Mathematical results on scale-free random graphs

B. Bollobás

January 16, 2003

1 Introduction

Recently there has been much interest in studying large-scale real-world networks and attempting to model their properties using random graphs. Although the study of real-world networks as graphs goes back some time, recent activity perhaps started with the paper of Watts and Strogatz [55] about the 'small-world phenomenon'. Since then the main focus of attention has shifted to the 'scale-free' nature of the networks concerned, evidenced by, for example, power-law degree distributions. It was quickly observed that the classical models of random graphs introduced by Erdős and Rényi [28] and Gilbert [33] are not appropriate for studying these networks, so many new models have been introduced. The work in this field falls very roughly into the following categories.

1. Direct studies of the real-world networks themselves, measuring various properties such as degree-distribution, diameter, clustering, etc.

2. Suggestions for new random graph models motivated by this study.

3. Computer simulations of the new models, measuring their properties.

4. Heuristic analysis of the new models to predict their properties.

5. Rigorous mathematical study of the new models, to prove theorems about their properties.

Although many hundreds of interesting papers have been written in this area (see, for example, the surveys [2, 27]), so far almost all of this work comes under 1-4; to date there has been very little rigorous mathematical work in the field. Our main aim in this article is to present some of this mathematical work, including several new results. Even an overview of the work in 1-4 lies outside our scope, so we shall present only those models which have been made mathematically precise and for which results have been proved, and mention only a few heuristic results for comparison with the theorems we present. For similar reasons, we cannot even survey the 'classical' theory of random graphs, referring the reader instead to [11] and [38]. However, we shall briefly describe the classical models, as well as some results relevant for comparison; much of the work on the new models has appeared in computer science and physics journals, and it may be that some of the authors are not aware of the related classical results.

The rest of this article is organized as follows. In the next section we briefly describe the classical models of random graphs. In section 3 we state some theorems about these models chosen for comparison with recent results about the new models. Section 4 is a brief digression concerning the Watts-Strogatz 'small-world' model. The rest of the article concerns 'scale-free' models; a brief introduction is given in section 5. These models fall into two types. The first takes a power-law degree distribution as given, and then generates a graph with this distribution. Such models will not be considered here. The second type arises from attempts to *explain* the power law starting from basic assumptions about the growth of the graph. In section 6 we describe the Barabási-Albert (BA) model, noting that their definition does not make mathematical sense. A precisely defined model, the 'LCD model', along the lines of the BA model is described in section 7, followed by a generalization due to Buckley and Osthus [20] in the next section. In these and the following few sections we concentrate on the degree distribution, presenting results showing that the models are indeed scale-free. Sections 9 and 10 present such results for the 'copying' models of Kumar, Raghavan, Rajagopalan, Sivakumar, Tomkins and Upfal [40], and the very general models defined by Cooper and Frieze [24]. Section 11 describes a model for directed graphs with 'preferential attachment' using both in- and out-degrees, and gives the power laws for in- and out-degree distribution.

At this point we return to the LCD model, presenting results about properties other than degree sequence: the clustering coefficient is discussed in section 12, the diameter in section 13 and 'robustness' in section 14.

The last section concerns a special case of the BA model that had been studied considerably earlier; that of scale-free trees. In section 15, we present results for small subgraphs (useful for the LCD model) and distance distribution.

Finally, in section 16 we conclude with a few remarks.

2 Classical models of random graphs

The theory of random graphs was founded by Erdős and Rényi in a series of papers published in the late 1950s and early 1960s. Erdős and Rényi set out to investigate what a 'typical' graph with n labelled vertices and M edges looks like. They were not the first to study statistical properties of graphs; what set their work apart was the probabilistic point of view: they considered a probability space of graphs and viewed graph invariants as random variables. In this setting powerful tools of probability theory could be applied to what had previously been viewed as enumeration questions.

Throughout this section, and indeed the rest of this article, we consider models of *labelled* graphs. Although in the end one may choose to ignore the labels, the models are naturally defined as generating graphs on a certain set of distinguishable vertices, rather than isomorphism classes of graphs. For definiteness it is often convenient to assume that, when the graph has n vertices, the vertex set is $[n] = \{1, 2, \dots, n\}$.

In modern notation Erdős and Rényi considered the space $\mathcal{G}_{n,M}$ of all $\binom{N}{M}$ graphs with vertex set $[n]$ having M edges, where $N = \binom{n}{2}$ is the number of all possible edges between vertices in $[n]$. The set $\mathcal{G}_{n,M}$ is made into a probability space by taking the elements of $\mathcal{G}_{n,M}$ equiprobable; $G_{n,M}$ will denote a random element of this space. We are interested in what happens as $n \to \infty$, with $M = M(n)$ a function of n. We say that $G_{n,M}$ has a certain property $\mathcal{P}$ *with high probability* (**whp**) if

$$\Pr(G_{n,M} \text{ has } \mathcal{P}) \to 1$$

as $n \to \infty$. (Here and in what follows it is always understood that M is a function of n. The case when M is constant as $n \to \infty$ is rather uninteresting.) Following Erdős and Rényi, it is customary to say that a *typical* random graph $G_{n,M}$ has property $\mathcal{P}$ if $G_{n,M}$ has $\mathcal{P}$ **whp**.

One of the main discoveries of Erdős and Rényi was that as $M = M(n)$ increases, the structure of a typical $G_{n,M}$ tends to change suddenly. The following is a simple but fundamental result from [28] about connectedness.

Theorem 1. *Let $M_\omega = \frac{n}{2}(\log n + \omega)$, where $\omega = \omega(n)$ is a function of n. If $\omega \to -\infty$ then a typical G_{n,M_ω} is disconnected, while if $\omega \to \infty$, a typical G_{n,M_ω} is connected.*

In the 1950s, Austin, Fagen, Penney and Riordan [4], Gilbert [32, 33], and Riddell and Uhlenbeck [50] also studied statistical properties of graphs, but their approach was very different, using generating functions to obtain exact enumeration formulae and then approximating these. The results obtained this way were much weaker than those of Erdős and Rényi.

