes the graph distance metric on a graph G, then for the sequence of random graphs $(\mathcal{G}_n)_{n\in\mathbb{N}}$ obtained via the PAFUD or PAFFD constructions with $m > 2$ or the PAFRO construction described in

Chapter 2,

$$\frac{d_{\mathcal{G}_n}(v_1^n, v_2^n)}{\log\log n} \xrightarrow{\mathbb{P}} \frac{1}{|\log(\tau-2)|}, \tag{5.0.2}$$

where v_1^n, v_2^n are two vertices selected uniformly at random from $[n]$ without replacement. We note that, most likely, an additional assumption on the fitness distribution is required as well, namely that its support is bounded away from zero. This implies that newly introduced vertices (with in-degree zero) cannot be arbitrarily unattractive, which could cause these vertices to form longer paths. When all vertices have a fitness of at least $\delta > 0$, however, we expect typical distances to behave as in (5.0.2). This agrees with the behaviour of typical distances in affine preferential attachment models (we again refer to [73] for an overview), but shows that in the strong disorder regime, when $\alpha \in (2, 1+\theta_m)$ and hence $\tau = \alpha < 1 + \theta_m$, typical distances are significantly shorter compared to the affine model in which the affine parameter δ is set to equal $\mathbb{E}[\mathcal{F}]$. This is caused by the fact that high degrees in the strong disorder regime are significantly larger, as described by Theorem 2.6 in Chapter 2, so that the graph is better connected and shorter graph distances between vertices can be observed.

In the latter case, when the fitness random variables follow an infinite mean power-law distribution and we consider the PAFUD or PAFFD constructions with $m > 2$ or the PAFRO construction described in Chapter 2, we conjecture typical distances to be $2, 3$ or 4 with high probability, similar to what is observed for typical distances in the configuration model with infinite mean degrees [52]. We also conjecture this to be true for typical distances in weighted recursive graphs with $m > 1$, as we observe from the proof of Proposition 6.1, Equation (6.2), that the (mean) degrees in the PAF model and the WRG model behave the same in the large graph limit. These short distances are caused by the emergence of a small set of vertices with extremely large fitness (or weights, in the WRG model) which enables them to obtain very high degree and hence connect vertices with very short paths.

For the diameter of the graph, we would expect to observe similar behaviour, though we do not have an intuitive argument as to why this is the case. That is, in the weak disorder regime the diameter is of the same order as in affine preferential attachment models, in the strong disorder regime the diameter is of the same order, but converges to a smaller constant, and in the extreme disorder the diameter is bounded with high probability.

The height of the preferential attachment tree with additive fitness is studied by Sénizergues in [128] and by Sénizergues and Pain in [121]. By the equivalence between PAF trees and the WRT model discussed in Section 1.4, the results on the height of the WRT model hold for PAF trees as well in the weak disorder regime. The equivalence between the two models does not hold in the strong and extreme disorder regime, and here nothing is known about the height of the PAF tree. In the strong disorder, though the largest degrees in the graph grow at a faster rate compared to affine preferential attachment models, the maximum degree has a label of order n rather than $\mathcal{O}(1)$. It is not quite clear what the effect of both of these observations is on the height of the tree. In the extreme disorder regime, the tree is essentially made up of many stars, which are the vertices with the highest fitness values. We expect the height in this model to be bounded, just like typical distances are bounded.

Finally, the local weak limit for the PAF model as introduced by Lo [95] is only proved to hold when the fitness distribution has bounded support, though Lo states that fitness distributions with a exponentially decaying tail can also be allowed. More precise and involved techniques are required to deal with the more heavy-tailed fitness distributions, and we expect the same local weak limit to arise for any fitness distribution with finite mean. When the fitness distribution has infinite mean, it is not quite clear what the local weak limit should be.

