Social Network Analysis

LM Informatics 2017-18, Course 72677
University of Bologna

Giovanni Rossi
Department of Computer Science and Engineering
University of Bologna, email: roxyjean@gmail.com

25 Sept - 15 Dec, 2017

1 Introduction

- Networks, consisting of a set of nodes a set of pairs of such nodes, are nowadays employed for modeling a wide variety of phenomena in different fields including computer science, sociology, engineering, physics and biology, to the extent that over the last few decades a "new science of networks" has emerged [24]. With respect to previous applications of graph theory in electrical engineering or in sociology, where networks either need to be designed satisfying certain requisites or else are used for essentially descriptive purposes, the main distinctive feature of the new science of networks is that it focuses on real-world ones, possibly evolving over time due to the autonomous behaviour of individual nodes, and in any case formalizing the available knowlodge about a complex system whose whole functioning and structure are sought to be understood [20]. Typical complex networks, consisting of really many nodes and many pairs of nodes, are the Internet, the World Wide Web and PPI (protein-to-protein interaction) networks.

- The novel science is thus concerned both with theoretical questions and findings as well as with empirical observations, and in particular also aims at modeling real complex networks by reproducing/comparing them via/with variations of the traditional Erdős-Rény random graph model [3, 19, 26]. In fact, although formalizing so diverse phenomena, complex networks still generally display some common features, like the notorious "small-world". However, social networks are different in two main respects detailed in the sequel [25].

- The purpose of this course is to eventually explore the objectives and methods of community detection in complex networks, which not only means searching for cohesive groups in social networks, but also refers, more broadly and importantly, to uncovering the modular structure which is found to characterize all types of complex networks [12]. Roughly speaking, communities or modules are regions (i.e. spanned or induced subgraphs, see below) where the density of links between nodes is significantly high, namely higher than expected [22].

- Community detection is thus closely related to graph clustering [28], and both topics will be looked at from the combinatorial optimization perspective. More precisely, the fundamental example given by graph clustering via modularity maximization [5, 21] is going to be analyzed as a maximum-weight set partitioning problem, with the instance in form of a quadratic pseudo-Boolean (set) function [4] assigning scores to subsets of nodes. Within such a general framework, alternative (i.e. non-modularity-based) pseudo-Boolean functions assigning cluster scores to node subsets may be constructed, possibly cubic rather than quadratic, with the general aim to finely detect (in spanned or generated subgraphs) some specific sought features. Provided enough time, the setting may be also interpreted in terms of fuzzy and overlapping community detection [1, 17, 18, 27, 30, 31, 33, 34].

Contents

2 Simple graphs

- The following definitions can be found in [2, 11]. As usual, "maximal" and "minimal" refer either to real (and in particular integer) numbers, or else (and more often) to inclusion $\supseteq$. In this latter case, if a (sub)set or a (sub)graph is a maximal one with a certain property, then any of its proper subsets or subgraphs does not have that property (and the same for minimality, mutatis mutandis).

- The elementary type of network is a simple graph $G = (N, E)$, namely an ordered pair where the first element is a set $N = \{1, \ldots, n\}$ of (indices or labels of) nodes or vertices, while the second element consists of edges or links or arcs and is a subset $E \subseteq N_2 = \{\{i, j\} : 1 \leq i < j \leq n\}$ of the $\binom{n}{2}$-set of unordered pairs of nodes. Observe that $(\cdot, \cdot)$ and $\{\cdot, \cdot\}$ denote respectively ordered and unordered pairs.

- In terms of power sets $2^N = \{A : A \subseteq N\}$ and $2^{N_2} = \{E : E \subseteq N_2\}$, clearly $N_2 = \{A : A \in 2^N, |A| = 2\}$, and there are $|2^{N_2}| = 2^{\binom{n}{2}}$ different simple graphs on n labeled vertices.

- A graph is not simple when it is directed and/or with loops and/or weighted. Formally, $G = (N, E)$ is directed when $E \subseteq N \times N$ consists of ordered pairs $(i, j) \neq (j, i)$ of vertices $i, j \in N$ (rather than unordered ones $\{i, j\} = \{j, i\}$), and has loops if $(i, i) \in E$ for some $i \in N$. In other terms, in directed graphs the edge set is a binary relation on vertex set N. Also, a graph is weighted when there are real-valued weights $w : E \to \mathbb{R}$ on edges. In addition to this basic setting, one may conceive more complex non-simple (possibly weighted) graphs whose vertices are partitioned into types and/or have attributes.

- In the sequel the concern is with simple graphs $G = (N, E), E \in 2^{N_2}$, possibly endowed with a $[0, 1]$-valued weight function $w : E \to [0, 1]$ on edges. The ensemble of all these possibly [0,1]-weighted simple graphs may be regarded from the following geometric perspective. Recall that 2^{N_2} bijectively corresponds to the $2^{\binom{n}{2}}$-set $\{0, 1\}^{\binom{n}{2}}$ of extreme points (or vertices) of the $\binom{n}{2}$-dimensional unit hypercube $[0, 1]^{\binom{n}{2}}$ through characteristic functions $\chi_E : N_2 \to \{0, 1\}, E \in 2^{N_2}$ defined by

$$\chi_E(\{i, j\}) = \begin{cases} 1 \text{ if } \{i, j\} \in E, \\ 0 \text{ if } \{i, j\} \in N_2 \backslash E. \end{cases}$$

  In this view, simple graphs $G = (N, E)$ correspond to extreme points $\chi_E \in \{0, 1\}^{\binom{n}{2}}$ of the $\binom{n}{2}$-cube, while $[0, 1]$-weighted graphs $G_w = (N, E_w)$ correspond to other points $\mathbf{w} = \left(w(\{i, j\}_1), \ldots, w\left(\{i, j\}_{\binom{n}{2}}\right)\right) \in [0, 1]^{\binom{n}{2}}$ of the cube or fuzzy edge (sub)sets, with $w(\{i, j\}) = 0$ if $\{i, j\} \in N_2 \backslash E$. This means indexing the $\binom{n}{2}$ axes (or elements of the canonical basis) of $\mathbb{R}^{\binom{n}{2}} \supset [0, 1]^{\binom{n}{2}}$ by the $\binom{n}{2}$ unordered pairs $\{i, j\} \in N_2$.

- What follows applies to simple graphs $G = (N, E), E \in 2^{N_2}$. A subgraph $G' = (N', E') \subseteq G$ satisfies $N' \subseteq N, E' \subseteq E$. Special attention shall be

devoted to those $2^n - 2$ subgraphs spanned or induced by vertex subsets $A, \emptyset \subset A \subset N$, denoted by $G(A) = (A, E(A))$, namely with vertex set A and edge set $E(A) = \{\{i,j\} : E \ni \{i,j\} \subseteq A\}$ consisting of the edges whose ends are both in A.