The model of random graphs introduced by Gilbert [33] (precisely at the time that Erdős and Rényi started their investigations of $\mathcal{G}_{n,M}$) is, perhaps, even more fundamental than $\mathcal{G}_{n,M}$, and is more convenient to use. To define Gilbert's model, $\mathcal{G}_{n,p}$, let $\{X_{ij} : 1 \le i < j \le n\}$ be an array of iid Bernoulli random variables, with $\Pr(X_{ij} = 1) = p$ and $\Pr(X_{ij} = 0) = 1 - p$, and let $G_{n,p}$ be the random graph on $[n]$ in which two vertices i and j are adjacent if $X_{ij} = 1$. Less formally, to construct a random $G_{n,p} \in \mathcal{G}_{n,p}$, put in edges with probability p, independently of each other. Again p is often a function of n, though the case p constant, $0 < p < 1$, makes perfect sense. For $M \sim pN$ the models $\mathcal{G}_{n,M}$ and $\mathcal{G}_{n,p}$ are almost interchangeable. (Note that, as usual, we commit a harmless abuse of notation, using $\mathcal{G}_{n,\cdot}$ for two different models. There is no danger of confusion, as $M \to \infty$ while $0 < p < 1$.)

Since the early 1960s several other 'classical' models of random graphs have been introduced. A *graph process* $\tilde{G}_n = (G_{n,t})_{t=0}^N$ on $[n]$ is a nested sequence of graphs, $G_{n,0} \subset G_{n,1} \subset \cdots \subset G_{n,N}$ such that $G_{n,t}$ has precisely t edges. The space $\tilde{\mathcal{G}}_n$ of *random graph processes* consists of all $N!$ graph processes on $[n]$, endowed with the uniform (normalized counting) measure. Note that this notation is consistent with that used earlier: the distribution of $G_{n,t}$, a random graph process stopped at time t, is precisely the distribution of $G_{n,t}$ as an element of $\mathcal{G}_{n,t}$. A random graph process has a natural interpretation as a dynamic Markov process; given $G_{n,0}, \ldots, G_{n,t}$, at the next step $G_{n,t+1}$ is obtained by adding one of the $N-t$ remaining possible edges to $G_{n,t}$ uniformly at random. In studying $\tilde{\mathcal{G}}_n$ one is mostly interested in the *hitting times* of certain properties (those preserved by adding edges), that is, the random variable given by the minimal t for which $G_{n,t}$ has the property. For example, Theorem 1 claims that **whp** the hitting time of connectedness is at least $\frac{n}{2}(\log n - \omega(n))$ and at most $\frac{n}{2}(\log n - \omega(n))$ whenever $\omega(n) \to \infty$. In fact, **whp**, the hitting time of connectedness is precisely the hitting time of having no isolated (degree 0) vertices.

To get a random element $G_{n,k\text{-out}}$ of the space $\mathcal{G}_{n,k\text{-out}}$, join each vertex i to k other vertices chosen at random and take the union of all these edges. Equivalently, let $\vec{G}_{n,k\text{-out}}$ be the random directed graph obtained by sending arcs from each vertex to a set of k other vertices chosen uniformly at random; the random graph $G_{n,k\text{-out}}$ is the underlying simple graph of $\vec{G}_{n,k\text{-out}}$. Note that each $G_{n,k\text{-out}}$ has at least $kn/2$ and at most kn edges; although kn is much smaller than $\frac{n}{2}\log n$, the threshold of connectedness given by Theorem 1, for all $k \geq 2$, **whp** $G_{n,k\text{-out}}$ is connected.

The space $\mathcal{G}_{n,r\text{-reg}}$ is simply the set of all r-regular graphs on $[n]$ with the uniform measure. Although this space is very easy to define, for larger values of r it is not so easy to study.

The study of random graphs really took off in the mid 1970s; since then several thousand papers have been written on the topic. Many of the results are presented in the monographs [11] and [38].

3 Results for classical random graphs

In this brief review it would be impossible to survey even the more important results about classical random graphs; all we shall do is present some results that are analogous to a number of results about scale-free random graphs we shall present later.

In addition to discovering the prevalence of 'phase transitions' for numerous properties of random graphs, Erdős and Rényi [29] proved that the component structure of a random graph process undergoes a sudden change around time $t \sim n/2$. This result about the emergence of the 'giant component' is the single most important theorem of Erdős and Rényi about random graphs. Here we state it only in a simplified form.

Theorem 2. *Let $c > 0$ be a constant, and set $p = c/n$. If $c < 1$ then* **whp** *every component of $G_{n,p}$ has order $O(\log n)$. If $c > 1$ then* **whp** *$G_{n,p}$ has a component with $(\alpha(c) + o(1))n$ vertices, where $\alpha(c) > 0$, and all other components have $O(\log n)$ vertices.*

Considerably more precise results have been proved by Bollobás [10], Łuczak [42], and Janson, Knuth, Łuczak and Pittel [37]. The component of order $\Theta(n)$ whose existence is guaranteed by Theorem 2 is usually called the *giant component*. If

c is considerably larger than 1, then the giant component has a large robust (highly connected) subgraph.

For p constant, the degree sequence of $G_{n,p}$ is close to a sequence of n iid Binomial random variables with probability p and mean np. (A very strong precise result along these lines is given in [46].) For $p = c/n$, where c is constant, the degree sequence is well approximated by a sequence of n iid Poisson random variables with mean c. In particular, one has the following very weak result.

Theorem 3. *Let X_k be the number of vertices of degree k in $G_{n,p}$ where $p = c/n$, with $c > 0$ constant. Then for $k = 0, 1, \ldots$*

$$\Pr\left((1-\epsilon)\frac{c^k e^{-c}}{k!} \leq \frac{X_k}{n} \leq (1+\epsilon)\frac{c^k e^{-c}}{k!}\right) \to 1$$

as $n \to \infty$.