Weighted recursive graphs

In Chapter 3 we studied the properties of the degree distribution and maximum degree of the weighted recursive graph. We expect that the convergence in probability of the rescaled maximum degree, as in Equation (2.15) can be improved to almost sure convergence in the Gumbel-SV sub-case. Moreover, it would be interesting to see if higher-order asymptotics can be obtained for the maximum degree in the Gumbel-RV and Gumbel-RaV sub-cases beyond what is presented in Theorems 2.11 and 2.13. Also, in the Gumbel-RV sub-case, we are not able to obtain a second-order correction term when $\tau > 1$. This is mainly due to the fact that fluctuations of the degrees around their conditional means are of higher-order than the second-order asymptotics of the conditional means themselves when $\tau > 1$. A different approach, possibly as developed in Chapter 4 could resolve this and allow for a more refined asymptotic understanding of the maximum degree in this case.

The main question of interest related to Chapter 3 is whether the results for the maximum degree in Theorem 2.8 can be obtained under weaker assumptions on the vertex-weight distributions in the Gumbel maximum domain of attraction (MDA) as well. In Chapter 3 we considered three different sub-cases, but these are not exhaustive. That is, there are many distributions which do not satisfy either of the SV, RV or RaV sub-cases but do belong to the Gumbel MDA. As we state in Remark 2.9(v), we expect WRGs with vertex-weight with such distributions to exert similar behaviour at least to first order. However, the second (and higher) order behaviour of the maximum for such distributions is most likely different. For particular choices of vertex-weight distributions we expect it to be possible to obtain similar results with the same techniques developed in Chapter 3, though we deem it unlikely to be able to obtain results for a more general representation of vertex-weights distributions in the Gumbel MDA, for example using the representation by Resnick [126, Corollary 1.7].

In the Bounded case, we discuss the growth rate of the location of the maximum degree in Conjecture 2.10, something we were unable to prove. Intuitively, the maximum degree grows much faster than the mean degree of any vertex, so it arises from a large deviation event in which a vertex is able to increase its degree significantly beyond its mean degree. Which vertices are best equipped depends on an intricate balance: it is most likely for 'old' vertices (i.e. vertices with a 'small' label) to have a degree that is sufficiently large (i.e. of the order $\log n/\log\theta_m$) to attain the maximum degree in the graph, but there are not many vertices with a 'small' label. At the same time, there are many 'young' vertices, but as they are so young it is even less likely for one of these vertices to attain the maximum degree. In the end, we expect vertices with labels of the order n^β, with $\beta := 1-(\theta_m-1)/(\theta_m\log\theta_m)$ to satisfy this balance of age and number of vertices.

In Chapter 4 we develop more refined techniques to analyse the degree distribution and large degrees in weighted recursive trees (WRT). Despite this more precise approach, we require the additional assumption that the vertex-weights are bounded away from zero. As this has no influence on the behaviour of high degrees at all, and since we only need this assumption for one particular inequality, we expect this assumption to be purely of a technical nature. We hope to be able to omit this assumption in the future. Furthermore, as is the case in Chapter 3, we wonder if more general assumptions on the vertex-weight distribution can be made in the case the distribution belongs to the

Gumbel MDA, and whether higher-order corrections for the maximum degree can be obtained for such distributions as well.

Also, though the focus in Chapter 4 is on vertex-weight distributions with bounded support, one could try to use the same approach to obtain higher-order asymptotics of the maximum degree when the vertex-weight distribution has unbounded support. This would require a more careful proof of mostly Proposition 5.1 in Chapter 4. However, for these techniques to be used to obtain a better understanding of the asymptotics of the maximum degree, more precise asymptotic expressions of the limiting degree distribution, as in Theorem 2.7 in Chapter 3, would be required as well.

we recall that at the end of Section 4 in Chapter 4 we discuss how a Poisson point process limit arises as the higher-order limit of the rescaled high degrees. Though we are only able to establish this for vertex-weight distributions with an atom at one and for two specific other examples, we conjecture this result to hold universally. More detailed results on the asymptotic expression of the limiting degree distribution p_k, as presented in Theorem 2.7 in Chapter 3, would be required to confirm or refute this claim.