- A fundamental subgraph is a $i-j$-path $\mathfrak{P}_{ij} = (N_{\mathfrak{P}_{ij}}, E_{\mathfrak{P}_{ij}}) \subseteq G$ (for any $\{i,j\} \in N_2$), whose vertices $i_l \in N_{\mathfrak{P}_{ij}} = \{i = i_1, i_2, \ldots, i_{k+1} = j\}$ are all distinct and with edge set $E_{\mathfrak{P}_{ij}} = \{\{i_l, i_{l+1}\} : 1 \leq l \leq k\}$. The length of $\mathfrak{P}_{ij}$ is the number k of its edges. A $i-j$-path is geodesic or shortest if its length is minimal. The distance $dist_G(ij)$ between i and j in G equals the length of a geodesic $i-j$-path if there is one, and ∞ otherwise. In these terms, G is connected if $dist_G(ij) < \infty$ for all $\{i,j\} \in N_2$, and disconnected otherwise.

- A component is a maximal connected subgraph, where maximality entails that if $G' \subseteq G$ is a component (of G), then $G' = G(A)$ is the subgraph spanned by some vertex subset A. Connected graphs evidently have only one component. In fact, if $G(A_1), \ldots, G(A_k)$ are the components of $G = (N, E)$, then $\{A_1, \ldots, A_k\}$ is a partition of N as well as $\{E(A_1), \ldots, E(A_k)\}$ is a partition of E, as detailed hereafter.

2.1 Closure of graphs: partitions

- A partition $P = \{A_1, \ldots, A_{|P|}\}$ of N is a family of subsets $A_1, \ldots, A_{|P|} \neq \emptyset$ of N, or "blocks" of P, satisfying $A_l \cap A_k = \emptyset, 1 \leq l < k \leq |P|$ and $A_1 \cup \cdots \cup A_{|P|} = N$.

- In the complete graph $K_N = (N, N_2)$ all $\binom{n}{2}$ unordered pairs of vertices are edges. Extending this notation, for $\emptyset \subset A \subseteq N$, let $K_A = (A, A_2)$ be the complete graph on A, where $A_2 = \{\{i,j\} : A \supseteq \{i,j\} \in N_2\}$. A clique in a graph G is a maximal complete subgraph $(A, A_2) \subseteq G$. Clearly if $K_A \subseteq G$, then $K_A = G(A)$. Note that isolated nodes, namely those $i \in N$ (if any) such that $\{i,j\} \notin E$ for all $j \in N \backslash i$, are complete subgraphs (but of course non-maximal ones as long as $E \neq \emptyset$, hence any graph surely has at least one clique).

- Partitions P of N bijectively correspond to those very special graphs G on N whose components $G(A_1), \ldots, G(A_k)$ are each a complete subgraph (hence a clique), i.e. $G(A_l) = K_{A_l}, 1 \leq l \leq k$. The way to say this in combinatorial theory is the following. For any graph $G = (N, E)$, define its closure $\bar{G} = (N, \bar{E})$ to be the graph obtained from G by adding all edges within each component. Then, partitions may be regarded as those graphs $G = \bar{G}$ that coincide with their closure, thereby giving rise to the so-called "polygon matroid" defined on the edges of K_N [2].

- Coming to the components $G(A_1), \ldots, G(A_k)$ of any graph G, clearly every vertex as well as every edge are included in exactly one component, entailing that $\{A_1, \ldots, A_k\}$ and $\{E(A_1), \ldots, E(A_k)\}$ are partitions respectively of N and E.

- As already observed, there are $2^{\binom{n}{2}}$ graphs on n labeled vertices, while the (Bell) number of partitions of a n-set [2, 16], denoted by $\mathcal{B}_n$, obeys

recursion $\mathcal{B}_0 := 1$ and $\mathcal{B}_n = \sum_{0 \leq k < n} \binom{n-1}{k} \mathcal{B}_k$. By considering how many different non-closed graphs G are mapped via $G \to \bar{G}$ into each closed graph $\bar{G}$, it may be seen that $2^{\binom{n}{2}} - \mathcal{B}_n > 0$ $(n > 2)$ is rapidly incerasing. From the above geometric perspective, where graphs $G = (N, E)$ are extreme points $\chi_E \in \{0,1\}^{\binom{n}{2}}$ of the $\binom{n}{2}$-cube $[0,1]^{\binom{n}{2}}$ (through characteric functions χ_E of edge sets $E \in 2^{N_2}$), this means that closed graphs $G = \bar{G}$ identify only a proper subset $\{\chi_E : E = \bar{E}\} \subset \{0,1\}^{\binom{n}{2}}$ of such extreme points $(n > 2)$.

- In order to deal with graph clustering from the combinatorial optimization perspective (see above), in the sequel both subsets and partitions of N shall play a fundamental role. Accordingly, recall that $(2^N, \supseteq)$ is a poset (partially ordered set) and $(2^N, \cap, \cup)$ is the Boolean lattice of subsets (of a n-set). Denoting by $\mathcal{P}^N$ the $\mathcal{B}_n$-set of partitions of N, for $P, Q \in \mathcal{P}^N$ let $P \geqslant Q$ if for every block $B \in Q$ there is a block $A \in P$ such that $A \supseteq B$, in which case P is coarser than Q, or equivalently Q is finer than P. (Coarsening $\geqslant$ differs from greater-or-equal $\geq$ between real quantities.) Then, $(\mathcal{P}^N, \geqslant)$ is another main poset, while $(\mathcal{P}^N, \wedge, \vee)$ is the geometric lattice of partitions (of a n-set), where $\wedge$ and $\vee$ respectively are the meet or "coarsest-finer-than" and the join or "finest-coarser-than" operators [2].

- Most objective function-based graph clustering methods rely on maximizing (or minimizing) partition functions $f : \mathcal{P}^N \to \mathbb{R}$ sometimes called "additive" or "additively separable" [12, 13, 14], in that $f(P) = \sum_{A \in P} v(A)$ for all $P \in \mathcal{P}^N$, where $v : 2^N \to \mathbb{R}$ assigns weights to subsets and thus is the instance of a maximum-weight set partitioning problem.

2.2 Connectivity

- Another fundamental subgraph is the cycle, which is a $i - j$-path where $i = j$. A graph with no cycles is a forest, while a tree is a connected forest, hence forests have trees for components.

- Removing a vertex subset $A \in 2^N$ from a graph $G = (N, E)$ means removing also all edges with one or both ends in A. Thus if G_{-A} obtains from G by removing all vertices $i \in A$ and $A^c = N \backslash A$ is the complement of A (in 2^N), then $G_{-A} = G(A^c)$ is the subgraph spanned by A^c.

- For $k \geq 0$, a graph G on n vertices is k-connected if $k < n$ is the minimum number $k = |A|$ of vertices whose removal makes G_{-A} either disconnected or else the complete graph $K_{\{i\}}$ on only one vertex $i \in N$ (hence any graph on $n > 0$ vertices is 0-connected). The greatest integer k such that G is k-connected is the connectivity $\kappa(G)$ of G. Hence $\kappa(G) = 0$ if and only if either G is disconnected, or else $n = 1$, while if $\kappa(G) = 1$ then G simply is connected. Trees obviously provide the typical case where connectivity is 1. For all edges $\{i, j\} \in E$, spanned subgraphs $G(\{i, j\}) = K_{\{i,j\}}$ are the only minimal subgraphs with connectivity 1. In fact, if $G \neq K_{\{i,j\}}$ and $\kappa(G) = 1$, then G has at least one cutvertex, namely a vertex whose removal results in a graph $G_{-\{i\}}$ which is disconnected. On the other hand, the simplest example of a 2-connected graph is the cycle, and in

fact in any 2-connected graph every vertex belongs to at least one cycle. Finally, the complete graph has connectivity $\kappa(K_N) = n - 1$.

