Generating large scale-free networks with the Chung–Lu random graph model

Dario Fasino and Arianna Tonetto
Department of Mathematics, Computer Science and Physics
University of Udine, Udine, Italy

Francesco Tudisco
Gran Sasso Science Institute, L’Aquila, Italy

Abstract
Being able to produce synthetic networks by means of generative random graph models and scalable algorithms is a recurring tool-of-the-trade in network analysis, as it provides a well founded basis for the statistical analysis of various properties in real-world networks. In this paper, we illustrate how to generate large random graphs having a power-law degree profile by means of the Chung–Lu model. In particular, we are concerned with the fulfillment of a fundamental hypothesis that must be placed on the model parameters, without which the generated graphs loose all the theoretical properties of the model, notably, the controllability of the expected node degrees and the absence of correlations between the degrees of two nodes joined by an edge. We provide explicit formulas for the model parameters in order to generate random graphs which fulfil a number of requirements on the behavior of the smallest, largest, and average expected degrees and have several desirable properties, including a power-law degree distribution with any prescribed exponent larger than 2, the presence of a giant component and no potentially isolated nodes.

AMS subject classifications. 68R10, 90B15
Keywords. Scale-free networks, random graphs, Chung–Lu model.

1 Introduction
Networks are ubiquitous in the modern society as a large number of biological, social, engineering, and physical systems have been successfully modeled and implemented as networks. Understanding structures and dynamics defined over such networks has thus become a prevalent challenge across many disciplines. In particular, a problem of central importance in network analysis is to be able to produce random graphs that resemble certain fundamental structural properties that emerge from the empirical observation of real-world networks. This problem is not only theoretically interesting, but also of practical relevance. In fact, a variety of generative network models have been proposed in the literature and have been successfully used in a range of applications, including the analysis of the network dynamics, understanding common properties across diverse type of networked systems and pointing out the self-organization mechanisms at the basis of the formation of complex networks, see e.g., [22, 28] or the survey [5].

1
Several fundamental structural properties have emerged from the extensive statistical analysis of empirical network data that has been carried out over the years. Among the most prominent ones we name the small-world effect, the emergence of clusters, motifs and meso-scale structures and the degree distribution. In particular, starting from the seminal work by Barabási and Albert [2], a wealth of empirical studies has shown that the node degrees of many interesting real-world networks follow a power-law distributions with exponent $2 < \gamma < 3$ [12, 25, 27].

A graph or network is considered to have a power-law degree distribution with exponent $\gamma$ if the number $n_k$ of nodes having degree $k$ can be approximated by $\alpha k^{-\gamma}$, for some coefficient $\alpha$ that may depend on the size of the graph. In these networks a small but not negligible fraction of the nodes has very large degree. These nodes are called hubs. On the other hand, real-world networks having a power-law degree distribution are often sparse, that is, the average degree is much smaller than the size of the network. Furthermore, even though networks often undergo an evolution, growing in size during time due to the addition of new nodes and edges, the average degree remains roughly constant, see e.g., [14].

However, both empirical and analytical considerations on the power law distribution $n_k \approx \alpha k^{-\gamma}$ show that only when the exponent $\gamma$ lies in between 2 and 3 a growing network can have some hubs and be sparse at the same time. More precisely, if we let the network size grow unboundedly, depending on the value of $\gamma$ we have different behaviors: if $0 < \gamma < 2$ then the average degree diverges and the network cannot be sparse, while if $\gamma > 3$ then the degree variance is bounded and no large hub can appear [3, §4.4].

Networks having a power-law degree profile are called scale-free, due to the fact that the power law $f(k) = \alpha k^{-\gamma}$ fulfills the identity $f(ck) = C f(k)$, for some constant $C$ depending on $c$ but independent on $k$, so that the degree distribution remains unchanged, apart from a multiplicative factor, under scaling of the independent variable. One of the earliest and most used generative models for scale-free networks is the Barabási–Albert model [2]. This model generates graphs that evolve in time and is easily described by a simple node-level self-organization rule, the so-called preferential attachment. The generation process is initiated from a small subgraph, whose precise structure is asymptotically not influential on the degree profile. At each step of the generation, a new node is added to the network and is connected to $k$ pre-existing nodes, where $k$ is a fixed integer. Such $k$ nodes are chosen with a probability that is proportional to their current degree. With this rule, the degree distribution asymptotically follows a power law with exponent $\gamma = 3$, where the average degree is $k$ and the largest degree grows as $\mathcal{O}(\sqrt{n})$ on average, for a network with $n$ nodes.

Many other generative models for random scale-free networks have been developed since then. For example, the original preferential attachment model has been generalized along many directions, including the introducing of node deletion [24], node attractivity [13] and more general attachment rules [21, 20]. Notably, the generative model introduced by Dorogovtsev, Mendes and Samukhin in [13] depends on a parameter that can be tuned to adjust the degree distribution into any power law with exponent $\gamma \geq 2$. These models provide a justification for the emergence of power-law degree distributions for preferential attachment growth processes. Moreover, they allow to predict the behavior of a number of quantitative properties of large scale-free networks, and can be used as a baseline to detect real-world networks that deviate from this paradigm. For example, important conclusions on the presence of statistical correlations between degrees of neighboring nodes in power-law graphs are shown in [22] and rely on computational experiments with both random and real-world scale-free networks.

Even though a wealth of generative models for scale-free random networks is now available, the investigation of methods for modeling networks either analytically or numerically is still an active research direction in network science [19, 23, 30]. In fact, no generative model fits the ever-changing needs
of complex network analysis. Moreover, some models are difficult to analyze analytically, others may introduce (or miss) certain structural properties as degree correlations, node clustering or the occurrence of certain subgraphs. Thus, depending on the application and the context, one model can be preferred over another.

In this work, we focus on a random graph model originally proposed by Chung and Lu in \cite{7,8,11} and further thoroughly analyzed in \cite{9}. This model, which we refer to as the Chung–Lu model, is very flexible and conceptually very simple. The model depends on a vector \( w \) whose elements \( w_1, \ldots, w_n \) set out the node degrees in expectation, under suitable hypotheses. In other words, given a degree sequence \( w \), the model can generate random graphs with \( n \) nodes whose expected degrees are exactly \( w_1, \ldots, w_n \).

Our main goal is to explore to what extent large scale-free networks can be generated by means of the Chung–Lu random graph model. In particular, we consider random graphs drawn from a Chung–Lu model with vector \( w \) defined as

\[
    w_i = c(i_0 + i)^{1/(1-\gamma)}, \quad i = 1, \ldots, n
\]

and we show that, with a suitable choice of the parameters \( c \) and \( i_0 \), one can generate random graphs that have a number of desired properties. For example, they can have expected degrees with a prescribed power-law distribution \( n_k = \mathcal{O}(k^{-\gamma}) \), they can have specified average and largest expected degrees, allowing for graphs that are both sparse and have hubs, and they can have a giant component, i.e., a connected subgraph with a number of nodes that scales linearly with the number of nodes of the entire network.