Finally, for both the PAF models discuss in Chapter 2 and the WRG and WRT models discussed in Chapters 3 and 4 it is possible to use other techniques to analyse the behaviour of these models. Most notably, in the case of PAF trees and the WRT model, embedding these evolving trees in continuous-time branching processes (CTBP) could be especially fruitful in extending the results presented in this thesis. The theory of CTBPs is generally considered to be very strong and useful in many applications, among which analysing evolving tree models. Among others, CTBPs are used by Senkevich et. al [129] to study competing growth processes, Fountoulakis and Iyer [56] to study a general class of weighted recursive trees, Bhamidi [14] and Banerjee and Bhamidi [11] to study a general class of preferential attachment models and Garavaglia et. al [60] to study a preferential attachment tree model with multiplicative fitness and ageing.

Especially in the case of the more general description of vertex-weight distributions in the Gumbel MDA, as discussed as a possible improvement of the results presented in Chapters 3 and 4, this could be particularly useful (a more general description of vertex-weight distribution in the Gumbel MDA is used in [129], for example).

Finally, beyond extending the results presented in this thesis for the WRG model, it would be interesting to study the local weak limit of the WRG model. The techniques developed by Lo to construct the local weak limit of the PAF model in [95] should also suffice to obtain the local weak limit for the WRG model. Especially in the case of infinite mean vertex-weights, it would be exciting to see whether the local weak limit for the WRG and PAF models is the same.

Chapter 6

Bibliography

[1] M. A. Abdullah, M. Bode, and N. Fountoulakis. Typical distances in a geometric model for complex networks. *Internet Math.*, page 38, 2017.

[2] D. Achlioptas, A. Clauset, D. Kempe, and C. Moore. On the bias of traceroute sampling: or, power-law degree distributions in regular graphs. *J. ACM*, 56(4):Art. 21, 28, 2009.

[3] L. Addario-Berry and L. Eslava. High degrees in random recursive trees. *Random Structures Algorithms*, 52(4):560–575, 2018.

[4] L. Addario-Berry and K. Ford. Poisson-Dirichlet branching random walks. *Ann. Appl. Probab.*, 23(1):283–307, 2013.

[5] W. Aiello, A. Bonato, C. Cooper, J. Janssen, and P. Prał at. A spatial web graph model with local influence regions. *Internet Math.*, 5(1-2):175–196, 2008.

[6] R. Albert, H. Jeong, and A.-L. Barabási. Diameter of the world-wide web. *nature*, 401(6749):130–131, 1999.

[7] T. Antunović, E. Mossel, and M. Z. Rácz. Coexistence in preferential attachment networks. *Combin. Probab. Comput.*, 25(6):797–822, 2016.

[8] K. B. Athreya. Preferential attachment random graphs with general weight function. *Internet Math.*, 4(4):401–418, 2007.

[9] K. B. Athreya, A. P. Ghosh, and S. Sethuraman. Growth of preferential attachment random graphs via continuous-time branching processes. *Proc. Indian Acad. Sci. Math. Sci.*, 118(3):473–494, 2008.

[10] L. Backstrom, P. Boldi, M. Rosa, J. Ugander, and S. Vigna. Four degrees of separation. In *Proceedings of the 4th Annual ACM Web Science Conference*, pages 33–42, 2012.

[11] S. Banerjee and S. Bhamidi. Persistence of hubs in growing random networks. *Preprint, arXiv:2004.13785*, 2020.

[12] A.-L. Barabási. Love is all you need: Clauset's fruitless search for scale-free networks. *Blog post available at https://www. barabasilab. com/post/love-is-all-you-need*, page 20, 2018.

[13] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[14] S. Bhamidi. Universal techniques to analyze preferential attachment trees: Global and local analysis. *Preprint available at http://www. unc. edu/ bhamidi*, 2007.

[15] G. Bianconi and A.-L. Barabási. Bose-einstein condensation in complex networks. *Physical review letters*, 86(24):5632, 2001.

[16] M. Boguñá, F. Papadopoulos, and D. Krioukov. Sustaining the internet with hyperbolic mapping. *Nature communications*, 1(1):1–8, 2010.

[17] B. Bollobás. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European J. Combin.*, 1(4):311–316, 1980.

[18] B. Bollobás, C. Borgs, J. Chayes, and O. Riordan. Directed scale-free graphs. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms (Baltimore, MD, 2003)*, pages 132–139. ACM, New York, 2003.