In a graph G, the *distance* $d(u, v)$ between two vertices u and v is the length (number of edges) of the shortest path between them. The *diameter* $\mathrm{diam}(G)$ of a connected graph G is the maximum distance between two vertices; a disconnected graph is taken to have infinite diameter. The diameter of a random graph has been studied by a great many people, including Burtin [21, 22], Bollobás [9] and Bollobás and de la Vega [14]. If $pn/\log n \to \infty$ and $\log n/\log(pn) \to \infty$ then **whp** the diameter of $G_{n,p}$ is asymptotic to $\log n/\log(pn)$. In the range we are interested in here, corresponding to the $\Theta(n)$ edges in scale-free random graphs, $G_{n,p}$ is disconnected, so the the diameter of $G_{n,k\text{-out}}$ or $G_{n,r\text{-reg}}$ is more relevant. Let us state a weak form of a result from [14].

Theorem 4. *Let $r \geq 3$ and $\epsilon > 0$ be fixed. Then*

$$\Pr\left((1-\epsilon)\frac{\log n}{\log(r-1)} \leq \mathrm{diam}(G_{n,r\text{-}reg}) \leq (1+\epsilon)\frac{\log n}{\log(r-1)}\right) \to 1$$

as $n \to \infty$.

As we shall see, results vaguely resembling Theorem 4 hold for scale-free random graphs. More or less by definition, the results corresponding to Theorem 3 are rather different.

4 The Watts-Strogatz 'small-world' model

In 1998, Watts and Strogatz [55] raised the possibility of constructing random graphs that have some of the important properties of 'real-world' networks. The real-world networks they considered included neural networks, the power grid of the western United States and the collaboration graph of film actors. Watts and Strogatz noticed that these networks were 'small-world' networks: their diameters were considerably smaller than those of regularly constructed graphs (such as lattices, or grid graphs) with the same number of vertices and edges. More precisely, Watts and Strogatz found that real-world networks tend to be highly clustered, like lattices, but have small diameters, like random graphs. That large social networks have rather small diameters had been noticed considerably earlier, in the 1960s, by Milgram [47] and others, and was greatly popularized by Guare's popular play 'six degrees of separation' in 1990.

The importance of the Watts and Strogatz paper is due to the fact that it started the active and important field of modelling large-scale networks by random graphs defined by simple rules. As it happens, from a mathematical point of view, the experimental results in [55] were far from surprising.

Instead of the usual diameter $\mathrm{diam}(G)$ of a graph G, Watts and Strogatz considered the average distance

$$L(G) = \sum_{\{u,v\}\subset V, u\neq v} d(u,v)/\binom{n}{2},$$

where V is the vertex set of G and n is the number of vertices. Clearly $L(G) \leq \mathrm{diam}(G)$, but in 'most' cases $L(G)$ is not much smaller than $\mathrm{diam}(G)$. (For example, for $G_{n,r\text{-reg}}$, $r \geq 3$, **whp** these quantities are asymptotically equal.)

To measure the 'cliquishness' of a graph, for a graph G and vertex v, let $C_v(G)$ be the proportion of pairs of neighbours of v that are themselves neighbours, and let $C_1(G)$ be the average of $C_v(G)$ as v runs over the vertices. In section 12 we shall give a more formal definition of this *clustering coefficient* $C_1(G)$, together with a variant of it.

For a random r-regular graph, $C_1(G_{n,r\text{-reg}}) \sim \frac{r-1}{n}$, while

$$\mathrm{diam}(G_{n,r\text{-reg}}) \sim \log n/\log(r-1):$$

the clustering coefficient is small, and so is the diameter. On the other hand, as pointed out by Watts and Strogatz, many real-world networks tend to have a largish clustering coefficient *and* small diameter. To construct graphs with these properties, Watts and Strogatz suggested starting with a fixed graph with large clustering coefficient and 'rewiring' some of the edges.

To be precise, let G be the graph C_n^r, the r^{th} power of an n-cycle, where $n > 2r$. Thus G is a $2r$-regular graph of order n; two vertices are joined in G if their distance in the n-cycle C_n is at most r. For $n = 2rs$, $s \geq 2$, say, we have $\mathrm{diam}(G) = s$, and $L(G) \sim s/2$ as $s \to \infty$, while $C_1(G) = \frac{3(r-1)}{2(2r-1)}$. Let $G(p)$ be the random graph obtained from G by deleting each edge at random with probability p, independently of the other edges, and then adding the same number of edges back at random. Almost equivalently, $G(p)$ is obtained from G by 'rewiring' a proportion p of the edges. What Watts and Strogatz found was that, even for a small value of p, $L(G(p))$ drops down to $O(\log n)$, while $C_1(G(p))$ stays close to $3/4$; the introduction of a small number of random edges reduces the diameter to $O(\log n)$.

Following this observation, much research was devoted to the 'surprising' phenomenon that the introduction of a little randomness makes the diameter small (while various other graph invariants remain essentially unchanged). In fact, it is far from surprising that a few random edges superimposed on a connected ground graph give a graph of small diameter. For example, Bollobás and

Chung [13] proved that a random matching added to a cycle gives a graph whose diameter is about that of a random cubic graph. Similarly, for $c > 0$, adding cn random edges to a tree of order n results in a graph of diameter $O(\log n)$. These results (though not the precise constant given in [13]) are particular instances of a general phenomenon which has been known much longer; they follow from the fact that the diameter of $G_{n,r}$-reg (or of the giant component of $G_{n,p}$, $p = c/n$) is $O(\log n)$.

The graphs obtained by rewiring some of the edges of a power of a cycle do not resemble large-scale real-world networks, although they share some of their characteristics. To model these networks, it is desirable to define new families of random graphs rather different from from the classical models. This is the topic of the next several sections.

5 Scale-free models

In 1999, Faloutsos, Faloutsos and Faloutsos [30] suggested certain 'scale-free' power laws for the graph of the Internet, and showed that these power laws fit the real data very well. In particular, they suggested that the degree distribution follows a power law, in contrast to the Poisson distribution for classical random graphs given in Theorem 3. This was soon followed by work on rather vaguely described random graph models aiming to explain these power laws, and others seen in features of many real-world networks.

In fact, power-law distributions had been observed considerably earlier; in particular, in 1926 Lotka [41] claimed that citations in academic literature follow a power law, and in 1997 Gilbert [34] suggested a probabilistic model supporting 'Lotka's law'. Other early investigations into power-law distributions are due to Simon [51] and Zipf [56].