- Similar definitions apply to the edge-connectivity, although it seems plain that any subset of edges can be removed while leaving unaffected all vertices. In particular, an edge is a bridge if its removal augments (by 1) the number of components. Equivalently, the bridges are all (and only) those edges (if any) that belong to no cycle. In trees every edge is a bridge.

- A block $\mathfrak{B} \subseteq G$ (of G) is a maximal connected subgraph with no cutvertices, where again maximality entails $\mathfrak{B} = G(A)$ for some $A \in 2^N$. Hence any subgraph $G' \subseteq G$ is a block (i.e. $G' = \mathfrak{B}$) if and only if
  - either G' is a maximal 2-connected subgraph,
  - or else $G' = G(\{i,j\})$ where $\{i,j\} \in E$ is a bridge (i.e. $G(\{i,j\})$ is the subgraph spanned by pair $\{i,j\} \in 2^N$, where $\{i,j\} \in E$ is also an edge and in paticular a bridge, thus i and j are cutvertices),
  - or else $G' = K_{\{i\}}$ where i is an isolated vertex: $\{i,j\} \in N_2 \backslash E = E^c$ for all $j \in N \backslash i$.

- Each edge being contained in exactly one block, if $\mathfrak{B}_1, \ldots, \mathfrak{B}_k$ are all the blocks of $G = (N, E)$, then their edge sets $E'_l, 1 \leq l \leq k$ collectively constitute a partition $\{E'_1, \ldots, E'_k\}$ of E (where $|E'| = 1$ when $E' = \{i,j\}$ is a bridge). In this view, a block is the 2-connected analog of a component. In fact, for those $\mathcal{B}_n$ graphs $G = \bar{G}$ that coincide with their closure (see above), blocks and components are the same.

- Menger (1927) (and max-flow min-cut) theorem may be summarized as follows. For any $\{i,j\} \in N_2$ such that $dist_G(ij) < \infty$, define $i-j$-paths $\mathfrak{P}^1_{ij}, \ldots, \mathfrak{P}^k_{ij} \subset G$ to be independent if the intersection of the node sets of any two of them contains only i and j. That is, $N_{\mathfrak{P}^l_{ij}} \cap N_{\mathfrak{P}^{l'}_{ij}} = \{i,j\}$ for all $1 \leq l < l' \leq k$. Then, a graph is k-connected if any two of its vertices can be joined by k independent paths.

2.3 Adjacency and Laplacian matrices

- The adjacency matrix $\mathcal{A} = \mathcal{A}_G \in \{0,1\}^{n \times n}$ of a simple graph $G = (N, E)$ has entries

$$a_{ij} = \begin{cases} 1 \text{ if } \{i,j\} \in E, \\ 0 \text{ if } \{i,j\} \in E^c = N_2 \backslash E, \end{cases} \quad \text{for } 1 \leq i,j \leq n.$$

  Hence $a_{ij} = a_{ji}$ and $a_{ii} = 0$ for all $\{i,j\} \in N_2$. On the other hand, if there may be loops and edges are directed, i.e. if $E \subseteq N \times N$ (or equivalently $E \in 2^{N \times N}$), then the adjacency matrix is not symmetric and may have 1s on the main diagonal, in that

$$a_{ij} = \begin{cases} 1 \text{ if } (i,j) \in E, \\ 0 \text{ if } (i,j) \in E^c = (N \times N) \backslash E, \end{cases} \quad \text{for } 1 \leq i,j \leq n.$$

  In other terms, the adjacency matrix of simple graphs has only $\binom{n}{2}$ valid entries, namely those above the main diagonal, and therefore essentially

coincides with the characteristic function χ_E of edge sets (see above). From this view, the ensemble of all directed graphs admitting loops and with [0,1]-valued weights on edges/loops is $[0,1]^{n^2}$.

- The adjacency matrix is the simplest one that is usually associated with graphs. In fact, algebraic graph theory [15] studies graphs by using algebraic properties of associated matrices, and in particular spectral graph theory [6, 7] studies the relation between graph properties and the spectrum (i.e. the eigenvalues and eigenvectors) of the adjacency and Laplacian matrices. The spectrum of the adjacency matrix of a graph is thus commonly referred to as the spectrum of that graph, while the Lapalacian spectrum clearly refers to the Laplacian matrix $\mathcal{L} = \mathcal{L}_G = (\ell_{ij})_{1\leq i,j\leq n}$, whose integer entries are

$$\mathcal{L}_{ij} = \begin{cases} -a_{ij} \text{ if } i \neq j, \\ d_G(i) \text{ if } i = j, \end{cases} \quad \text{for } 1 \leq i, j \leq n,$$

  where $d_G(i) = \sum_{j\in N} a_{ij}$ is the degree of vertex i in G (vertices of directed graphs have both in and out degrees, as detailed in the references).

- The general target of spectral graph theory thus is to obtain information from the spectra of the adjacency and Laplacian (and related) matrices. In objective function-based spectral graph clustering and community detection methods, where the goal is to find partitions $P = \{A_1, \ldots, A_k\} \in \mathcal{P}^N$ of vertices maximizing ratio $|E(A_1) \cup \cdots \cup E(A_k)|/|E|$, the information contained in spectra is primarily used for selecting a range for the number of clusters [32]. Perhaps the most immediate example comes from the simplest graph clustering problem, given by a closed graph $G = \bar{G}$, whose components $G(A_1), \ldots, G(A_k)$ are each a complete subgraph (see above). Its spectrum consists of eigenvalues $(-1)^{|A_l|-1}$ and $(|A_l| - 1)^1$; similarly the eigenvalues of the Laplacian spectrum are 0^k and $|A_l|^{|A_l|-1}$ (for $1 \leq l \leq k$), where multiplicities are indicated as exponents. Note that $r(P) = \sum_{1\leq l\leq k}(|A_l| - 1) = n - k$ is the rank function for the geometric lattice $(\mathcal{P}^N, \wedge, \vee)$ (of partitions $P = \{A_1, \ldots, A_k\}$ of N [2]). In general, the multiplicity of 0 as an eigenvalue of $\mathcal{L}$ counts the number k of components, and the associated eigenvectors are linear combinations of the characteristic functions $\chi_{A_1}, \ldots, \chi_{A_k}$ of these components' vertex sets.