[19] B. Bollobás, S. Janson, and O. Riordan. The phase transition in inhomogeneous random graphs. *Random Structures Algorithms*, 31(1):3–122, 2007.

[20] B. Bollobás, S. Janson, and O. Riordan. Sparse random graphs with clustering. *Random Structures Algorithms*, 38(3):269–323, 2011.

[21] B. Bollobás and O. Riordan. The diameter of a scale-free random graph. *Combinatorica*, 24(1):5–34, 2004.

[22] B. Bollobás, O. Riordan, J. Spencer, and G. Tusnády. The degree sequence of a scale-free random graph process. volume 18, pages 279–290. 2001.

[23] C. Borgs, J. Chayes, C. Daskalakis, and S. Roch. First to market is not everything: an analysis of preferential attachment with fitness. In *STOC'07—Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, pages 135–144. ACM, New York, 2007.

[24] K. A. Borovkov and V. Vatutin. Trees with product-form random weights. *Discrete Mathematics & Theoretical Computer Science*, 2006.

[25] K. A. Borovkov and V. A. Vatutin. On the asymptotic behaviour of random recursive trees in random environments. *Adv. in Appl. Probab.*, 38(4):1047–1070, 2006.

[26] T. Britton, M. Deijfen, and A. Martin-Löf. Generating simple random graphs with prescribed degree distribution. *J. Stat. Phys.*, 124(6):1377–1397, 2006.

[27] A. D. Broido and A. Clauset. Scale-free networks are rare. *Nature communications*, 10(1):1–10, 2019.

[28] F. Caravenna, A. Garavaglia, and R. van der Hofstad. Diameter in ultra-small scale-free random graphs. *Random Structures Algorithms*, 54(3):444–498, 2019.

[29] D. G. Champernowne. A model of income distribution. *The Economic Journal*, 63(250):318–351, 1953.

[30] W. C. Chen and W. C. Ni. Internal path length of the binary representation of heap-ordered trees. *Inform. Process. Lett.*, 51(3):129–132, 1994.

[31] F. Chung and L. Lu. The average distances in random graphs with given expected degrees. *Proc. Natl. Acad. Sci. USA*, 99(25):15879–15882, 2002.

[32] F. Chung and L. Lu. Connected components in random graphs with given expected degree sequences. *Ann. Comb.*, 6(2):125–145, 2002.

[33] A. Clauset and C. Moore. Accuracy and scaling phenomena in internet mapping. *Physical Review Letters*, 94(1):018701, 2005.

[34] A. Clauset, C. R. Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *SIAM Rev.*, 51(4):661–703, 2009.

[35] M. Deijfen, H. van den Esker, R. van der Hofstad, and G. Hooghiemstra. A preferential attachment model with random initial degrees. *Ark. Mat.*, 47(1):41–72, 2009.

[36] S. Dereich. Preferential attachment with fitness: unfolding the condensate. *Electron. J. Probab.*, 21:Paper No. 3, 38, 2016.

[37] S. Dereich, C. Mailler, and P. Mörters. Nonextensive condensation in reinforced branching processes. *Ann. Appl. Probab.*, 27(4):2539–2568, 2017.

[38] S. Dereich, C. Mönch, and P. Mörters. Typical distances in ultrasmall random networks. *Adv. in Appl. Probab.*, 44(2):583–601, 2012.

[39] S. Dereich, C. Mönch, and P. Mörters. Distances in scale free networks at criticality. *Electron. J. Probab.*, 22:Paper No. 77, 38, 2017.

[40] S. Dereich and P. Mörters. Random networks with sublinear preferential attachment: degree evolutions. *Electron. J. Probab.*, 14:no. 43, 1222–1267, 2009.

[41] S. Dereich and P. Mörters. Random networks with concave preferential attachment rule. *Jahresber. Dtsch. Math.-Ver.*, 113(1):21–40, 2011.

[42] S. Dereich and M. Ortgiese. Robust analysis of preferential attachment models with fitness. *Combin. Probab. Comput.*, 23(3):386–411, 2014.

[43] L. Devroye and S. Janson. Long and short paths in uniform random recursive dags. *Ark. Mat.*, 49(1):61–77, 2011.