The degree distribution of the graph of telephone calls seems to follow a power law as well; motivated by this, Aiello, Chung and Lu [1] proposed a model for 'massive graphs'. This model ensures that the degree distribution follows a power law by *fixing* a degree sequence in advance to fit the required power law, and then taking the space of random graphs with this degree sequence. Thus their approach is very different from the models we are interested in, where the aim is to understand how power laws might arise, by finding simple rules that generate random graphs satisfying such laws.

In the next sections we present several of these models, concentrating for the moment on the degree sequence. Later in the article we return to one particular model, the LCD model, presenting results about several other properties.

6 The Barabási-Albert model

Perhaps the most basic and important of the 'scale-free' random graph models, i.e., models producing power-law or 'scale-free' behaviour from simple rules, is the 'BA model'. This was introduced by Barabási and Albert [5] in 1999:

> ... starting with a small number (m_0) of vertices, at every time step we add a new vertex with $m(\leq m_0)$ edges that link the new vertex to m different vertices already present in the system. To incorporate preferential attachment, we assume that the probability Π that a new vertex will be connected to a vertex i depends on the connectivity k_i of that vertex, so that $\Pi(k_i) = k_i/\sum_j k_j$. After t steps the model leads to a random network with $t+m_0$ vertices and mt edges.

The basic motivation is to provide a highly simplified model of the growth of, for example, the world-wide web. New sites (or pages) are added one at a time, and link to earlier sites chosen with probabilities depending on their current 'popularity'; this is the principle that 'popularity is attractive'; this principle presumably plays a role in the growth of real networks in a wide range of contexts. It is customary to call this the 'preferential attachment' rule. Barabási and Albert themselves, and many other people, gave experimental and heuristic results about the BA model; we will return to a few of these later. From a mathematical point of view, however, the description above, repeated in many papers, does not make sense.

The first problem is getting started: how do we take probabilities proportional to the degrees when these are all zero? Perhaps it makes sense to ignore the explicit start from no edges given by Barabási and Albert, and start instead from a small graph G_0 with no isolated vertices, hoping that the choice of G_0 makes little difference. While for many properties G_0 turns out not to matter, for others it matters very much. For example, in the case $m = 1$ the BA model describes the growth of a tree, *provided* G_0 *is a tree.* If G_0 is disconnected, say, then at all later stages the graph produced will also be disconnected. For general m the initial graph G_0 also has significant lasting effects, for example on the expected maximum degree, which can change by a constant factor when G_0 is changed.

The second problem is with the preferential attachment rule itself, and arises only for $m \geq 2$; when we add a new vertex, say the $t+1^{\text{st}}$, we must join it to a random set N_{t+1} of m earlier vertices. In our notation, working always with graphs on $\{1, 2, \ldots\}$, the BA model says only that, for $1 \leq i \leq t$,

$$\Pr(i \in N_{t+1}) = md_t(i)/\sum_{j=1}^{t} d_t(j), \tag{1}$$

where $d_t(i)$ is the degree of vertex i in the growing graph at time t. (Actually, as can be seen from the quotation above, Barabási and Albert give this formula without the factor of m. If we assume their formula is intended to hold separately for each edge added, then (1) follows. However, their description does not allow us to add edges one by one independently, as it is specified that the edges go to different vertices.) To fully describe the model, we must specify the distribution of N_{t+1}, i.e., the probability that $N_{t+1} = S$ for each of the $\binom{t}{m}$ possible sets S of earlier vertices. This distribution is not uniquely specified by giving the marginal probabilities that $i \in N_{t+1}$ for each earlier vertex i. To see this note, for example, that the distribution of N_{t+1} has $\binom{t}{m} - 1$ degrees of freedom (the $\binom{t}{m}$ probabilities must add up to 1) while there are only t marginal probabilities specified by the BA description. Again one might hope that the exact choice does not make much difference, and again this turns out to be false. As shown by the following result, there is a range of models fitting the BA description with very different properties.

Theorem 5. *Let $f(n)$, $n \geq 2$, be any integer valued function with $f(2) = 0$ and $f(n) \leq f(n+1) \leq f(n) + 1$ for every $n \geq 2$, such that $f(n) \to \infty$ as $n \to \infty$. Then there is a random graph process $T^{(n)}$ satisfying* (1) *such that, with probability 1, $T^{(n)}$ has exactly $f(n)$ triangles for all sufficiently large n.*

In less formal language, Theorem 5 says, for example, that if you want $\log n$ triangles when the graph has n vertices, there is a precise model satisfying the BA description (except for the start, which cannot be satisfied) which achieves this. Similarly, if you want n^α triangles for any $0 < \alpha \leq 1$, or any other plausible function. Thus the clustering coefficient (see section 12) may also be tuned. The only tiny caveat is that you may be forced to create a finite number of triangles at the start. Note that this is different from the result in [36], which considers a model outside the Barabási-Albert definition (triangles are created by adding edges between existing vertices).

Proof. We give only an outline of the proof. We will work entirely with simple graphs, with no loops or multiple edges, starting with $T^{(2)}$ a single edge. When adding a new vertex v to a simple graph and joining it to two distinct existing vertices, x and y, the number of triangles either remains the same, or goes up by one. It goes up by one if and only if xy is an edge. Restating the theorem, we must show that given $T^{(n)}$ we have two ways choosing x and y to define $T^{(n+1)}$, each satisfying the Barabási-Albert preferential attachment rule (1): one where xy is always an edge of $T^{(n)}$, and one where, except perhaps for finitely many steps near the start, it never is.

The first case is easy: to guarantee a new triangle, take xy to be a random edge of $T^{(n)}$. By definition of degree, the probability that a particular vertex w is chosen as one of x and y is just the degree $d(w)$ of w in $T^{(n)}$ over the total number $(2n-3)$ of edges of $T^{(n)}$, so (1) is satisfied.