The paper is organized as follows. In the next subsection we collect some preliminaries. Then, in Section \( \textbf{2} \) we present the main features of the Chung–Lu model and propose a simple and efficient procedure to generate random graphs from this model. We provide the MATLAB implementation of the graph generator and showcase its computing time performance. In Section \( \textbf{3} \) we discuss the main results of this work. First, we prove that a naive selection of the vector \( w \) is unable to respect a basic assumption of the model when the size of the graph becomes large. Then, we provide explicit expressions for parameters that allow to generate sequences of power-law graphs with exponent larger than 2 having a prescribed average degree, in expectation. As a consequence, we demonstrate the possibility of generating large scale-free graphs having almost surely a connected component comprising a significant fraction of all nodes.

1.1 Notations and basic results

We use standard graph theoretical notations and definitions (cf. \cite{9}). A (undirected) graph or network \( G \) consists of a finite set \( V \) of nodes (or vertices) and a set \( E \) of edges. For notational simplicity, we identify \( V \) with \( \{1, 2, \ldots, n\} \). Each edge \( e \in E \) is an unordered pair of nodes. If \( e = \{i, j\} \) then we say that nodes \( i \) and \( j \) are adjacent, and that \( e \) is incident to \( i \) (and also to \( j \)). An edge having the form \( \{i, i\} \) is called loop. The degree of a node \( i \in V \) is the number of edges incident to \( i \), denoted by \( d_i \). A loop contributes as one edge to the node degree. If \( d_i = 0 \) then we say that node \( i \) is isolated. For a vector \( w = (w_1, \ldots, w_n) \) we write \( \text{mean}(w) \) and \( \text{mean}_2(w) \) to denote its first and second order means, \( \text{mean}(w) = \frac{1}{n} \sum_{i=1}^{n} w_i \) and \( \text{mean}_2(w) = \frac{1}{n} \sum_{i=1}^{n} w_i^2 \), respectively.

In the sequel, we will repeatedly use the following elementary result: If \( f \) is a non-increasing function, then

\[
    \int_{i_0 + 1}^{n+i_0+1} f(x) \, dx \leq \sum_{i=1}^{n} f(i_0 + i) \leq f(i_0 + 1) + \int_{i_0 + 1}^{n+i_0} f(x) \, dx.
\]
2 The Chung–Lu random graph model

The Chung–Lu random graph model is one of the most widespread random graph models. A detailed description and analysis of such model is presented in [9], which collects important earlier results concerning the number and size of connected components [7, 10], average distance and diameter [8, 6], and spectral properties of the associated adjacency matrix [11]. An independent appearance of the same model can be found in [29], where it has been introduced to analyze connectivities in protein-protein interaction networks. In [1] Aiello, Chung and Lu described an earlier random graph model for power law degree distributions which is equivalent asymptotically to the Chung–Lu model.

The Chung–Lu model is usually denoted by $G(w)$ where $w = (w_1, \ldots, w_n)$ is a vector of non-negative real numbers which define the model. A random graph $G \in G(w)$ is a graph with $n$ nodes, whose edges are generated independently from one another according to the following rule: For $i, j = 1, \ldots, n$, the probability of having an edge between nodes $i$ and $j$ is

$$p_{ij} = w_i w_j \rho, \quad \rho = 1/(\sum_{k=1}^{n} w_k).$$

(2)

For mathematical convenience, loops are usually allowed. Hence, the expected number of edges attached to node $i$, that is, the expected degree of node $i$ is

$$\sum_{j=1}^{n} p_{ij} = \rho w_i \sum_{j=1}^{n} w_j = w_i.$$

(3)

In other words, the expected degree of each node in a graph $G \in G(w)$ is equal to the corresponding coefficient in the vector $w$. As noted in [20], the Chung–Lu model is the only random graph model where the probability of having an edge between nodes $i$ and $j$ is the product $f(w_i)f(w_j)$ of separate functions of the expected degrees of nodes $i$ and $j$. As a consequence, this is the only random graph model that does not introduce correlations among the degrees of the nodes joined by an edge. For that reason, the Chung–Lu model represents the fundamental “null model” for finding community structures in networks [15, 17, 24], since tightly interconnected node sets can be revealed as deviations from an uncorrelated random graph. Recall that the existence of nontrivial correlations among such degrees in Barabási–Albert networks is well known, see e.g., [20, 22] or [25, Sect. 7.2].

Finally, we mention that the Chung–Lu model is the basic building block of more recent and advanced random graph models, such as the degree corrected stochastic block model [16, 18] and the block two-level Erdős–Rényi model [19], which have been proposed as models for random graphs having a prescribed average degree distribution and integrating clustering effects and community structures that appear in social networks.

2.1 Admissible expected degree sequences

For certain vectors $w$ the number $p_{ij}$ defined in (2) can exceed 1. In order to preserve its probabilistic meaning, this issue is usually avoided by specifying the constraint $\max_{i=1\ldots,n} w_i^2 \leq \sum_{k=1}^{n} w_k$ on $w$, in such a way that $0 \leq p_{ij} \leq 1$. Other choices are possible; for example, various authors set $p_{ij}$ as the minimum between $w_i w_j / \sum_k w_k$ and 1, but in that case the identity (3) is no longer valid and all interesting theoretical properties of the model are lost. For this reason, in the present work we adopt the following definition.
Definition 1. Let $w \in \mathbb{R}^n$. We say that $w$ is admissible if $w_1 \geq w_2 \geq \ldots \geq w_n \geq 0$ and

$$w_1^2 \leq \sum_{i=1}^{n} w_i.$$ 

Moreover, we denote by $A_n$ the set of all admissible $n$-vectors.

Note that, if $w \in A_n$, then $w_1 \leq \sum_{i=1}^{n} w_i$, so all the entries of $w$ are bounded by $n$. While the ordering of the $w_i$'s is not essential in the definition of $G(w)$, we adopt it for simplicity of further computations.

2.2 An efficient generator of Chung–Lu random graphs

Since real-world networks are often very large, the availability of efficient random graph generators is crucial for practical purposes. The obvious algorithm based on (2) considers each node pair and generates the corresponding edge according to the prescribed probability. This is the approach implemented in the function sticky of the MATLAB package CONTEST [31], which is probably the earliest implementation of a $G(w)$-type random graph generator. The resulting computational cost is $O(n^2)$ for a graph with $n$ nodes, which is unsuitable for large graphs. An efficient implementation of the Chung–Lu model is included within the block two-level Erdős–Rényi (BTER) algorithm [19], which has been designed with the goal of producing random graphs resembling certain social network properties. A more efficient algorithm has been described in [23].