[44] L. Devroye and J. Lu. The strong convergence of maximal degrees in uniform random recursive trees and dags. *Random Structures Algorithms*, 7(1):1–14, 1995.

[45] P. S. Dodds, R. Muhamad, and D. J. Watts. An experimental study of search in global social networks. *science*, 301(5634):827–829, 2003.

[46] S. Dommers, R. van der Hofstad, and G. Hooghiemstra. Diameters in preferential attachment models. *J. Stat. Phys.*, 139(1):72–107, 2010.

[47] E. Eisenberg and E. Y. Levanon. Preferential attachment in the protein network evolution. *Physical review letters*, 91(13):138701, 2003.

[48] P. Erdős and A. Rényi. On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 5:17–61, 1960.

[49] P. Erdős and A. Rényi. On the evolution of random graphs. *Bull. Inst. Internat. Statist.*, 38:343–347, 1961.

[50] P. Erdős and A. Rényi. On the strength of connectedness of a random graph. *Acta Math. Acad. Sci. Hungar.*, 12:261–267, 1961.

[51] G. Ergün and G. J. Rodgers. Growing random networks with fitness. *Physica A: Statistical Mechanics and its Applications*, 303(1-2):261–272, 2002.

[52] H. van den Esker, R. van der Hofstad, G. Hooghiemstra, and D. Znamenski. Distances in random graphs with infinite mean degrees. *Extremes*, 8(3):111–141 (2006), 2005.

[53] L. Eslava. Depth of vertices with high degree in random recursive trees. *arXiv preprint arXiv:1611.07466*, 2016.

[54] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. In *The Structure and Dynamics of Networks*, pages 195–206. Princeton University Press, 2011.

[55] A. D. Flaxman, A. M. Frieze, and J. Vera. A geometric preferential attachment model of networks. *Internet Math.*, 3(2):187–205, 2006.

[56] N. Fountoulakis and T. Iyer. Condensation phenomena in preferential attachment trees with neighbourhood influence. *arXiv preprint arXiv:2101.02734*, 2021.

[57] N. Fountoulakis, T. Iyer, C. Mailler, and H. Sulzbach. Dynamical models for random simplicial complexes. *arXiv preprint arXiv:1910.12715*, 2019.

[58] N. Fountoulakis, P. van der Hoorn, T. Müller, and M. Schepers. Clustering in a hyperbolic model of complex networks. *Electron. J. Probab.*, 26:Paper No. 13, 132, 2021.

[59] N. Freeman and J. Jordan. Extensive condensation in a model of preferential attachment with fitness. *Electron. J. Probab.*, 25:Paper No. 68, 42, 2020.

[60] A. Garavaglia, R. van der Hofstad, and G. Woeginger. The dynamics of power laws: fitness and aging in preferential attachment trees. *J. Stat. Phys.*, 168(6):1137–1179, 2017.

[61] J. L. Gastwirth and P. K. Bhattacharya. Two probability models of pyramid or chain letter schemes demonstrating that their promotional claims are unreliable. *Oper. Res.*, 32(3):527–536, 1984.

[62] E. N. Gilbert. Random graphs. *Ann. Math. Statist.*, 30:1141–1144, 1959.

[63] W. Goh and E. Schmutz. Limit distribution for the maximum degree of a random recursive tree. *J. Comput. Appl. Math.*, 142(1):61–82, 2002. Probabilistic methods in combinatorics and combinatorial optimization.

[64] C. Goldschmidt and B. Haas. A line-breaking construction of the stable trees. *Electronic Journal of Probability*, 20:1–24, 2015.

[65] A. Grauer, L. Lüchtrath, and M. Yarrow. Preferential attachment with location-based choice: Degree distribution in the noncondensation phase. *arXiv preprint arXiv:1905.08481*, 2019.

[66] L. Gugelmann, K. Panagiotou, and U. Peter. Random hyperbolic graphs: degree sequence and clustering. In *International Colloquium on Automata, Languages, and Programming*, pages 573–585. Springer, 2012.

[67] L. de Haan and A. Ferreira. *Extreme value theory*. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2006. An introduction.