For the second case we must assign non-negative weights $p_{xy} = p_{yx}$ to pairs $\{x, y\} \subset V(T^{(n)})$ with p_{xy} zero for every edge, such that $\sum_{y \neq x} p_{xy} = d(x)/(2n-3)$. Then $\sum_{\{x,y\}} p_{xy} = 1$, so we may take p_{xy} as the probability of joining the new vertex to x and y. Such an assignment is possible under very mild conditions; for example, the maximum degree of $T^{(n)}$ being at most $n/3$ is more than sufficient. It is easy to check that in any process satisfying (1), the maximum degree is at most $O(n^{1/2})$ **whp**, so the result follows. □

An extreme case of the process above, in which a triangle is added at every step, was actually considered by Dorogovtsev and Mendes [27] (section IX C), without noting that it satisfies the Barabási-Albert description. In fact, it is introduced there as a simpler alternative model for easier analysis.

As seen from the example above, in order to prove results about the BA model, one must first decide on the details of the model itself. In the next section we present one particular choice for how to do this which has several advantages.

7 The LCD model and $G_m^{(n)}$

In this section we define precisely a random graph model introduced in [16] satisfying the vague description given by Barabási and Albert. It turns out to be convenient to allow multiple edges and loops; there will not be very many of these, and in any case there seems no reason to exclude them from the point of view of the interpretation: one web site may contain several links to another, for example, or links to itself.

Consider a fixed sequence of vertices $v_1, v_2, \ldots$. (Later we shall take $v_i = i$; the general case simplifies the notation when we merge vertices.) Let us write $d_G(v)$ for the degree of the vertex v in the graph G. We define inductively a random graph process $(G_1^{(t)})_{t\geq 0}$ so that $G_1^{(t)}$ is a graph on $\{v_i : 1 \leq i \leq t\}$, as follows: start with $G_1^{(0)}$ the empty 'graph' with no vertices, or with $G_1^{(1)}$ the graph with one vertex and one loop. Given $G_1^{(t-1)}$, form $G_1^{(t)}$ by adding the vertex v_t together with a single edge between v_t and v_i, where i is chosen randomly with

$$\Pr(i = s) = \begin{cases} d_{G_1^{(t-1)}}(v_s)/(2t-1) & 1 \leq s \leq t-1, \\ 1/(2t-1) & s = t. \end{cases} \tag{2}$$

In other words, send an edge e from v_t to a random vertex v_i, where the probability that a vertex is chosen as v_i is proportional to its degree at the time, counting e as already contributing one to the degree of v_t. (We shall see why this is convenient later.) For $m > 1$, add m edges from v_t one at a time, counting the previous edges as well as the 'outward half' of the edge being added as already contributing to the degrees. We choose this precise rule because it leads to the following equivalent definition: define the process $(G_m^{(t)})_{t\geq 0}$ by running the process $(G_1^{(t)})$ on a sequence $v'_1, v'_2, \ldots$, and forming the graph $G_m^{(t)}$ from $G_1^{(mt)}$ by identifying the vertices $v'_1, v'_2, \ldots, v'_m$ to form v_1, identifying $v'_{m+1}, v'_{m+2}, \ldots, v'_{2m}$ to form v_2, and so on.

For the rest of the article we shall take $v_i = i$, so $G_m^{(t)}$ is a graph on $[t] = \{1, 2, \ldots, t\}$. Note that the edges of $G_m^{(t)}$ have a natural orientation, from later vertices to earlier vertices, so ij is oriented from i to j if $i > j$. However, as for studies of the BA model, we shall generally treat the graph as unoriented. For these models the orientation is not very interesting (indeed it may be reconstructed from the graph even if the vertex labels are not given).

In addition to satisfying the basic mathematical criterion of being precisely specified, the process $G_m^{(t)}$ has several useful properties. One is that $G_m^{(t)}$ can be defined in terms of $G_1^{(mt)}$, a much simpler object, so questions about $G_m^{(t)}$ can be re-written in terms of $G_1^{(mt)}$, and results can be proved this way. Another very important property is the following: while the process $G_1^{(t)}$ is dynamic, the distribution of the graph $G_1^{(n)}$ obtained at a particular time $t = n$ has a simple *static* description, the *linearized chord diagram* or LCD description, given in [16]:

> An *n-pairing* is a partition of the set $\{1, 2, \ldots, 2n\}$ into pairs, so there are $(2n)!/(n!2^n)$ n-pairings. These objects are sometimes thought of as *linearized chord diagrams* (or *LCDs*) [52, 15], where an LCD with n chords consists of $2n$ distinct points on the x-axis paired off by semi-circular chords in the upper half plane. Two LCDs are considered to be the same when one can be turned into the other by moving the points on the x-axis without changing their order. Thinking of pairings as LCDs, we shall talk of chords and their left and right endpoints. We form a graph $\phi(L)$ from an LCD L as follows: starting from the left, identify all endpoints up to and including the first right endpoint reached to form vertex 1. Then identify all further endpoints up to the next right endpoint to form vertex 2, and so on. For the edges, replace each chord by an edge joining the vertex corresponding to its right endpoint to that corresponding to its left endpoint.

As stated in [16], if L is chosen uniformly at random from all $(2n)!/(n!2^n)$ LCDs with n chords (i.e., n-pairings), then $\phi(L)$ has the same distribution as a random $G_1^{(n)}$ defined via the process $G_1^{(t)}$ given earlier.

> To see this note that L can be obtained by taking a random LCD L' with $n-1$ chords and adding a new chord whose right endpoint is to the right of all $n-1$ chords, and whose left endpoint lies in one of the $2n-1$ possible places, each chosen with equal probability. This corresponds to adding a new vertex to $\phi(L')$ and joining it to another vertex with probabilities according to the degrees, exactly as in the description of $(G_1^{(n)})$.

A simple result proved in [19] using the LCD description, which can also be proved in other ways, concerns the degree sequence. We write $\#_m^n(d)$ for the number of vertices of $G_m^{(n)}$ with *in-degree* equal to d, i.e., with (total) degree $m + d$.