- For simple graphs both $\mathcal{A}, \mathcal{L} \in \mathbb{R}^{n\times n}$ are symmetric, the latter also being positive semidefinite and singular. Denote their (real) eigenvalues respectively by $\lambda_1^{\mathcal{A}} \leq \cdots \leq \lambda_n^{\mathcal{A}}$ and $0 = \lambda_1^{\mathcal{L}} \leq \cdots \leq \lambda_n^{\mathcal{L}}$. What follows is succintly excerpted from [6, Sections 1.3,1.4,1.7].
  - $\mathrm{tr}(\mathcal{A}) = \sum_{1\leq k\leq n} \lambda_k^{\mathcal{A}} = 0$ as well as $\mathrm{tr}(\mathcal{L}) = \sum_{1\leq k\leq n} \lambda_k^{\mathcal{L}} = \sum_{i\in N} d_G(i) = 2|E|$;
  - A $i-j$-path where vertices need not be distinct is a $i-j$-walk, and when $i = j$ the walk is closed. The number of $i-j$-walks of length k is the i,j entry $(1 \leq i, j \leq n)$ of matrix $\mathcal{A}^k$. Hence the i,i entry of $\mathcal{A}^2$ equals $d_G(i)$ and $\mathrm{tr}(\mathcal{A}^2) = 2|E|$, while $\mathrm{tr}(\mathcal{A}^3)$ is six times the number of triangles (or complete subgraphs on three vertices).

– Necessary and sufficient condition

$$\lambda_1^{\mathcal{A}} = \cdots = \lambda_n^{\mathcal{A}} = 0 \Leftrightarrow E = \emptyset \Leftrightarrow 0 = \lambda_1^{\mathcal{L}} = \cdots = \lambda_n^{\mathcal{L}}$$

applies.

– For a graph G on $n > 1$ vertices, the second smallest Laplacian eigenvalue $\lambda_2^{\mathcal{L}}$ is the algebraic connectivity of G. Since the multiplicity of $0 = \lambda_1^{\mathcal{L}}$ equals the number of components, then $\lambda_2^{\mathcal{L}} \geq 0$, with equality if and only if G is disconnected. The algebraic connectivity is monotone, in that it does not decrease if edges are added, and also is a lower bound for the (vertex) connectivity: $\kappa(G) \geq \lambda_2^{\mathcal{L}}$ (see [10]).

2.4 Random graphs

- The random graph $\mathcal{G}(n,p)$ is the probability space whose elements are the $2^{\binom{n}{2}}$ simple graphs $G = (N, E)$ on a n-set of labeled vertices, and where $p \in [0,1]$ is the probability that any two vertices $\{i,j\} \in N_2$ are paired by an edge $\{i,j\} \in E$, thereby providing $\binom{n}{2}$ mutually independent events. Every edge set $E \in 2^{N_2}$ thus realizes with probability

$$p^{|E|}(1-p)^{\binom{n}{2}-|E|},$$

and in particular $p = \frac{1}{2}$ induces the uniform probability distribution over 2^{N_2}, as each of the $2^{\binom{n}{2}}$ edge sets realizes with probability $2^{-\binom{n}{2}}$.

- To get familiar with random graphs, the following simple exercise may be useful: determine the probability that in (any realization of) $\mathcal{G}(n,p)$

  1. there are exactly k edges ($0 \leq k \leq \binom{n}{2}$);
  2. there is some (i.e. at least one) isolated vertex;
  3. there is some complete subgraph $K_A, 1 < |A| \leq n$.

- The traditional way to study random graphs is by setting $p = p(n)$ and then considering certain properties of graphs (like, say, being connected) while searching for the corresponding threshold function $p^*(n)$, namely such that as $n \to \infty$

  - if $p(n) < p^*(n)$, then (any realization of) $\mathcal{G}(n, p(n))$ almost surely does not have the chosen property, while

  - if $p(n) > p^*(n)$, then $\mathcal{G}(n, p(n))$ almost surely has the property.

  In fact, $\mathcal{G}(n,p) = p\chi_{N_2} + (1-p)\chi_{\emptyset}$ can be thought of as evolving from empty fo full along the main diagonal of the $\binom{n}{2}$-cube $[0,1]^{\binom{n}{2}}$ (see above), while $p = p(n)$ evolves through ever increasing functions of n. Then, examples of threshold functions (taken from [29, p. 14]) are:

  – at $p^*(n) = n^{-2}$ edges appear, meaning that $p^*(n) = n^{-2}$ is the threshold function for non-emptyness;

  – at $p^*(n) = n^{-\frac{3}{2}}$ edges with a common end (which is thus a cutvertex) appear;

- at $p^*(n) = n^{-1-\frac{1}{k}}$ (with arbitrary k but fixed) trees with $k+1$ vertices appear;
- at $p^*(n) = n^{-1}$ triangles appear, as do cycles of every fixed length k;
- $p^*(n) = n^{-1} \ln n$ is the threshold function for connectedness;
- at $p^*(n) = n^{-\frac{2}{3}}$ complete subgraphs on four vertices appear;
- at $p^*(n) = n^{-\frac{2}{k-1}}$ (with arbitrary k but fixed) complete subgraphs K_A on $|A| = k$ vertices appear;
- $p^*(n) = n^{-\frac{1}{2}} (\ln n)^{\frac{1}{2}}$ is the threshold function for the property that every pair of vertices $\{i, j\} \in N_2$ has a common neighbor, namely some $h \in N \backslash \{i, j\}$ such that $\{i, h\}, \{j, h\} \in E$.

- As for complex network analysis, firstly random graphs have been a term of comparison, to see whether some empirical evidence found in the former is or is not coherent with the corresponding expectation in the latter. The probability $p_k^{\mathcal{G}}$ that a (i.e. any) vertex $i \in N$ in the random graph $\mathcal{G}(n, p)$ has degree $d_i^{\mathcal{G}} = k$ $(0 \leq k < n)$ is given (again) by the binomial distribution

$$p_k^{\mathcal{G}} = \binom{n-1}{k} p^k (1-p)^{n-1-k}.$$

Hence the expected or mean degree is

$$\langle k \rangle = z := \sum_{0 \leq k < n} k p_k^{\mathcal{G}} = (n-1)p \simeq np$$

while the variance is $(n-1)p(1-p)$. For $n \gg kz$ sufficiently large, the degree distribution $p_k^{\mathcal{G}}$ becomes the Poisson one:

$$p_k^{\mathcal{G}} = \frac{z^k e^{-z}}{k!}.$$

- Both these (binomial and Poisson) distributions are strongly peaked about their mean z, and have been compared with the degree distribution

$$p_G^k = \frac{|\{i : d_G(i) = k\}|}{n} \ (0 \leq k < n)$$

of real networks G, such as social ones and (portions of) the Internet and World Wide Web. Real degree distributions have been found to obey power-laws, which are basically those that do not fit (well enough) the central limit theorem, hence where the number of outliers, although small, still constitutes a non-negligible fraction of the whole. In real networks, the essential fact is that some vertices, whose number is a small but non-negligible fraction of n, have a very large degree. This is detailed hereafter

3 Power-laws and degree distributions

- Many quantities $x \in [x_{\min}, \infty)$, including wealth (but not height) across individuals, are distributed according to a power-law, with probability or

density $p(x)$ of values remaining non-negligible even when x is very large. In fact, a common way to "recognize" power-law distributions is by observing that they display a $\ln x, \ln p(x)$-plot fitted by a straight (negatively sloped) line. In turn, this entails *"scale invariance"*, namely that when comparing the densities at $p(x)$ and at some $p(cx)$, where c is a constant, they are always proportional (i.e. at all x). That is, $p(cx) \propto p(x)$ (or $p(cx) = f(c)p(x)$). Hence the relative likelihood between small and large events is the same, no matter what choice of "small" is made. In other terms, the density "scales", whence the name "scale-free" networks for those where the degree distributution obeys a power-law.