The computational cost of that algorithm is $O(n + m)$ on average, where $m = \frac{1}{2} \sum w_i^2$ is the expected number of edges in $G(w)$. For sparse graphs, this cost is essentially linear in $n$.

We propose here yet another generator for graphs in $G(w)$. The main advantage of our generator is its purely linear algebra implementation, which runs very efficiently on MATLAB due to the absence of explicit for-loops. Our algorithm is based on some ideas laid out in [4] [19] and, as for the method proposed in [23], has a running time essentially proportional to the number of edges in the graph. The algorithm implements the principle called “ball dropping” in [30]. Initially, the algorithm generates two random vectors, $I$ and $J$, whose entries are node indices. In each vector, the node index $i \in \{1, \ldots, n\}$ appears $\ell w_i/(2m)$ times on average, where $\ell$ is the length of the random vectors. Therefore, edges are generated by joining nodes $I(k)$ and $J(k)$ for $k = 1, \ldots, \ell$. Obviously, the procedure may produce repeated edges, which are removed at the end of the algorithm. To counteract the removal, the vector length is set to $\lceil m + e \rceil$ where $e = \frac{1}{2} (\sum w_i)^2 / (\sum w_i^2)$ is an estimate on the number of multiple edges that are generated in the first step (assuming $w \in A_n$, of course). Figure 1 shows the MATLAB implementation of this algorithm, which is also available via the repository https://github.com/ftudisco/scalefreechunglu. For completeness, we also provide a PYTHON version of the graph generator.

Figure 2 demonstrates the running time of the algorithm for various random graphs in the $G(w)$ model. We performed a series of experiments in MATLAB v.2019b on a laptop PC endowed with a i7-8550U processor and 1.80GHz CPU clock. We generated graphs with $n$ nodes where $n = 100 \cdot 2^{k-1}$ for $k = 1, \ldots, 14$ and the vectors $w$ are chosen using three different degree distributions: a constant distribution with $w_i = 4$, a random distribution where each $w_i$ is set by a pseudo-random generator uniform in $[0, 8]$, and a power law distribution with $\gamma = 3$ and expected average degree $\text{mean}(w) = 4$, obtained by the formula $w_i = 2 \sqrt{n/i}$. In all experiments, the expected number of edges is $4n$, even
3 Scale-free random graphs in the Chung–Lu model

Owing to the pervasiveness of scale-free networks in the real world, and the fact that the Chung–Lu model allows us to choose in advance the expected degree distribution of random graphs, it is natural to ask whether or not it is possible to generate large power-law networks with arbitrary exponent from $G(w)$, by a suitable choice of the parameter $w \in A_n$. The main goal of this section is to answer that question. In particular, we address the possibility of generating large graphs having prescribed statistical

Fig. 1: MATLAB function that, given the vector of expected degrees $w = (w_1, \ldots, w_n)$, generates the adjacency matrix of a random graph $G \in G(w)$.

Fig. 2: Mean performance of the random graph generator over 10 random trials. Each line represents a sequence of random graphs generated by a different degree distribution: constant $w_i = 4$ (red circles), uniform in $[0, 8]$ (blue crosses), power law with $\gamma = 3$ and expected average degree $\text{mean}(w) = 4$ (magenta triangles). Left: Running time vs number of edges. Right: Average time per generated edge. The timings are the averages over 10 runs per each dimension $n$ and degree profile.
properties, such as the exponent of the power law and the average degree.

Let \( n_k \approx \alpha k^{-\gamma} \) be the degree distribution of a scale-free network \( G \) with \( n \) vertices and \( \gamma > 1 \). Then, for large \( n \), the number \( N(k) \) of nodes with degree greater than or equal to \( k \) can be approximated by

\[
N(k) = \sum_{i=k}^{\infty} n_i \approx \alpha \int_{k}^{\infty} x^{-\gamma} \, dx = \frac{\alpha}{\gamma - 1} k^{1-\gamma}. \tag{4}
\]

For further reference, we note incidentally that the largest degree \( d_{\text{max}} \) for nodes in \( G \) can be estimated by imposing \( N(d_{\text{max}}) = 1 \), giving

\[
d_{\text{max}} \approx \left( \frac{\alpha}{\gamma - 1} \right)^{1/(\gamma - 1)}. \tag{5}
\]

On the other hand, if \( G \in G(w) \) with \( w \in A_n \) then it is reasonable to assume \( N(w_i) \approx i \), since the expected degree of nodes \( 1, \ldots, i \) is at least \( w_i \). Then, solving for \( w_i \) the approximate identity \( i \approx \alpha w_i^{1-\gamma}/(\gamma - 1) \), coming from (4), we obtain

\[
w_i \approx ci^{-\gamma}, \quad c = \left( \frac{\gamma - 1}{\alpha} \right)^{1/(\gamma - 1)}. \tag{6}
\]

This construction will be somewhat extended in the following result, which is based on an idea found in [8] and [9, §5.7]. Basically, a new parameter is introduced to shift the index \( i \). As we will see later on, the presence of that parameter is very useful to overcome some limitations on the structure of the networks arising from (6). At the same time, that parameter may affect the number of nodes having small degree, resulting in a degree profile that partially deviates from the estimate (5). We will illustrate and discuss an example of this phenomenon in Figure 3.

**Lemma 1.** Let \( n \in \mathbb{N} \) and \( \gamma > 1 \). Let \( w \in \mathbb{R}^n \) be a vector such that

\[
w_i = c(i_0 + i)^{-\gamma}, \quad i = 1, \ldots, n, \tag{7}
\]

for some positive constant \( c \) and \( i_0 > -1 \). If \( w \in A_n \) then a graph \( G \in G(w) \) has an expected degree distribution that follows a power law with exponent \( \gamma \). Namely, for \( k \geq w_n \), the number of nodes with expected degree \( k \) is approximately \( \alpha k^{1-\gamma} \gamma^{\gamma-1-1} \).