Theorem 6. *Let $m \geq 1$ and $\epsilon > 0$ be fixed, and set*

$$\alpha_{m,d} = \frac{2m(m+1)}{(d+m)(d+m+1)(d+m+2)}.$$

Then with probability tending to 1 as $n \to \infty$ we have

$$(1-\epsilon)\alpha_{m,d} \leq \frac{\#_m^n(d)}{n} \leq (1+\epsilon)\alpha_{m,d}$$

for every d in the range $0 \leq d \leq n^{1/15}$.

This result gives a rigorous justification of the power-law dependence of the degrees described in [6].

Let us remark that in the case $m = 1$, essentially this result had been proved much earlier by Szymański [54] in a slightly different context (see section 15).

In the next few sections we describe other scale-free models for which power-law degree distribution has been proved.

8 The Buckley-Osthus model

Two groups, Dorogovtsev, Mendes and Samukhin [26] and Drinea, Enachescu and Mitzenmacher [25], introduced a variation on the BA model in which vertices have an 'initial attractiveness': the probability that an old vertex is chosen to be a neighbour of the new vertex is proportional to its in-degree plus a constant 'initial attractiveness', which we shall write as am. The case $a = 1$ is just the BA model, since there total degree is used, and each out-degree is m. Buckley and Osthus [20] made this more general model precise along the lines of the LCD model; for a fixed positive integer a, they define a process $H_{a,1}^{(t)}$ exactly as $G_1^{(t)}$ is defined above, but replacing (2) with

$$\Pr(i = s) = \begin{cases} \frac{d^{in}_{H_{a,1}^{(t-1)}}(v_s)+a}{(a+1)t-1} & 1 \le s \le t-1, \\ \frac{a}{(a+1)t-1} & s = t. \end{cases}$$

Note that when $a = 1$ the definition of $H_{a,1}^{(t)}$ reduces exactly to that of $G_1^{(t)}$. As for $G_m^{(t)}$, a process $H_{a,m}^{(t)}$ is defined in [20] by identifying vertices in groups of m. Buckley and Osthus established that the degree distribution of this model also obeys a power law. Let us write $\#^n_{a,m}(d)$ for the number of vertices of $H_{a,m}^{(n)}$ with in-degree d.

Theorem 7. *Let $m \ge 1$ and $a \ge 1$ be fixed integers, and set*

$$\alpha_{a,m,d} = (a+1)(am+a)!\binom{d+am-1}{am-1}\frac{d!}{(d+am+a+1)!}.$$

Let $\epsilon > 0$ be fixed. Then **whp** *we have*

$$(1-\epsilon)\alpha_{a,m,d} \le \frac{\#^n_{a,m}(d)}{n} \le (1+\epsilon)\alpha_{a,m,d}$$

for all d in the range $0 \le d \le n^{1/100(a+1)}$. In particular, **whp** *for all d in this range we have*

$$\frac{\#^n_{a,m}(d)}{n} = \Theta(d^{-2-a}).$$

The proof is rather difficult, as the equivalent for $H_{a,1}^{(t)}$ of the LCD model for $G_1^{(t)}$ is much more complicated. Dorogovtsev, Mendes and Samukhin [26] gave a non-rigorous argument for a weaker form of this result, where the range of d considered is bounded.

9 The copying model

Around the same time as the BA model, Kumar, Raghavan, Rajagopalan, Sivakumar, Tomkins and Upfal [40] gave rather different models to explain the observed power laws in the web graph. The basic idea is that a new web page is often made by copying an old one, and then changing some of the links. Let us define one of these models by quoting almost verbatim from [40]:

> The linear growth copying model is parametrized by a *copy factor* $\alpha \in (0,1)$ and a constant out-degree $d \geq 1$. At each time step, one vertex u is added, and u is then given d out-links for some constant d. To generate the out-links, we begin by choosing a 'prototype' vertex p uniformly at random from V_t (the old vertices). The i^{th} out-link of u is then chosen as follows. With probability α, the destination is chosen uniformly at random from V_t, and with the remaining probability the out-link is taken to be the i^{th} out-link of p. Thus, the prototype is chosen once in advance. The d out-links are chosen by α-biased independent coin flips, either randomly from V_t, or by copying the corresponding out-link of the prototype.
>
> The intuition behind this model is the following. When an author decides to create a new web page, the author is likely to have some topic in mind. The choice of prototype represents the choice of topic—larger topics are more likely to be chosen. The Bernoulli copying events reflect the following intuition: a new viewpoint about the topic will probably link to many pages 'within' the topic (i.e., pages already linked to by existing resource lists about the topic), but will also probably introduce a new spin on the topic, linking to some new pages whose connection to the topic was previously unrecognized.

As for the BA model, it turns out that the degree distribution does follow a power law. Let $N_{t,r}$ be the expected number of vertices of degree r in the graph formed by the model outlined above at time t (when the graph has t vertices). Among other results, the following was proved in [40].

Theorem 8. *For $r > 0$, the limit $P_r = \lim_{t\to\infty} N_{t,r}/t$ exists, and satisfies*

$$P_r = P_0 \prod_{i=1}^{r} \frac{1 + \alpha/(i(1-\alpha))}{1 + 2/(i(1-\alpha))}$$

and

$$P_r = \Theta\left(r^{-\frac{2-\alpha}{1-\alpha}}\right).$$

When one looks only at the degree sequence, this copying model behaves very similarly to models with preferential attachment; we shall return to this in the next section. In other ways, however, the model is essentially different. An obvious example is that copying will give rise to many more dense bipartite subgraphs; part of the original motivation was to explain the appearance of these in the web graph.

10 The Cooper-Frieze model

Recently, Cooper and Frieze [24] introduced a model with many parameters which includes the models of the last three sections as special cases, and proved a very general result about the power-law distribution of degrees. In the undirected case, the model describes a (multi-)graph process $G(t)$, starting from $G(0)$ a single vertex with no edges. Their attachment rule is a mixture of preferential (by degree) and uniform (uniformly at random, or 'u.a.r'). Quoting from [24]:

> Initially, at step $t = 0$, there is a single vertex v_0. At any step $t = 1, 2, ..., T, ...$ there is a birth process in which either new vertices or new edges are added. Specifically, either a procedure NEW is followed with probability $1 - \alpha$, or a procedure OLD is followed with probability α. In procedure NEW, a new vertex v is added to $G(t-1)$ with one or more edges added between v and $G(t-1)$. In procedure OLD, an existing vertex v is selected and extra edges are added at v.
>
> The recipe for adding edges typically permits the choice of initial vertex v (in the case of OLD) and the terminal vertices (in both cases) to be made either u.a.r or according to vertex degree, or a mixture of these two based on further sampling. The number of edges added to vertex v at step t by the procedures (NEW, OLD) is given by distributions specific to the procedure. The details of these choices are given below.