- If x is a *continuous* random variable with power-law distribution, then

$$p(x) = Cx^{-\alpha} \text{ for } x \geq x_{\min} \text{ and } \alpha > 1.$$

As for the normalization constant C, firstly note that a straight (negatively sloped) $\ln x, \ln p(x)$-plot means $\ln p(x) = -\alpha \ln x + c$ (with $\alpha > 0$), and therefore taking the exponential of both sides

$$p(x) = Cx^{-\alpha}, \text{ where } C = e^c.$$

More precisely, C is determined via normalization

$$1 = \int_{x_{\min}}^{\infty} p(x)\mathrm{d}x = C \int_{x_{\min}}^{\infty} x^{-\alpha}\mathrm{d}x = \frac{C}{1-\alpha} \left[x^{-\alpha+1}\right]_{x_{\min}}^{\infty}.$$

For $\alpha > 1$ (as otherwise the right-hand side diverges),

$$C = (\alpha - 1)x_{\min}^{\alpha-1},$$

and thus the proper normalized expression for the power-law density is

$$p(x) = \frac{\alpha - 1}{x_{\min}} \left(\frac{x}{x_{\min}}\right)^{-\alpha}.$$

"Some distributions follow a power law for part of their range but are cut off at high values of x. That is, above some value they deviate from the power law and fall off quickly towards zero. If this happens, then the distribution may be normalizable no matter what the value of the exponent α. Even so, exponents less than unity are rarely, if ever, seen." [23].

The mean value of x is

$$\langle x \rangle = \int_{x_{\min}}^{\infty} xp(x)\mathrm{d}x = C \int_{x_{\min}}^{\infty} x^{-\alpha+1}\mathrm{d}x = \frac{(\alpha-1)x_{\min}^{\alpha-1}}{2-\alpha} \left[x^{-\alpha+2}\right]_{x_{\min}}^{\infty},$$

hence infinite for $\alpha \leq 2$. For $\alpha > 2$,

$$\langle x \rangle = \frac{(\alpha-1)x_{\min}}{\alpha-2}.$$

Similarly,

$$\langle x^2 \rangle = \frac{C}{3-\alpha} \left[x^{-\alpha+3}\right]_{x_{\min}}^{\infty}$$

is infinite for $\alpha \leq 3$, while for $\alpha > 3$

$$\langle x^2 \rangle = \frac{(\alpha-1)x_{\min}^2}{\alpha-3}.$$

- If $x = k$ is a discrete random variable $k \in \{x_{\min}, x_{\min}+1, \ldots\} \subseteq \mathbb{N}$ (hence like the degree of nodes in real complex networks), then its density p_k may be defined to obey a power-law in the following two ways.

  (a) Firstly, by (simply) setting $p_k = Ck^{-\alpha}$ for some $\alpha > 1$ as well as $x_{\min} = k_{\min} = 1$, and next normalizing according to

  $$1 = \sum_{k=1}^{\infty} p_k = C \sum_{k=1}^{\infty} k^{-\alpha} = C\zeta(\alpha), \text{ or } C = \frac{1}{\zeta(\alpha)},$$

  the density is

  $$p_k = \frac{k^{-\alpha}}{\zeta(\alpha)}, \text{ where } \zeta(\alpha) = \sum_{k=1}^{\infty} k^{-\alpha}$$

  is the Riemann (zeta) ζ-function [2]. If, realistically, $k_{\min} > 1$, then $p_k = 0$ if $k < k_{\min}$, while

  $$p_k = \frac{k^{-\alpha}}{\zeta(\alpha, k_{\min})} \text{ if } k \geq k_{\min}, \text{ where}$$

  $$\zeta(\alpha, k_{\min}) = \sum_{k=k_{\min}}^{\infty} k^{-\alpha}$$

  is the (normalizing) incomplete or generalized ζ-function.

  (b) Secondly, the density may be set equal to

  $$p_k = C\frac{\Gamma(k)\Gamma(\alpha)}{\Gamma(k+\alpha)}, \text{ where}$$

  $$\Gamma(t) = \int_0^{\infty} t^{\alpha-1} e^{-t} \mathrm{d}t$$

  is the gamma function, which for positive integers k is $\Gamma(k) = (k-1)!$. In fact, $\frac{\Gamma(k)\Gamma(\alpha)}{\Gamma(k+\alpha)}$ is the Legendre-beta function, which for large k is similar to power-law $k^{-\alpha}$, thus providing a desired asymptotic behaviour. Normalization yields a very simple form for constant C:

  $$1 = \sum_{k=1}^{\infty} p_k = C \sum_{k=1}^{\infty} \frac{\Gamma(k)\Gamma(\alpha)}{\Gamma(k+\alpha)} = \frac{C}{\alpha - 1}, \text{ or } C = \alpha - 1, \text{ hence}$$

  $$p_k = (\alpha - 1)\frac{\Gamma(k)\Gamma(\alpha)}{\Gamma(k+\alpha)},$$

  with expectation or mean

  $$\langle k \rangle = \sum_{k=1}^{\infty} k p_k = \frac{\alpha - 1}{\alpha - 2}.$$

  Also,

  $$\langle k^2 \rangle = \sum_{k=1}^{\infty} k^2 p_k = \frac{(\alpha-1)^2}{(\alpha-2)(\alpha-3)}$$

- Although it may be difficult to detect and/or simulate power-law distributions [9, 23, 8], still several studies now agree that complex, possibly social networks undoubtedly display a degree distribution following a power-law (see [19] and the references there provided). Accordingly, the random graph model has been turned into the so-called "configuration model" [20], namely a graph in all respects random apart from the degree distribution, which is fixed by a power-law [26] (where this latter mimics those found empirically). However, even when sharing a common (power-law) degree distribution, still real networks deviate from the theoretical expectations computable for the configuration model in two fundamental respects:

  (a) the degree correlation between adjacent vertices, also called assortative mixing if positive (like in social networks), and disassortative mixing if negative (like in most non-social networks); the configuration model displays much less degree correlation than that observed empirically;

  (b) the clustering coefficient, also referred to as "transitivity", in that it is the expectation that a (i.e. any) triple of connected vertices spans a complete subgraph (or triangle); its empirical values are sensibly greater then the theoretical ones for the configuration model.