**Proof.** As shown in (3), the expected degree of the \( i \)-th vertex of \( G \) is \( w_i \). For \( x \geq w_n \) let \( N(x) \) be the number of nodes with expected degree greater than or equal to \( x \). Since the sequence (7) is strictly decreasing, we have \( N(w_i) = i \). Inverting the relation (7) we obtain

\[
i = \left( \frac{w_i}{c} \right)^{1-\gamma} - i_0.
\]

Hence, for \( k = 1, \ldots, n \) it holds

\[
N(k) = \left( \frac{k}{c} \right)^{1-\gamma} - i_0.
\]
Consequently, the number of nodes with expected degree $k$ is approximately

$$n_k = N(k) - N(k+1) = \frac{k^{1-\gamma} - (k+1)^{1-\gamma}}{c^{1-\gamma}} \approx \frac{\gamma - 1}{c^{1-\gamma}} k^{-\gamma},$$

where the last passage comes from the mean value theorem. \hfill \square

The foregoing lemma does not ensure that the vector $w$ in (7) belongs to $A_n$. That condition can be met by a suitable choice of the constant $c$, as shown in the following result in the simplest case $i_0 = 0$.

**Theorem 1.** For all $\gamma > 1$ and for all $n \in \mathbb{N}$ let $w_i = ci^{-p}$ for $i = 1, \ldots, n$, where $p = 1/(\gamma - 1)$, $0 < c \leq c_{\text{max}}$ and

$$c_{\text{max}} = \begin{cases} (1 - (n + 1)^{1-p})/(p - 1) & \gamma \neq 2 \\ \log(n + 1) & \gamma = 2. \end{cases}$$

Then $w \in A_n$.

**Proof.** If $\gamma \neq 2$ then $p \neq 1$ and from the leftmost inequality in (11) we have

$$w_i^2 = c^2 \leq c \frac{1 - (n + 1)^{1-p}}{p - 1} = c \int_1^{n+1} x^{-p} \, dx \leq c \sum_{i=1}^{n} i^{-p} = \sum_{i=1}^{n} w_i.$$

Thus the claim follows from Lemma 11. Analogously, when $\gamma = 2$ we have

$$w_i^2 = c^2 \leq c \log(n + 1) = c \int_1^{n+1} \frac{1}{x} \, dx \leq c \sum_{i=1}^{n} \frac{1}{i} = \sum_{i=1}^{n} w_i,$$

so $w \in A_n$ and the proof is complete. \hfill \square

It is worth noting that the inequality $c \leq c_{\text{max}}$ appearing in Theorem 11 is asymptotically tight. Indeed, suppose that we set $c > (n^{1-p} - p)/(1 - p)$ and $\gamma > 2$. Then, using the rightmost inequality in (11) we obtain

$$w_i^2 > c \frac{n^{1-p} - p}{1 - p} = c \left( 1 + \int_1^{n+1} x^{-p} \, dx \right) \geq c \sum_{i=1}^{n} i^{-p} = \sum_{i=1}^{n} w_i,$$

hence in this case $w \notin A_n$. When $\gamma = 2$ the analogous conclusion follows by setting $c > \log n + 1$, and for $1 < \gamma < 2$ by assuming $c > 1/(p - 1)$. In summary, we have the following result.

**Corollary 1.** Let $p = 1/(\gamma - 1)$ and $w_i = ci^{-p}$ for $i = 1, \ldots, n$. If $w \in A_n$ then $c \leq \hat{c}$ where

$$\hat{c} = \begin{cases} 1/(p - 1) & 1 < \gamma < 2 \\ \log n + 1 & \gamma = 2 \\ (n^{1-p} - p)/(1 - p) & \gamma > 2. \end{cases}$$

In the proof of Theorem 11 for simplicity we have only shown that admissible vectors of the form (7) with $i_0 = 0$ exist. However, in a similar way, it is possible to show that there are Chung–Lu scale-free networks obtained from admissible vectors of the same form, for every $i_0 > -1$. Indeed, if $\gamma \neq 2$ then it’s enough to choose $c$ and $i_0$ such that

$$c \leq (1 + i_0)^2p \frac{(n + i_0 + 1)^{1-p} - (1 + i_0)^{1-p}}{1 - p}.$$  \hfill (8)
In fact, choosing $c$ as in (8) we obtain
\[
w_i^2 = c^2(1 + i_0)^{-2p} \leq c \frac{(n + i_0 + 1)^{1-p} - (1 + i_0)^{1-p}}{1-p} = \int_{1+i_0}^{n+i_0+1} cx^{-p} \, dx \leq \sum_{i=1}^{n} w_i.
\]

For $\gamma = 2$ the condition analogous to (8) is
\[
c \leq (i_0 + 1)^2 (\log(n + i_0 + 1) - \log(i_0 + 1)).
\]

The introduction of $i_0$ allows us to prescribe the average expected degree $d$ and the largest expected degree $m$ in a Chung–Lu scale-free graph, under appropriate hypotheses. The next result, which is based on an idea found in [8] and [9, p. 109], explains how to achieve this goal by a suitable choice of the parameters $i_0$ and $c$ in (7).

**Theorem 2.** Let $n \in \mathbb{N}$ and $\gamma > 2$. Suppose that $d = d(n)$ and $M = M(n)$ are two nondecreasing functions of $n$ such that $0 < d(n) \leq M(n) \leq n$,
\[
\lim_{n \to \infty} \frac{d(n)}{M(n)} = 0,
\]

and there exists a constant $0 < \eta < 1$ independent on $n$ such that
\[
\eta nd(n) \geq M(n)^2.
\]

For $i = 1, \ldots, n$ let $w_i = c(i_0 + i)^{-p}$ where $p = 1/(\gamma - 1)$,
\[
c = c(n) = (1 - p)d(n)n^p, \quad i_0 = i_0(n) = n \left( \frac{(1 - p)d(n)}{M(n)} \right)^{1/p} - 1.
\]

Then,

1. for $n$ sufficiently large it holds $w \in \mathcal{A}_n$

2. any $G \in G(w)$ has an expected degree distribution that follows a power law with exponent $\gamma$, that is, for $k \geq w_n$, the number of nodes with expected degree $k$ is $n_k \approx \alpha k^{\gamma - 1}$ with $\alpha = (\gamma - 1)c^{\gamma - 1}$

3. the largest expected degree of any $G \in G(w)$ is $M(n)$ and the average expected degree is asymptotically $d(n)$, in the sense that $\lim_{n \to \infty} \frac{\text{mean}(w)}{d(n)} = 1$.

**Proof.** To keep notation simple, we sometimes omit the explicit dependence on $n$ of $M$, $d$, and other variables to be introduced in the proof.