[68] J. Haslegrave and J. Jordan. Preferential attachment with choice. *Random Structures Algorithms*, 48(4):751–766, 2016.

[69] J. Haslegrave and J. Jordan. Non-convergence of proportions of types in a preferential attachment graph with three co-existing types. *Electron. Commun. Probab.*, 23:Paper No. 54, 12, 2018.

[70] J. Haslegrave, J. Jordan, and M. Yarrow. Condensation in preferential attachment models with location-based choice. *Random Structures Algorithms*, 56(3):775–795, 2020.

[71] E. Hiesmayr and U. Işlak. Asymptotic results on Hoppe trees and their variations. *J. Appl. Probab.*, 57(2):441–457, 2020.

[72] R. van der Hofstad. *Random graphs and complex networks. Vol. 1.* Cambridge Series in Statistical and Probabilistic Mathematics, [43]. Cambridge University Press, Cambridge, 2017.

[73] R. van der Hofstad. *Random graphs and complex networks*, volume 2. Book in preparation, http://www.win.tue.nl/~rhofstad/NotesRGCNII.pdf, 2021+.

[74] R. van der Hofstad and G. Hooghiemstra. Universality for distances in power-law random graphs. *J. Math. Phys.*, 49(12):125209, 14, 2008.

[75] R. van der Hofstad, G. Hooghiemstra, and P. Van Mieghem. Distances in random graphs with finite variance degrees. *Random Structures Algorithms*, 27(1):76–123, 2005.

[76] R. van der Hofstad, G. Hooghiemstra, and D. Znamenski. Distances in random graphs with finite mean and infinite variance degrees. *Electron. J. Probab.*, 12:no. 25, 703–766, 2007.

[77] R. van der Hofstad, P. Van Der Hoorn, N. Litvak, and C. Stegehuis. Limit theorems for assortativity and clustering in null models for scale-free networks. *Adv. in Appl. Probab.*, 52(4):1035–1084, 2020.

[78] C. Holmgren and S. Janson. Fringe trees, Crump-Mode-Jagers branching processes and m-ary search trees. *Probab. Surv.*, 14:53–154, 2017.

[79] T. Iyer. Degree distributions in recursive trees with fitnesses. *Preprint, arXiv:2005.02197*, 2020.

[80] E. Jacob and P. Mörters. Spatial preferential attachment networks: power laws and clustering coefficients. *Ann. Appl. Probab.*, 25(2):632–662, 2015.

[81] L. F. James. Generalized mittag leffler distributions arising as limits in preferential attachment models. *arXiv preprint arXiv:1509.07150*, 2015.

[82] S. Janson. Asymptotic degree distribution in random recursive trees. *Random Structures Algorithms*, 26(1-2):69–83, 2005.

[83] H. Jeong, Z. Néda, and A.-L. Barabási. Measuring preferential attachment in evolving networks. *EPL (Europhysics Letters)*, 61(4):567, 2003.

[84] J. Jordan. Preferential attachment graphs with co-existing types of different fitnesses. *J. Appl. Probab.*, 55(4):1211–1227, 2018.

[85] J. Jorritsma and J. Komjáthy. Distance evolutions in growing preferential attachment graphs. *arXiv preprint arXiv:2003.08856*, 2020.

[86] J. Jorritsma and J. Komjáthy. Weighted distances in scale-free preferential attachment models. *Random Structures Algorithms*, 57(3):823–859, 2020.

[87] J. M. Kleinberg, R. Kumar, P. Raghavan, S. Rajagopalan, and A. S. Tomkins. The web as a graph: measurements, models, and methods. In *Computing and combinatorics (Tokyo, 1999)*, volume 1627 of *Lecture Notes in Comput. Sci.*, pages 1–17. Springer, Berlin, 1999.

[88] J. Komjáthy and B. Lodewijks. Explosion in weighted hyperbolic random graphs and geometric inhomogeneous random graphs. *Stochastic Processes and their Applications*, 130(3):1309–1367, 2020.

[89] P. L. Krapivsky and S. Redner. Organization of growing random networks. *Physical Review E*, 63(6):066123, 2001.

[90] P. L. Krapivsky and S. Redner. Choice-driven phase transition in complex networks. *J. Stat. Mech. Theory Exp.*, (4):P04021, 14, 2014.