4 Configuration model

- Like the random graph $\mathcal{G}(n,p)$ is the probability space where any of the $2^{\binom{n}{2}}$ simple graphs on n labeled vertices may realize, similarly the configuration model [20] is the probability space $\mathcal{G}(n,(p_k)^{\alpha}_{k_{\min}})$ where only a proper subset of such graphs may realize. Specifically, for any discrete power-law distribution $(p_k)^{\alpha}_{k_{\min}}$ (i.e. with parameters $k_{\min},\alpha$) of type (a) or (b) above, a fixed *degee sequence* $d(1),\ldots,d(n)$ is generated where $d(i), i \in N$ are realizations of independent random variables identically distributed according to p_k, i.e. $Prob[d(i)=k]=p_k$ for all i. Then, in the configuration model the only graphs $G'=(N,E')$ that may realize, each with equal (uniform) probability, are those with such a fixed degree sequence $d_{G'}(i)=d(i)$ for all i. The fixed degree sequence may well be some $d_G(1),\ldots,d_G(n)$ observed in a real network G. In any case, the n realizations/observations must be such that $\sum_{i\in N} d(i) = 2|E'|$ is even.

- The probability space obtains by associating with each node i the number $d(i)$ of its "stubs", i.e. edges ending in i, and then placing the *uniform distribution* over all and only those *orderings or permutations* of the total $\sum_{i\in N} d(i) = d_{tot}$ stubs satisfying the following admissibility condition.

  For $1 \le k \le d_{tot}/2$, the $2k$-th and $2k-1$-th stubs in the ordering cannot:

  (i) be associated with the same node,

  (ii) be associated with any two distinct nodes $i,j \in N$ such that the $2k'$-th and $2k'-1$-th stubs have already been associated with i,j at some $k'<k$.

  The resulting (random) graph thus has for edges all $d_{tot}/2$ pairs of consecutive stubs (in the random order), hence (i) is the loop-free condition, while (ii) assures that there are no multiple edges. In this way, each graph

with the fixed degree sequence realizes with equal (uniform) probability given by the ratio of the number $\prod_{i \in N} d(i)!$ of different stub orderings yielding that graph, to the total number of admissible stub orderings.

- The configuration model primarily constitutes a benchmark for comparison with real networks. More precisely, like the theoretical values of the traditional random graph enable to see that real networks have power-law rather than Poisson/binomial degree distributions, similarly the theoretical values of the configuration model enable to see that, apart from degree distributions, real networks remain different from randomly generated ones. In particular, as already outlined, the difference concerns both the (expected) degree correlation between adjacent nodes, and the clustering coefficient.

4.1 Neighbors at increasing distances

- For every node $i \in N$, denote by $N_i^m = \{j : j \in N, dist_G(ij) = m\}$ the set of m-neighbors of i, namely nodes j at distance m from i in a given network G. Hence $N_i^0 = \{i\}$ and $|N_i^1| = d_G(i)$, while $N_i^\infty \neq \emptyset$ if G is disconnected (see above). Neighbors simply are 1-neighbors.

- The mean number $|N_i^1|$ of neighbours of a randomly chosen vertex i in the configuration model with degree distribution $(p_k)_{k_{\min}}^{\alpha}$ is simply the average degree $z_1 = \langle k \rangle = \sum_{k \geq k_{\min}} k p_k$ ($= z$ in previous sections).

- As for the average number $|N_i^2|$ of 2-neighbors of i, firstly note that the probability distribution of the degree of the vertex to which any edge leads is *proportional* to kp_k. In fact, a randomly chosen edge is more likely to end in nodes with higher degree, in precise proportion to nodes' degree. This means that the probability that any neighbor $j \in N_i^1$ has degree k is

$$Prob[d(j) = k] = \frac{k p_k}{\sum_{l \geq k_{\min}} l p_l} =: q_{k-1}.$$

This is the probability that j is linked to $k-1$ nodes $i' \neq i$ or 2-neighbors of i. Accordingly, the average degree of j is

$$\sum_{k \geq k_{\min}} k q_k = \sum_{k \geq k_{k_{\min}}} \frac{k(k+1)p_{k+1}}{\sum_{l \geq k_{\min}} p_l l} = \sum_{k \geq k_{k_{\min}}} \frac{k(k-1)p_k}{\sum_{l \geq k_{\min}} p_l l} = \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle}.$$

The mean number z_2 of i's 2-neighbors thus obtains by multplying this ratio by the average degree z_1 itself:

$$z_2 = \langle k^2 \rangle - \langle k \rangle.$$

By substituting the Poisson degree distribution $p_k = \frac{z^k e^{-z}}{k!}$ into this expression, the mean number of 2-neighbors in the random graph $\mathcal{G}(n,p)$ is found to be $z_2 = \langle k \rangle^2$, i.e. the square of the mean of $|N_i^1|$. On the other hand, for power-law distributions $(p_k)_{k_{\min}}^{\alpha}$ the first term $\langle k^2 \rangle$ dominates, and thus z_2 is much closer to the mean of the square degree (rather than to the square of the mean).

- Coming to $|N_i^m|$, at any distance from i the degree distribution for any node $j \in N_i^m$ remains given by q_k above. Hence the mean number z_m of m-neighbors of i satisfies recursion

$$z_m = \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} z_{m-1} = \frac{z_2}{z_1} z_{m-1},$$

and thus reiterating

$$z_m = \left(\frac{z_2}{z_1} \right)^m z_1.$$

- In the random graph $\mathcal{G}(n, p(n))$, probability $p^*(n) = n^{-1}$ is the threshold function for the appeareance not only of cycles (see above), but also of a *giant component*, namely one largest component containining a finite fraction S of the total number $n \to \infty$ vertices. That is, its size nS scales *linearly* with the size of the whole graph. Since in the random graph $z_1 = np$, the threshold for the (Poisson distributed) mean degree is $z_1^* = 1$. The analog threshold for the configuration model (or random graph with given degree distribution) is in terms of ratio $\frac{z_2}{z_1}$. In particular, depending on whether $z_2 > z_1$ or not, the mean number of m-neighbors either diverges or converges exponentially as m becomes large. Hence the average total number $\sum_{k>0} |N_i^k|$ of neighbours of vertex i (i.e. at all distances) is finite if $z_2 < z_1$ or infinite if $z_2 > z_1$ (as $n \to \infty$). If this number is finite, then clearly there can be no giant component, while if it is infinite then there *must* be a giant component. In other terms, the threshold for the appeareance of a giant component in the random graph with given degree distribution is $z_2^*/z_1^* = 1$. Rearranged in terms of $z_1 = \langle k \rangle$ and $z_2 = \langle k^2 \rangle - \langle k \rangle$, threshold $z_2^* = z_1^*$ takes form

$$\frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} = 1 \text{ or } \langle k^2 \rangle - 2\langle k \rangle = 0 = \sum_{k \geq k_{\min}} p_k k(k-2).$$