Firstly, note that (12) implies that the largest expected degree is
\[
w_1 = c(i_0 + 1)^{-p} = (1 - p)d(n) \left( n \left( \frac{(1 - p)d(n)}{M(n)} \right)^{1/p} \right)^{-p} = M.
\]

By hypotheses, we can write $i_0 + 1 = n\theta$ where $\theta = \theta(n)$ is a function of $n$ such that $\theta(n) > 0$ and $\lim_{n \to \infty} \theta(n) = 0$. Define
\[
\epsilon_n = \sum_{i=1}^{n} (i_0 + i)^{-p} - \int_{i_0}^{n+i_0+1} x^{-p} \, dx.
\]
From (1) we have $\epsilon_n > 0$ and

$$\epsilon_n \leq (i_0 + 1)^{-p} - \int_{n+i_0}^{n+i_0+1} x^{-p} \, dx < (i_0 + 1)^{-p}.$$  

In particular $c\epsilon_n < M$. Henceforth, from (11), we have $c\epsilon_n/nd = O(1/\sqrt{nd})$. Moreover,

$$\sum_{i=1}^{n} w_i = c \sum_{i=1}^{n} (i + i_0)^{-p} = c \left( \int_{i_0+1}^{n+i_0+1} x^{-p} \, dx + \epsilon_n \right)$$

$$= \frac{c}{1-p} \left( (n + i_0 + 1)^{1-p} - (i_0 + 1)^{1-p} \right) + c\epsilon_n$$

$$= d\eta \left( (1+\theta)^{1-p} - \theta^{1-p} \right) + c\epsilon_n.$$  

Hence,

$$\lim_{n \to \infty} \frac{1}{nd(n)} \sum_{i=1}^{n} w_i = \lim_{n \to \infty} \left( (1+\theta)^{1-p} - \theta^{1-p} \right) + \frac{c\epsilon_n}{nd} = 1$$

and we get the first part of the claim. Finally, if (11) holds then for sufficiently large $n$ we have

$$\sum_{i=1}^{n} w_i \geq nd(1-\theta^{1-p}) \geq nd\eta \geq M^2 = w_1^2.$$  

Hence $w \in \mathcal{A}_n$ and the proof is complete.

In Figure 3 we show a number of statistics some statistics on random graphs built by the algorithm in Figure 1 according to the formulas in (12) with $n = 10000$, $d = 15$ and $M = n^{0.45}$. The following table shows the values of other relevant parameters. Due to Theorem 2, the resulting vectors $w$ are admissible.

| $\gamma$ | $i_0$  | $w_{n_0}$ |
|----------|--------|------------|
| 2.1      | 146.29 | 1.3458     |
| 2.9      | 156.76 | 7.0473     |
Fig. 4: Average degree in scale-free Chung–Lu random graphs of sizes growing from $10^2$ to $10^6$. Each color represent a different exponent for the power law: red crosses for $\gamma = 2.2$, blue circles for $\gamma = 2.5$ and green triangles for $\gamma = 3$. Each point represents a different graph obtained from the formula (12) choosing $d = 2$ and $M(n) = \sqrt{n}$ (left) or $d = 10$ and $M(n) = 5\sqrt{n}$ (right).

The first and third panel in Figure 3 display both the expected (red line) and the actual (blue dots) degree profile of the graphs. The vertical dashed line indicates the value of $i_0$. The relevance of that value is clearly visible: Nodes whose index is less than $i_0$ have comparable degrees, whereas the largest degree variation is produced by the remaining nodes. The point where the vertical line crosses the continuous red line sets up the scale of the largest hubs in the network. The second and fourth panel show the degree distributions, that is, $n_k$ vs $k$ (blue dots), together with the expected power law (red line). The abscissa of the vertical dashed line is $w_n$, which bounds from below the range where the degree distribution is expected to follow the power law. In fact, the number of nodes whose degree is smaller than $w_n$ departs ostensibly from the $O(k^{-\gamma})$ behavior.

It is worth noting that the hypothesis (10) is not very restrictive. For example, for networks in the Barabási–Albert model the ratio between maximum and average degree is typically $O(\sqrt{n})$, and in real networks the degree of hub nodes is far greater than the average. Moreover, the condition (11) is almost optimal. Indeed, let mean($w$) = $(\sum_i w_i)/n$ be the expected average degree of a graph from $G(w)$ and note that the condition $w \in A_n$ is equivalent to $M^2 = w^2 \leq \sum_i w_i = n \text{mean}(w)$. Hence, if (11) is violated then no arbitrary large admissible vector $w$ can be obtained.

On the other hand, the condition (11) is quite stringent, at least in some scenarios. For example, if $d(n)$ is upper bounded by a constant then that condition implies that the largest expected degree $M(n)$ must grow not faster than $\sqrt{n}$. However, we observed in (13) that if the degree profile is a power law then the largest degree behaves as $d_{\text{max}} = O(n^p)$. Hence, the estimate (13) can be attained for large $n$ only if $p \leq \frac{1}{2}$, that is $\gamma \geq 3$.

Finally, notice that, from equation (13) it is clear that the average expected degree mean($w$) converges towards $d(n)$ as $O(\theta(n)^{1-p})$. Then, if $p$ is approximately 1, i.e., $\gamma$ is close to 2, the convergence can be very slow. This is shown in Figure 4 where we plot the average degree of different networks generated by the algorithm in Figure 1 as a function of the dimension of the graph $n$. As we can see from the graph, as $n$ increases the average degree converges from below to the parameter $d$ of the formula (12). Moreover, the convergence is faster when $\gamma$ is bigger (i.e. $p$ is smaller).
3.1 Conditions for the giant component

A very relevant element in the analysis of random graphs is the presence of giant components. It is customary to say that a network has a giant component if there is a connected subgraph comprising a significant fraction of all the nodes. More precisely, in a network whose number of nodes $n$ increases over time, a giant component is a connected subgraph whose size is $\mathcal{O}(n)$. Aiello, Chung and Lu [1] obtained very detailed results concerning existence, uniqueness and the size not only of the giant component, but also of the smaller connected components of random power-law graphs. In particular, they showed that a random power law graph with exponent $\gamma < \gamma_0 \approx 3.47875$ almost surely has a unique giant component.

However, the results in [1] are based on the fundamental assumption that the largest degree in a random power-law graph with $n$ nodes and exponent $\gamma > 1$ is roughly proportional to $n^{1/\gamma}$. That assumption is questionable. For example, in [3, §4.3] and [25, §3.3.2] it is argued that the largest degree behaves as $n^{1/(\gamma-1)}$, as we also discussed in (5).

More generally, for Chung–Lu random graphs with a generic expected degree sequence, a giant component appears if the expected average degree is larger than 1. Indeed, the following result holds, see [10] and [9, Thm. 6.14].

**Theorem 3.** Let $w \in A_n$, mean($w$) = $(\sum_i w_i)/n$ and $\text{mean}_2(w) = (\sum_i w_i^2)/(\sum_i w_i)$. If mean($w$) > 1 then almost surely a random graph $G \in G(w)$ has a unique giant component, which contains $n(\lambda_0 + o(1))$ mean($w$) edges, where $\lambda_0$ is the unique positive root of the equation

$$\sum_i w_i e^{-w_i \lambda} = (1 - \lambda) \sum_i w_i.$$

Moreover, if $\text{mean}_2(w) < 1 - \varepsilon$ for some $\varepsilon > 0$ independent on $n$ then almost surely there is no giant component in $G$.