The parameters fixed in advance are integers $j_0, j_1 \geq 1$, and real numbers α, β, γ, δ, $p_1, \ldots, p_{j_0}$, $q_1, \ldots q_{j_1}$ between 0 and 1, with $\alpha < 1$ and $\sum p_i = \sum q_i = 1$. The procedure for defining $G(t)$ from $G(t-1)$ is as follows (from [24]):

> Choice of procedure at step t.
>
> α: Probability that an OLD node generates edges.
>
> $1 - \alpha$: Probability that a NEW node is created.
>
> Procedure NEW
>
> $\boldsymbol{p} = (p_i : i \geq 1)$: Probability that new node generates i new edges.
>
> β: Probability that choices of terminal vertices are made uniformly.
>
> $1 - \beta$: Probability that choices of terminal vertices are made according to degree.
>
> Procedure OLD
>
> $\boldsymbol{q} = (q_i : i \geq 1)$: Probability that old node generates i new edges.
>
> δ: Probability that the initial node is selected uniformly.
>
> $1 - \delta$: Probability that the initial node is selected according to degree.
>
> γ: Probability that choices of terminal vertices are made uniformly.

$1-\gamma$: Probability that choices of terminal vertices are made according to degree.

In words:

> The model creates edges in the following way: An initial vertex v is selected. If the terminal vertex w is chosen u.a.r, we say v is *assigned* to w. If the terminal vertex w is chosen according to its vertex degree, we say v is *copied* to w. In either case the edge has an intrinsic direction (v, w), which we may choose to ignore. We note that sampling according to vertex degree is equivalent to selecting an edge u.a.r and then selecting an endpoint u.a.r.

Note that although this 'copying' mechanism is not at all that of the 'copying model' described in the previous section, as pointed out by Cooper and Frieze, as far as the evolution of the degree sequence is concerned, the two are more or less interchangeable. Note also that the mixture of uniform and preferential attachment is easily seen to be equivalent to the preferential attachment with constant initial attractiveness considered in section 8.

Given the generality of the Cooper-Frieze model, it is not surprising that their result is rather difficult to state. Quoting again from [24], we must first start with some notation.

> **Notation**
> Let $\mu_p = \sum_{j=0}^{j_0} jp_j$, $\mu_q = \sum_{j=0}^{j_1} iq_j$ and let $\theta = 2((1-\alpha)\mu_p + \alpha\mu_q)$. To simplify subsequent notation, we transform the parameters as follows:
>
> $$\begin{aligned} a &= 1 + \beta\mu_p + \frac{\alpha\gamma\mu_q}{1-\alpha} + \frac{\alpha\delta}{1-\alpha}, \\ b &= \frac{(1-\alpha)(1-\beta)\mu_p}{\theta} + \frac{\alpha(1-\gamma)\mu_q}{\theta} + \frac{\alpha(1-\delta)}{\theta}, \\ c &= \beta\mu_p + \frac{\alpha\gamma\mu_q}{1-\alpha}, \\ d &= \frac{(1-\alpha)(1-\beta)\mu_p}{\theta} + \frac{\alpha(1-\gamma)\mu_q}{\theta}, \\ e &= \frac{\alpha\delta}{1-\alpha}, \qquad f = \frac{\alpha(1-\delta)}{\theta}. \end{aligned}$$
>
> We note that
> $$c + e = a - 1 \text{ and } b = d + f. \tag{3}$$
> Now define the sequence $(d_0, d_1, ..., d_k, ...)$ by $d_0 = 0$ and for $k \geq 1$
> $$d_k(a+bk) = (1-\alpha)p_k + (c+d(k-1))d_{k-1} + \sum_{j=1}^{k-1}(e+f(k-j))q_j d_{k-j}. \tag{4}$$
> Since $a \geq 1$, this system of equations has a unique solution.

Statement of results

The main quantity we study is the random variable $D_k(t)$, the number of vertices of degree k at step t. We let $\overline{D}_k(t) = \mathbb{E}(D_k(t))$. We prove that for small k, $\overline{D}_k(t) \approx d_k t$ as $t \to \infty$.

Theorem 9. *There exists a constant $M > 0$ such that for $t, k = 1, 2, \ldots,$*

$$|\overline{D}_k(t) - td_k| \leq Mt^{1/2} \log t.$$

The number of vertices $\nu(t)$ at step t is **whp** asymptotic to $(1-\alpha)t$, see ... It follows that

$$\bar{d}_k = \frac{d_k}{1-\alpha}.$$

The next theorem summarizes what we know about the d_k:

Theorem 10.

(i) $Ak^{-\zeta} \leq d_k \leq B \min\{k^{-1}, k^{-\zeta/j_1}\}$ *where* $\zeta = (1 + d + f\mu_q)/(d + f)$.

(ii) *If* $j_1 = 1$ *then* $d_k \sim Ck^{-(1+1/(d+f))}$.

(iii) *If* $f = 0$ *then* $d_k \sim Ck^{-(1+1/d)}$.

(iv) *If the* SOLUTION CONDITIONS *hold then*

$$d_k = C\left(1 + O\left(\frac{1}{k}\right)\right)k^{-x},$$

where C is constant and

$$x = 1 + \frac{1}{d + f\mu_q}. \tag{5}$$

We say that $\{q_j : j = 1, ..., j_1\}$ is *periodic* if there exists $m > 1$ such that $q_j = 0$ unless $j \in \{m, 2m, 3m, \ldots\}$.