4.2 Small-world effect

- The popular "small-world" effect, which historically refers mostly to social networks, is basically the finding that even for relatively small values of m, still union (of disjoint sets) $\bigcup_{0 \leq k \leq m} N_i^k$ already contains a very large fraction of the total number n of nodes. Both in the random graph $\mathcal{G}(n, p)$ and in the configuration model $\mathcal{G}(n, (p_k)_{k_{\min}}^{\alpha})$, the small-world effect clearly may occur only well above the thresholds for the appeareance of a giant component, hence respectively for $z_1 \gg 1$ and $z_2 \gg z_1$, as otherwise most pairs of vertices would be separated by an infinite distance. At these values (well above the thresholds) the average vertex-vertex distance is

$$\ell = \frac{\log(n/z_1)}{\log(z_2/z_1)} + 1.$$

This increases logarithmically, ence rather slowly, with n, entailing that even for very large networks the typical distance between any two nodes

is expected to be quite small. In particular, since $z_2 = z_1^2$ in the random graph $\mathcal{G}(n,p)$, then

$$\ell = \frac{\log(n/z_1)}{\log(z_1)} + 1 = \frac{\log n}{\log z_1} - \frac{\log z_1}{\log z_1} + 1 = \frac{\log n}{\log z_1}.$$

- For social networks, a small value of the average vertex-vertex distance is known as the small-world effect since the 60s. More recently, most types of real or synthetic networks have been observed to display the same effect. This is not surprising when considering the *diameter*

$$diam(G) = \max_{i,j \in N} dist_G(ij)$$

of any graph G with M edges, $0 \leq M \leq \binom{n}{2}$. The random graph $\mathcal{G}(n,M)$ is the probability space where any $G = (N,E)$ realizes with probability

$$Prob[G] = \begin{cases} \frac{M!\left(\binom{n}{2}-M\right)!}{\binom{n}{2}!} \text{ if } |E| = M, \\ 0 \text{ if } |E| \neq M. \end{cases}$$

If $M = M(n) < \binom{n}{2}$ satisfies (as $n \to \infty$) $\frac{2M^2}{n^3} - \log n \to \infty$, then almost every graph G in $\mathcal{G}(n,M)$ has diameter $diam(G) = 2$ (see [3, Corollary 10.11 (ii), p. 263]). Also, if functions $d = d(n), M = M(n)$ satisfy (as $n \to \infty$) (a) $\frac{\log n}{d} - 3\log\log n \to \infty$, (b) $2^{d-1}M^d n^{-d-1} - \log n \to \infty$, (c) $2^{d-2}M^{d-1}n^{-d} - \log n \to -\infty$, then almost every graph in $\mathcal{G}(n,M)$ has diameter d (see [3, Corollary 10.12 (ii), p. 263], while if only conditions (a) and (b) hold, then almost every graph has diameter $\leq d$). More generally, almost every graph with $M = M(n) \gg n-1$ edges has diameter $\leq c\log n$ for some constant $c = c(M)$. In conclusion, if the diameter increases as $\log n$ or slower, then so also must the average vertex–vertex distance, entailing that most networks with a sufficient number of edges shall display the small-world effect.

4.3 Degree correlation: assortativity coefficient

- When analyzing degree correlation, the concern basically is with all pairs of values k,l for vertex degrees, i.e. $0 \leq k,l < n$, and with the likelihood that a randomly chosen edge has ends with degrees k and l.

- For any network $G = (N,E)$ and all $0 \leq k,l < n-1$, consider the ratio

$$\rho_G^{kl} = \frac{|\{\{i,j\} : \{i,j\} \in E, d_G(i) = k+1, d_G(j) = l+1\}|}{|E|}$$

of the number of edges whose ends have degrees $k+1$ and $l+1$, to the total number $|E|$ of edges. Evidently,

$$\sum_{0 \leq k,l < n-1} \rho_G^{kl} = 1,$$

and thus

$$\bar{\rho}_G := \sum_{0 \leq k,l < n-1} kl\rho_G^{kl}$$

is the average over all edges of G of the product of their endnodes' degrees.

- As already mentioned, degree correlation in complex networks G is measured via comparison with the configuration model $\mathcal{G}(n, (p_k)^{\alpha}_{k_{\min}}$, with degree distribution $(p_k)^{\alpha}_{k_{\min}}$ similar or identical to that of G. In the configuration model, the mean $\langle\rho\rangle$ of $\bar{\rho}_G$ simply is

$$\langle\rho\rangle = \sum_{0\leq k,l<n-1} klq_kq_l, \text{ where}$$

$q_k = \dfrac{(k+1)p_{k+1}}{\sum_{l\geq k_{\min}} lp_l}$ is the excess degree distribution from Section 4.1.

Hence q_kq_l is the mean of ρ^{kl}_G in the configuration model $\mathcal{G}(n, (p_k)^{\alpha}_{k_{\min}})$ or probability that a randomly chosen edge has ends with degrees k and l.

- The comparison between any given network G and the configuration model $\mathcal{G}(n, (p_k)^{\alpha}_{k_{\min}})$ thus achieves by means of quantity

$$r_G = \frac{\bar{\rho}_G - \langle\rho\rangle}{\sigma^2_q} = \frac{1}{\sigma^2_q}\left[\sum_{0\leq k,l<n-1} kl\left(\rho^{kl}_G - q_kq_l\right)\right], \text{ where}$$

$\sigma^2_q = \sum_{k\geq k_{\min}} k^2q_k - \left(\sum_{k\geq k_{\min}} kq_k\right)^2$ is the variance of distribution q_k.

- The sign of r_G depends on the difference between $\bar{\rho}_G$ and its expectation or mean $\langle\rho\rangle$ in the configuration model. More precisely, if $r_G = 0$, then G displays no degree correlation. On the other hand, if $r_G < 0$, then G displays negative degree correlation or disassortative mixing. Finally, if $r_G > 0$, then G displays positive degree correlation or assortative mixing, which is precisely the case of (most) social networks. Hence r_G is called the *"assortativity coefficient"* (of G).

4.4 Clustering coefficient

- The clustering coefficient $cc(G)$ of a network $G = (N, E), E \in 2^{N_2}$ is

$$cc(G) = \frac{3 \times \text{number of triangles in } G}{\text{number of connected triples in } G} \in [0, 1],$$

where the number of triangles in G equals $\mathrm{tr}(\mathcal{A}^3)/6$, i.e. the trace of the third power of the adjacency matrix of G divided by 6 (see above), while a connected triple (in G) is a tree on three vertices (included in G). Every triangle thus corresponds to three connected triples. In words, the clustering coefficient is the ratio of three times the number of complete subgraphs $K_{\{i,j,h\}} \subseteq G$ spanned by triples $\{i, j, h\} \subseteq N$ of vertices, to the number of trees on three vertices included in G. Hence $cc(G)$ is the probability that by randomly choosing two edges $\{i, j\}, \{i, h\} \in E$ with a common end i, the other two ends j, h are also adjacent: $\{j, h\} \in E$.