Note that, in the preceding claim, it holds mean($w$) ≤ $\text{mean}_2(w)$ due to the Cauchy–Schwartz inequality. Whilst nothing is known about the existence of a giant component when mean($w$) < 1 < $\text{mean}_2(w)$, it is important to understand if it is possible to generate arbitrarily large networks with average degree greater than 1. In fact, from the simple inequality

$$\frac{\sum_i w_i^2}{\sum_i w_i} \leq \frac{\sum_i w_i^2}{\sum_i w_i} = \text{mean}_2(w)$$

it is immediate to derive that if $\text{mean}_2(w) < 1$ then $w \in A_n$. Thus, only a rather small subset of admissible vectors can produce a giant component. In particular, the construction defined in Theorem 1 severely constrains the possibility of having a giant component, as shown in the following result.

**Corollary 2.** Let $0 < p \leq 1$ and let the vector $w \in A_n$ be defined as $w_i = ci^{-p}$ for $i = 1, \ldots, n$. If mean($w$) > 1 then either $p \leq \frac{1}{2}$ or $\frac{1}{2} < p < 1$ and

$$n < 1/(1 - p)^{1/p}.$$

---

1 In probability theory, it is customary to say that a property depending on an integer $n$ holds almost surely if the probability that it holds tends to 1 as $n$ goes to infinity.

2 In [10] and [9, Thm. 6.14], the hypothesis $w \in A_n$ is tacitly assumed.
Proof. Firstly, we exclude the case $p = 1$. Indeed, in that case Corollary 1 implies $c \leq \log n + 1$, and for $n > 1$ it holds

$$\text{mean}(w) = \frac{c}{n} \sum_{i=1}^{n} i^{-1} \leq \left(\frac{\log n + 1}{n}\right)^2 < 1.$$ 

For $0 < p < 1$, again by Corollary 1 we have $c < n^{1-p}/(1-p)$. Hence,

$$\text{mean}(w) = \frac{c}{n} \sum_{i=1}^{n} i^{-p} < \frac{n^{-p}}{1-p} \left(1 + \int_{1}^{n} x^{-p} \, dx\right) < \frac{n^{1-2p}}{(1-p)^2}.$$ 

Letting $\text{mean}(w) > 1$ we obtain

$$n^{1-2p} > (1-p)^2.$$ 

(14)

This inequality holds for every $n \in \mathbb{N}$ if and only if $p \leq \frac{1}{2}$. On the other hand, if $p > \frac{1}{2}$ then (14) yields the upper bound $n < (1-p)^{2/(1-2p)}$ and the theorem is proved.

On the other hand, using the formula $w_i = c i^{-1/(\gamma-1)}$ for $\gamma \geq 3$, we can obtain vectors $w \in A_n$ with arbitrary large $n$ and such that the graphs in $G(w)$ have a prescribed expected average degree $d > 1$. Indeed, let $p = 1/(\gamma - 1)$ and $c = d(1-p)n^p$. If $0 < p < 1/2$, then for sufficiently large $n$ it holds

$$d(1-p)n^p \leq \frac{(n+1)^{1-p} - 1}{1-p},$$ 

so we have $w \in A_n$ by Theorem 1. Moreover, using essentially the same argument as the one of the proof of Theorem 2, it is not difficult to observe that the expected average degree of a graph in $G(w)$ is asymptotically equal to $d$. The same conclusions carry over the case $p = 1/2$, that is $\gamma = 3$, under the additional constraint $d \leq 4$. It is worth noting that, with this construction, we have

$$\frac{w_1}{w_n} = (n+1)^p,$$

that is, the ratio between the expected largest and smallest degrees behaves as the analytical estimate obtained in 5.

Altogether, the next result provides a way to construct sequences of large scale-free Chung-Lu random graphs with power law degree distribution $n_k = \alpha k^{-\gamma}$ and expected average degree larger than one, for every exponent $\gamma > 2$.

**Corollary 3.** Let $0 < p < 1$ and let $d$ be a fixed number larger than one. For $n \in \mathbb{N}$ define

$$M(n) = n^{(1-\varepsilon)/2},$$ 

where $0 < \varepsilon < 1$. Moreover, for $i = 1, \ldots, n$ let $w_i = c(i_0 + i)^{-p}$ where $c$ and $i_0$ are defined as in 12. Then for sufficiently large $n$ one has $w \in A_n$ and the expected average degree of a random graph from $G(w)$ is asymptotically $d$.

**Proof.** The limit (10) is trivially fulfilled. Moreover, $M(n)^2 = n^{1-\varepsilon} \leq n$, hence also the inequality (11) is valid with $\eta = 1/d$. Then the claim is a straightforward consequence of Theorem 2.
3.2 Avoiding potentially isolated nodes

In the analysis of complex networks, isolated nodes are of no interest and are usually discarded. Actually, the power law itself is ill-defined for zero-degree nodes. Since the parameters of the Chung–Lu model specify expected degrees, and not actual degrees, it is always possible that some nodes in a random graph $G \in \mathcal{G}(\mathbf{w})$ are isolated. In fact, if the graph $G$ is large and sparse then the effective degree of node $i$ is closely approximated by a Poisson random variable with parameter $w_i$, see e.g., [19] or [9, p. 110], so the probability of $i$ being isolated is about $e^{-w_i}$. Anyway, the chance of having $d_i = 0$ in a random graph drawn from $G(\mathbf{w})$ is obviously high when the expected value of $d_i$ is less than 1. Hence we say that node $i$ is potentially isolated if $w_i < 1$. In this paragraph we address the problem of avoiding the presence of potentially isolated nodes when the vector $\mathbf{w}$ is computed as in Theorem 2. For simplicity, we consider initially the case where $i_0 = 0$. Subsequently, we approach the general case. We introduce the following definition.

Definition 2. A vector $\mathbf{w} \in \mathcal{A}_n$ is called strongly admissible if $w_n \geq 1$.

It is immediate to see that if $\mathbf{w}$ is strongly admissible then the expected average degree of $G \in \mathcal{G}(\mathbf{w})$ is greater than 1. Hence, a direct consequence of Corollary 2 is that no strongly admissible vector can be obtained by the formula $w_i = ci^{-p}$ when $p = 1$, that is $\gamma = 2$. The case $p \neq 1$ is dealt with in the next result, which proves that a strongly admissible vector $\mathbf{w}$ of the form $w_i = ci^{-p}$ can exist for large $n$ only if $\gamma \geq 3$.