[91] P. L. Krapivsky, S. Redner, and F. Leyvraz. Connectivity of growing random networks. *Physical review letters*, 85(21):4629, 2000.

[92] D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat, and M. Boguñá. Hyperbolic geometry of complex networks. *Phys. Rev. E (3)*, 82(3):036106, 18, 2010.

[93] A. Lakhina, J. W. Byers, M. Crovella, and P. Xie. Sampling biases in IP topology measurements. In *IEEE INFOCOM 2003. Twenty-second Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE Cat. No. 03CH37428)*, volume 1, pages 332–341. IEEE, 2003.

[94] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graph evolution: Densification and shrinking diameters. *ACM transactions on Knowledge Discovery from Data (TKDD)*, 1(1):2–es, 2007, http://snap.stanford.edu/data/ca-CondMat.html.

[95] T. Y. Lo. Weak local limit of preferential attachment random trees with additive fitness. *arXiv preprint arXiv:2103.00900*, 2021.

[96] B. Lodewijks and M. Ortgiese. The maximal degree in random recursive graphs with random weights. *arXiv preprint arXiv:2007.05438*, 2020.

[97] B. Lodewijks and M. Ortgiese. A phase transition for preferential attachment models with additive fitness. *Electron. J. Probab.*, 25:Paper No. 146, 54, 2020.

[98] H. M. Mahmoud. Distances in random plane-oriented recursive trees. *J. Comput. Appl. Math.*, 41(1-2):237–245, 1992. Asymptotic methods in analysis and combinatorics.

[99] H. M. Mahmoud and R. T. Smythe. Asymptotic joint normality of outdegrees of nodes in random recursive trees. *Random Structures Algorithms*, 3(3):255–266, 1992.

[100] H. M. Mahmoud, R. T. Smythe, and J. Szymański. On the structure of random plane-oriented recursive trees and their branches. *Random Structures Algorithms*, 4(2):151–176, 1993.

[101] C. Mailler, P. Mörters, and A. Senkevich. Competing growth processes with random growth rates and random birth times. *Stochastic Process. Appl.*, 135:183–226, 2021.

[102] C. Mailler and G. Uribe Bravo. Random walks with preferential relocations and fading memory: a study through random recursive trees. *J. Stat. Mech. Theory Exp.*, (9):093206, 49, 2019.

[103] Y. Malyshkin and E. Paquette. The power of 2 choices over preferential attachment. *arXiv preprint arXiv:1311.1091*, 2013.

[104] Y. Malyshkin and E. Paquette. The power of choice combined with preferential attachment. *Electron. Commun. Probab.*, 19:no. 44, 13, 2014.

[105] S. S. Manna and P. Sen. Modulated scale-free network in euclidean space. *Physical Review E*, 66(6):066114, 2002.

[106] A. Meir and J. W. Moon. Recursive trees with no nodes of out-degree one. *Congr. Numer.*, 66:49–62, 1988. Nineteenth Southeastern Conference on Combinatorics, Graph Theory, and Computing (Baton Rouge, LA, 1988).

[107] S. Milgram. The small world problem. *Psychology today*, 2(1):60–67, 1967.

[108] M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. In *Proceedings of the Sixth International Seminar on Random Graphs and Probabilistic Methods in Combinatorics and Computer Science, "Random Graphs '93" (Poznań, 1993)*, volume 6, pages 161–179. Princeton University Press, 1995.

[109] M. Molloy and B. Reed. The size of the giant component of a random graph with a given degree sequence. *Combin. Probab. Comput.*, 7(3):295–305, 1998.

[110] J. W. Moon. The distance between nodes in recursive trees. In *Combinatorics (Proc. British Combinatorial Conf., Univ. Coll. Wales, Aberystwyth, 1973)*, pages 125–132. London Math. Soc. Lecture Note Ser., No. 13, 1974.

[111] T. F. Móri. On random trees. *Studia Sci. Math. Hungar.*, 39(1-2):143–155, 2002.

[112] T. F. Móri. The maximum degree of the Barabási-Albert random tree. *Combin. Probab. Comput.*, 14(3):339–348, 2005.