- In the configuration model $\mathcal{G}(n, (p_k)^{\alpha}_{k_{\min}})$ with same degree distribution $(p_k)^{\alpha}_{k_{\min}}$ as G, the mean of the clustering coefficient can be computed as

follows. If two neighbors $j, h \in N_i^1$ of the same vertex $i \in N$ have excess degrees k and l, then the probability that a randomly chosen edge links j and h is $2[k/(2|E|)][l/(2|E|)]$. The "mean number" of edges between j and h thus is $|E|$ times this quantity, or $kl/(2|E|)$. In fact, since the configuration model obtains by placing the uniform probability over the admissible orderings of the $2|E|$ stubs (see above), the probability that any two $2t$-th and $2t-1$-th consecutive positions ($t = 1, \ldots, |E|$) in an admissible random order are associated with j and h is $2[k/(2|E|)][l/(2|E|)]$, the first 2 counting the two ordered pairs $(j,h),(h,j)$. Hence the mean number of edges between j and h is this probability multiplied by the number $|E|$ of pairs of consecutive positions in an admissible random order of stubs. Since both vertices are neighbors of i, both k and l are distributed according to the excess degree density q_k (see above), and averaging over such a distribution the mean or expected clustering coefficient is

$$\langle cc \rangle = \frac{1}{2|E|} \left(\sum_{k \geq k_{\min}} k q_k \right)^2 = \frac{1}{2|E|} \left(\frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} \right)^2 = \frac{1}{n} \frac{\left(\langle k^2 \rangle - \langle k \rangle \right)^2}{\langle k \rangle^3},$$

where $2|E| = n\langle k \rangle$. This is the probability that for any two edges sharing a common end i in $\mathcal{G}(n, (p_k)_{k_{\min}}^{\alpha})$ their other two ends $j, h \in N_i^1$ are also adjacent. While $\langle cc \rangle$ explains sufficiently well the values of the clustering coefficient observed in non-social networks G (such as the Internet, World Wide Web and metabolic complexes), meaning that $\langle cc \rangle$ and $cc(G)$ are roughly the same, the clustering coefficient observed in social networks G takes much higher values than its expectation $\langle cc \rangle$ in the configuration model, i.e. $cc(G) \gg \langle cc \rangle$. Finally note that for the traditional random graph $\mathcal{G}(n,p)$ the mean clustering coefficient simply is $\langle cc \rangle = p$.

References

[1] B. Adamcsek, G. Palla, I. J. Farkas, I. Derényi, and T. Vicsek. CFinder: locating cliques and overlapping modules in biological networks. *Bioinformatics*, 22(8):1021–1023, 2006. https://doi.org/10.1093/bioinformatics/btl039.

[2] M. Aigner. *Combinatorial Theory (Reprint of the 1979 edition)*. Springer, 1997.

[3] B. Bollobás. *Random Graphs Second Edition*. Cambridge University Press, 2001.

[4] E. Boros and P. Hammer. Pseudo-Boolean optimization. *Discrete Applied Mathematics*, 123:155–225, 2002.

[5] U. Brandes, D. Delling, M. Gaertler, R. Görke, M. Hoefer, Z. Nikoloski, and D. Wagner. On modularity clustering. *IEEE Trans. on Knowledge and Data Engineering*, 20(2):172–188, 2007.

[6] A. E. Brower and W. H. Haemers. *Spectra of Graphs*. Springer, 2011.

[7] F. R. K. Chung. *Spectral Graph Theory*. American Mathematical Society, 1997.

[8] A. Clauset, C. R. Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703.

[9] A. Corral, A. Deluca, and R. F. i Cancho. A practical recipe to fit discrete power-law distributions. *arXiv*, 2012. arXiv:1209.1270v1.

[10] N. M. M. de Abreu. Old and new results on algebraic connectivity of graphs. *Linear Algebra and its Applications*, 423:53–73, 2007.

[11] R. Diestel. *Graph Theory*. Springer, 2000.

[12] S. Fortunato. Community detection in graphs. *Physics Reports*, 486(3-5):75–174, 2010.

[13] I. Gilboa and E. Lehrer. Global games. *International Journal of Game Theory*, (20):120–147, 1990.

[14] I. Gilboa and E. Lehrer. The value of information - an axiomatic approach. *Journal of Mathematical Economics*, 20(5):443–459, 1991.

[15] C. Godsil and G. Royle. *Algebraic Graph Theory*. Springer, 2001.

[16] R. L. Graham, D. E. Knuth, and O. Patashnik. *Concrete Mathematics - A Foundation for Computer Science (Second Edition)*. Addison-Wesley, 1994.

[17] X. Lei, S. Wu, L. Ge, and A. Zhang. Clustering and overlapping modules detection in PPI network based on IBFO. *PROTEOMICS*, 13(2):278–290, 2013. doi:10.1002/pmic.201200309.

[18] T. Nepusz, A. Petróczi, L. Négyessy, and F. Baszó. Fuzzy communities and the concept of bridgeness in complex networks. *Physics Review E*, 77(1):016107, 2008.

[19] M. E. J. Newman. Random graphs as models of networks. *arXiv*, 2002. arXiv:cond-mat/0202208v1.

[20] M. E. J. Newman. The Structure and Function of Complex Networks. *SIAM Review*, 45(2):167–256, 2003.

[21] M. E. J. Newman. Fast algorithm for detecting communities in networks. *Physics Review E*, 69(6):066133, 2004.

[22] M. E. J. Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103:8577–8582, 2006.

[23] M. E. J. Newman. Power laws, Pareto distributions and Zipf's law. *arXiv*, 2006. arXiv:cond-mat/0412004v3.

[24] M. E. J. Newman, A.-L. Barabási, and D. J. Watts. *The Structure and Dynamics of Networks*. Princeton University Press, 2006.

[25] M. E. J. Newman and J. Park. Why social networks are different from other types of networks. *Physical Review E*, 68(3):036122, 2003.

[26] M. E. J. Newman, D. J. Watts, and S. H. Strogatz. Random graph models of social networks. *Proceedings of the National Academy of Sciences PNAS*, 2002. doi: 10.1073/pnas.012582999.

[27] J. Reichardt and S. Bornholdt. Detecting fuzzy community structures in complex networks with a Potts model. *Physical Review Letters*, 93(21):218701, 2004.

[28] S. E. Schaeffer. Graph clustering. *Computer Science Review*, 1:27–64, 2007.

[29] J. Spencer. *The Strange Logic of random Graphs.* Springer, 2001.

[30] M. Szalay-Bekő, R. Palotai, B. Szappanos, I. A. Kovás, B. Papp, and P. Csermely. Hierarchical layers of overlapping network modules and community centrality. *Bioinformatics*, 28(16):2202–2204, 2012.

[31] J. Wang, J. Run, M. Li, and F.-X. Wu. Identification of hierarchical and overlapping functional modules in PPI networks. *IEEE Transactions on NanoBioscience*, 11(4):386–393, 2012. doi: 10.1109/TNB.2012.2210907.

[32] S. White and P. Smyth. A spectral clustering approach to finding communities in graphs. In H. Kargupta, J. Srivastava, C. Kamath, and A. Goodman, editors, *Proceedings of the 2005 SIAM Conference on Data Mining*, pages 274–285, 2005.

[33] H. Wu, L. Gao, J. Dong, and X. Jang. Detecting overlapping protein complexes by rough-fuzzy clustering in protein-protein networks. *Plos ONE*, 9(3-e91856), 2014. doi: 10.1371/journal.pone.0091856.

[34] S. Zhang, R.-S. Wang, and X.-S. Zhang. Identification of overlapping community structure in complex networks using fuzzy c-means clustering. *Phisica A*, 374:483–490, 2007.