Corollary 4. Let $p > 0$, $p \neq 1$, and let the vector $\mathbf{w} \in \mathcal{A}_n$ be defined as $w_i = ci^{-p}$ for $i = 1, \ldots, n$. If $\mathbf{w}$ is strongly admissible then either $p \leq \frac{1}{2}$ or $p > \frac{1}{2}$ and $n < n_{\text{max}}$ where

$$n_{\text{max}} = \begin{cases} (1-p)^{-\frac{1}{2p-1}} & 0 < p < 1, \\ (p-1)^{-1/p} & p > 1. \end{cases}$$  \hspace{1cm} (15)

Proof. When $p > 1$, that is $1 < \gamma < 2$, the claim follows directly from the condition $1 \leq w_n = cn^{-p} \leq n^{-p}/(p-1)$ coming from Corollary 1. When $0 < p < 1$, again by Corollary 1 we have $c < n^{-p}/(1-p)$. Hence,

$$w_n = cn^{-p} < \frac{n^{1-p}}{1-p}n^{-p} = \frac{n^{1-2p}}{1-p}.$$  \hspace{1cm} (16)

Letting $w_n \geq 1$ we obtain

$$n^{1-2p} > 1-p.$$  \hspace{1cm} (17)

This inequality holds for every $n \in \mathbb{N}$ if and only if $p \leq \frac{1}{2}$. On the other hand, if $w_n \geq 1$ and $p > \frac{1}{2}$ then (17) yields the upper bound in the claim. \hfill \square

The upper bound on $n$ defined by Corollary 4 severely constrains the size of a scale-free network without potentially isolated nodes. Indeed, Figure 3 illustrates the behavior of $n_{\text{max}}$ defined in (15) as a function of $\gamma = 1 + 1/p$. The graph of $n_{\text{max}}$ has a needle-like appearance in a neighborhood of $\gamma = 2$, where it is undefined, while lies in the order of a few thousands when $\gamma$ is in a rather wide range within the interval $(2,3)$.

Anyway, also in this case, the introduction of the additional parameter $i_0$ gives us more freedom in the construction of scale-free networks. In fact, our next results show that if we consider expected degree vectors $\mathbf{w}$ with entries $w_i = c(i_0 + i)^{-1/(\gamma-1)}$, then $G(\mathbf{w})$ can model arbitrarily large scale-free networks without any potentially isolated vertex.
1.2 1.4 1.6 1.8 2 2.2 2.4 2.6 2.8 3
10
0
10
5
Fig. 5: Plot of $n_{\text{max}}$ given by (15) for $1 < \gamma < 3$.

In the sequel, the entries of the vector $w \in A_n$ will be numbered as $w_0, \ldots, w_{n-1}$ rather than $w_1, \ldots, w_n$, while keeping the expression $w_i = c(i + i_0)^{-p}$. We adopt this renumbering to simplify notations in our final results. For example, the largest entry now is $w_0 = c_i - p_0$ and the inequality (8) is replaced by

$$c \leq i_0^{2p}(n + i_0)^{1-p} - i_0^{1-p} - p.$$ 

Accordingly, we assume $i_0 > 0$.

**Theorem 4.** Let $0 < p < 1$ and let $i_0 > 0$ be such that

$$i_0 \leq \frac{n}{2^{1/(1-p)} - 1},$$

$$i_0(i_0^{1/p} - 1) \geq n.$$ 

Moreover, let

$$c = \kappa i_0^{2p}((n + i_0)^{1-p} - i_0^{1-p}), \quad 1 \leq \kappa \leq 1/(1-p).$$

Then the vector $w \in \mathbb{R}^n$ with $w_i = c(i + i_0)^{-p}$ for $i = 0, \ldots, n-1$ is strongly admissible.

**Proof.** From (19) we have that $w \in A_n$, by the same passages carried out in (9). To complete the proof we have to prove that $w_{n-1} = c(n - 1 + i_0)^{-p} \geq 1$. Actually, we will show the stronger inequality $c(n + i_0)^{-p} \geq 1$. Indeed, simple passages yield the identity

$$c(n + i_0)^{-p} = \kappa i_0^{2p}((n + i_0)^{1-p} - i_0^{1-p})(n + i_0)^{-p}$$

$$= \kappa \frac{i_0}{(i_0 + n + 1)^{1/p}} \left( \frac{n + 1}{i_0 + 1} \right)^{1-p} - 1) = \kappa i_0^p(i^{1-p} - 1),$$

where we have set $r = (n/i_0) + 1$. Now, it is easy to verify that (17) implies $r^{1-p} \geq 2$. Moreover, from (18) we derive the inequality $r \leq i_0^{1/p}$. Finally, we obtain $c(n + i_0)^{-p} \geq \kappa \geq 1$ and the proof is complete.

On the basis of the previous theorem, we provide in the next corollary an explicit construction to generate large scale-free networks without potentially isolated nodes and arbitrary exponent $\gamma > 2$.

**Corollary 5.** For $0 < p < 1$ and $\eta \geq 1$ let $i_0 = \eta n^{\nu/(1+p)}$. Moreover, let $c$ be as in (19). Then for $n$ sufficiently large the vector $w \in \mathbb{R}^n$ with $w_i = c(i + i_0)^{-p}$ for $i = 0, \ldots, n-1$ is strongly admissible.
Proof. Since \( i_0 \) grows slower than \( n \), there exists \( N_1 \in \mathbb{N} \) such that for all \( n \geq N_1 \) the inequality (17) is fulfilled. Moreover,

\[
\lim_{n \to \infty} \frac{i_0(i_0^{1/p} - 1)}{n} = \lim_{n \to \infty} \frac{i_0^{(p+1)/p}}{n} = \eta^{(p+1)/p} \geq 1.
\]

Then there exists \( N_2 \in \mathbb{N} \) such that for all \( n \geq N_2 \) it holds \( i_0(i_0^{1/p} - 1)/n \geq 1 \), which is (18). Taking \( N = \max\{N_1, N_2\} \) all the hypotheses of Theorem 4 are fulfilled, and the claim follows.

We conclude with the following example showing that, under certain restrictions, the proposed construction can produce random scale-free networks similar to those produced by the Barabási–Albert model. For \( i = 1, \ldots, n \) consider the vector \( w \) with entries \( w_i = c^{-1/2} \) where \( 0 < c \leq 2\sqrt{n} + 1 - 2 \). By Theorem 4 \( w \in A_n \) and a graph \( G \in G(w) \) has an expected degree distribution following a power law with exponent \( \gamma = 3 \) and expected average degree \( d = 2c/\sqrt{n} \). In particular, the ratio between the largest and the average degree is approximately \( \sqrt{n}/2 \), as in the Barabási–Albert model. Furthermore, if \( c \geq \sqrt{n} \) then \( w \) is strongly admissible.