[113] T. Müller and M. Staps. The diameter of KPKVB random graphs. *Adv. in Appl. Probab.*, 51(2):358–377, 2019.

[114] H. S. Na and A. Rapoport. Distribution of nodes of a tree by degree. *Math. Biosci.*, 6:313–329, 1970.

[115] D. Najock and C. C. Heyde. On the number of terminal vertices in certain random trees with an application to stemma construction in philology. *J. Appl. Probab.*, 19(3):675–680, 1982.

[116] M. Newman. *Networks*. Oxford University Press, Oxford, 2018. Second edition of [MR2676073].

[117] M. E. Newman. Clustering and preferential attachment in growing networks. *Physical review E*, 64(2):025102, 2001.

[118] M. E. J. Newman. The structure and function of complex networks. *SIAM Rev.*, 45(2):167–256, 2003.

[119] I. Norros and H. Reittu. On a conditionally Poissonian graph process. *Adv. in Appl. Probab.*, 38(1):59–75, 2006.

[120] R. Oliveira and J. Spencer. Connectivity transitions in networks with super-linear preferential attachment. *Internet Math.*, 2(2):121–163, 2005.

[121] M. Pain and D. Sénizergues. Correction terms for the height of weighted recursive trees. *arXiv preprint arXiv:2101.01156*, 2021.

[122] B. Pittel. Note on the heights of random recursive trees and random m-ary search trees. *Random Structures Algorithms*, 5(2):337–347, 1994.

[123] D. d. S. Price. A general theory of bibliometric and other cumulative advantage processes. *Journal of the American society for Information science*, 27(5):292–306, 1976.

[124] D. d. S. Price. *Networks of scientific papers.* Princeton University Press, 2011.

[125] H. Prodinger and F. J. Urbanek. On monotone functions of tree structures. *Discrete Appl. Math.*, 5(2):223–239, 1983.

[126] S. I. Resnick. *Extreme values, regular variation and point processes.* Springer Series in Operations Research and Financial Engineering. Springer, New York, 2008. Reprint of the 1987 original.

[127] A. Rudas, B. Tóth, and B. Valkó. Random trees and general branching processes. *Random Structures Algorithms*, 31(2):186–202, 2007.

[128] D. Sénizergues. Geometry of weighted recursive and affine preferential attachment trees. *Preprint, arXiv:1904.07115*, 2019.

[129] A. Senkevich. *Competing growth processes with applications to networks.* PhD thesis, University of Bath, 2020.

[130] H. A. Simon. On a class of skew distribution functions. *Biometrika*, 42:425–440, 1955.

[131] R. Solomonoff and A. Rapoport. Connectivity of random nets. *Bull. Math. Biophys.*, 13:107–117, 1951.

[132] J. Szymański. On a nonuniform random recursive tree. In *Random graphs '85 (Poznań, 1985)*, volume 144 of *North-Holland Math. Stud.*, pages 297–306. North-Holland, Amsterdam, 1987.

[133] J. Szymański. On the maximum degree and the height of a random recursive tree. In *Random graphs '87 (Poznań, 1987)*, pages 313–324. Wiley, Chichester, 1990.

[134] A. Vázquez, R. Pastor-Satorras, and A. Vespignani. Large-scale topological and dynamical properties of the internet. *Physical Review E*, 65(6):066130, 2002.

[135] I. Voitalov, P. van der Hoorn, R. van der Hofstad, and D. Krioukov. Scale-free networks well done. *Physical Review Research*, 1(3):033034, 2019.

[136] P. Wan, T. Wang, R. A. Davis, and S. I. Resnick. Fitting the linear preferential attachment model. *Electron. J. Stat.*, 11(2):3738–3780, 2017.

[137] T. Wang and S. I. Resnick. Degree growth rates and index estimation in a directed preferential attachment model. *Stochastic Process. Appl.*, 130(2):878–906, 2020.

[138] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *nature*, 393(6684):440–442, 1998.

[139] G. U. Yule. A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, FRS. *Philosophical transactions of the Royal Society of London. Series B, containing papers of a biological character*, 213(402-410):21–87, 1925.