4 Conclusions

Power-law degree distributions are ubiquitous in real-world networks. The observed exponent often falls in the range \( 2 < \gamma < 3 \), which is the range where growing networks can show small average degree together with a non negligible quantity of hub nodes. Here, we investigate the possibility of generating large random graphs having a power-law degree distribution using the Chung–Lu generative model. This model is defined in terms of \( n \) parameters, \( w_1, \ldots, w_n \), where \( n \) is the number of nodes, and is particularly relevant as it is the only random graph model that does not introduce correlations between the degrees of two nodes connected by an edge. Under an appropriate condition, here called admissibility, the parameters \( w_i \) determine the node degrees, in expectation.

In this work, we analyze the possibility of using admissible parameters \( w_i \) to generate large networks having an expected power-law degree distribution, a giant component and no isolated nodes. Not surprisingly, the admissibility condition imposes severe restrictions to the resulting degree profile and, in some cases, also to the network size. In particular, we prove that, under appropriate conditions, the general formula \( w_i = c(i_0 + i)^{-p} \) can serve to generate arbitrarily large random graphs having a power-law degree profile with exponent \( \gamma = 1 + 1/p \), at least within certain exponent and degree ranges. Whilst our main focus is on the exponent range \( 2 < \gamma < 3 \), we also considered rather general exponents. Our main results provide explicit formulas for the coefficients \( c \) and \( i_0 \), in order to fulfil a number of desirable requirements on the connectivity of the network and the behavior of the smallest, largest, and average degrees, in expectation.

Acknowledgements

The work of DF and AT has been carried out in the framework of the departmental research project “ICON: Innovative Combinatorial Optimization in Networks”, Department of Mathematics, Computer Science and Physics (PRID 2017), University of Udine, Italy. Moreover, DF and FT have been partly supported by INdAM-GNCS.
References

[1] W. Aiello, F. Chung, and L. Lu. A random graph model for power law graphs. *Experiment. Math.*, 10:53–66, 2001.

[2] R. Albert and A. L. Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74:47–97, 2002.

[3] A. L. Barabási. *Network Science*. Cambridge University Press, 2016.

[4] V. Batagelj and U. Brandes. Efficient generation of large random networks. *Phys. Rev. E*, 71:036113, 2005.

[5] B. Bollobás and O. M. Riordan. Mathematical results on scale-free random graphs. In S. Bornholdt and H. G. Schuster, editors, *Handbook of Graphs and Networks*, pages 1–34. Wiley, 2005.

[6] F. Chung and L. Lu. The diameter of random sparse graphs. *Advances in Applied Math.*, 26:257–279, 2001.

[7] F. Chung and L. Lu. Connected components in random graphs with given expected degrees. *Ann. Combinatorics*, 6:125—145, 2002.

[8] F. Chung and L. Lu. The average distance in a random graph with given expected degrees. *Internet Mathematics*, 1:91–114, 2003.

[9] F. Chung and L. Lu. *Complex Graphs and Networks*. Number 107 in CBMS. Amer. Math. Soc., 2004.

[10] F. Chung and L. Lu. The volume of the giant component of a random graph with given expected degrees. *SIAM J. Discrete Math.*, 20:395–411, 2006.

[11] F. Chung, L. Lu, and V. Vu. The spectra of random graphs with given expected degrees. *Internet Mathematics*, 1:257–275, 2003.

[12] A. Clauset, C. Rohilla Shalizi, and M. G. J. Newman. Power-law distributions in empirical data. *SIAM Rev.*, 51:661–703, 2010.

[13] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin. Structure of growing networks with preferential linking. *Phys. Rev. Letters*, 85:4633–4636, 2000.

[14] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. *SIGCOMM Comput. Commun. Rev.*, 29(4):251–262, 1999.

[15] D. Fasino and F. Tudisco. An algebraic analysis of the graph modularity. *SIAM J. Matrix Anal. Appl.*, 35:997–1018, 2014.

[16] D. Fasino and F. Tudisco. The expected adjacency and modularity matrices in the degree corrected stochastic block model. *Spec. Matrices*, 6:110–121, 2018.

[17] D. Fasino and F. Tudisco. A modularity based spectral method for simultaneous community and anti-community detection. *Linear Algebra Appl.*, 542:605–623, 2018.
[18] B. Karrer and M. E. J. Newman. Stochastic blockmodels and community structure in networks. *Phys. Rev. E*, 83:016107, 2011.

[19] T. G. Kolda, A. Pinar, T. Plantenga, and C. Seshadhri. A scalable generative graph model with community structure. *SIAM J. Sci. Comput.*, 36:C424–C452, 2014.

[20] P. L. Krapivsky and S. Redner. Organization of growing random networks. *Phys. Rev. E*, 63:066123, 2001.

[21] P. L. Krapivsky, G. J. Rodgers, and S. Redner. Degree distributions of growing networks. *Phys. Rev. Lett.*, 86:5401, 2001.

[22] A. Litvak and R. van der Hofstad. Uncovering disassortativity in large scale-free networks. *Phys. Rev. E*, 87:022801, 2013.

[23] J. C. Miller and A. Hagberg. Efficient generation of networks with given expected degrees. In *Proceedings of the 8th International Conference on Algorithms and Models for the Web Graph*, WAW’11, pages 115–126. Springer-Verlag, 2011.

[24] C. Moore, G. Ghoshal, and M. E. J. Newman. Exact solutions for models of evolving networks with addition and deletion of nodes. *Phys. Rev. E*, 74:036121, 2006.

[25] M. E. J. Newman. The structure and function of complex networks. *SIAM Rev.*, 45:167–256, 2003.

[26] M. E. J. Newman. Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E*, 69:321–330, 2006.

[27] M. E. J. Newman. *Networks — An Introduction*. Oxford University Press, 2010.

[28] M. E. J. Newman, D. J. Watts, and S. H. Strogatz. Random graph models of social networks. *Proc. Natl Acad. Sci. USA*, 99:2566–2572, 2002.

[29] N. Przulj and D. J. Higham. Modelling protein-protein interaction networks via a stickiness index. *J. Roy. Soc. Interface*, 3:711–716, 2006.

[30] A. S. Ramani, N. Eikmeier, and D. F. Gleich. Coin-flipping, ball-dropping, and grass-hopping for generating random graphs from matrices of edge probabilities. *SIAM Rev.*, 61:549–595, 2019.

[31] A. Taylor and D. J. Higham. CONTEST: A controllable test matrix toolbox for MATLAB. *ACM Transactions on Mathematical Software*, 35:26:1–26:17, 2009.