Let

$$\phi_1(y) = y^{j_1} - \left(\frac{d + q_1 f}{b} y^{j_1 - 1} + \frac{q_2 f}{b} y^{j_1 - 2} + \cdots + \frac{q_{j_1} f}{b}\right).$$

Our SOLUTION CONDITIONS are:

S(i) $f > 0$ and either (a) $d + q_1 f > 0$ or (b) $\{q_j\}$ is not periodic.

S(ii) The polynomial $\phi_1(y)$ has no repeated roots.

In summary, these results say that the 'expected degree sequence' converges in a strong sense to the solution of a certain recurrence relation, and that under rather weak conditions, this solution follows a power law with an explicitly determined exponent and a bound on the error term. Cooper and Frieze also prove a simple concentration result, which we will not state, showing the number of vertices of a certain degree is close to its expectation.

In addition to the undirected case, Cooper and Frieze consider what one might call 'semi-directed' models, where one uses in- or out-degree for the preferential attachment rule, but not both. In the next section we describe a simple model which uses both.

11 Directed scale-free graphs

Although sometimes described in terms of directed graphs, with the exception noted above all the models presented so far are to all intents and purposes undirected, in the sense that the edge orientations are not used in an essential way in defining the model. As the real-world networks motivating scale-free random graphs are often directed, it makes sense to consider directed models, and it is natural to consider 'preferential attachment' which depends on in- and out-degrees. Such a model was introduced by Bollobás, Borgs, Chayes and Riordan in [12]:

> We consider a graph which grows by adding single edges at discrete time steps. At each such step a vertex may or may not also be added. For simplicity we allow multiple edges and loops. More precisely, let α, β, γ, δ_{in} and δ_{out} be non-negative real numbers, with $\alpha + \beta + \gamma = 1$. Let G_0 be any fixed initial graph, for example a single vertex without edges, and let t_0 be the number of edges of G_0. (Depending on the parameters, we may have to assume $t_0 \geq 1$ for the first few steps of our process to make sense.) We set $G(t_0) = G_0$, so at time t the graph $G(t)$ has exactly t edges, and a random number $n(t)$ of vertices. In what follows, *to choose* a vertex v of $G(t)$ *according to* $d_{out} + \delta_{out}$ means to choose v so that $\Pr(v = v_i)$ is proportional to $d_{out}(v_i)+\delta_{out}$, i.e., so that $\Pr(v = v_i) = (d_{out}(v_i) + \delta_{out})/(t + \delta_{out}n(t))$. To *choose* v *according to* $d_{in} + \delta_{in}$ means to choose v so that $\Pr(v = v_i) = (d_{in}(v_i) + \delta_{in})/(t + \delta_{in}n(t))$, where all degrees are measured in $G(t)$.
>
> For $t \geq t_0$ we form $G(t+1)$ from $G(t)$ according the the following rules:
>
> (A) With probability α, add a new vertex v together with an edge from v to an existing vertex w, where w is chosen according to $d_{in} + \delta_{in}$.
>
> (B) With probability β, add an edge from an existing vertex v to an existing vertex w, where v and w are chosen independently, v according to $d_{out} + \delta_{out}$, and w according to $d_{in} + \delta_{in}$.
>
> (C) With probability γ, add a new vertex w and an edge from an existing vertex v to w, where v is chosen according to $d_{out} + \delta_{out}$.
>
> The probabilities α, β and γ clearly should add up to one. To avoid trivialities, we will also assume that $\alpha + \gamma > 0$. When considering the web graph we take $\delta_{out} = 0$; the motivation is that vertices

added under step (C) correspond to web pages which purely provide content - such pages never change, are born without out-links and remain without out-links. Vertices added under step (A) correspond to usual pages, to which links may be later added. While mathematically it seems natural to take $\delta_{in} = 0$ in addition to $\delta_{out} = 0$, this gives a model in which every page not in G_0 has either no in-links or no out-links, which is rather unrealistic and uninteresting! A non-zero value of δ_{in} corresponds to insisting that a page is not considered part of the web until something points to it, typically one of the big search engines. It is natural to consider these edges from search engines separately from the rest of the graph, as they are of a rather different nature; for the same reason it is natural not to insist that δ_{in} be an integer. We include the parameter δ_{out} to make the model symmetric with respect to reversing the directions of edges (swapping α with γ and δ_{in} with δ_{out}), and because we expect the model to be applicable in contexts other than that of the web graph.

Our model allows loops and multiple edges; there seems no reason to exclude them. However, there will not be very many, so excluding them would not significantly affect our conclusions.

Note also that our model includes (a precise version of) the $m = 1$ case of the original model of Barabási and Albert as a special case, taking $\beta = \gamma = \delta_{out} = 0$ and $\alpha = \delta_{in} = 1$. We could introduce more parameters, adding m edges for each new vertex, or (as in [24]) a random number with a certain distribution, but one of our aims is to keep the model simple, and the main effect, of varying the overall average degree, can be achieved by varying β.

As for the other models, power law degree distribution is proved, this time for in- and out-degrees separately. Setting

$$c_1 = \frac{\alpha + \beta}{1 + \delta_{in}(\alpha + \gamma)} \quad \text{and} \quad c_2 = \frac{\beta + \gamma}{1 + \delta_{out}(\alpha + \gamma)},$$

and writing $x_i(t)$ for the number of vertices of $G(t)$ with in-degree i, and $y_i(t)$ for the number with out-degree i, the following result is proved in [12].

Theorem 11. *Let $i \geq 0$ be fixed. There are constants p_i and q_i such that $x_i(t) = p_i t + o(t)$ and $y_i(t) = q_i t + o(t)$ hold with probability 1. Furthermore, if $\alpha\delta_{in} + \gamma > 0$ and $\gamma < 1$, then as $i \to \infty$ we have*

$$p_i \sim C_{IN} i^{-X_{IN}},$$

where $X_{IN} = 1 + 1/c_1$ and C_{IN} is a positive constant. If $\gamma\delta_{out} + \alpha > 0$ and $\alpha < 1$, then as $i \to \infty$ we have

$$q_i \sim C_{OUT} i^{-X_{OUT}},$$

with $X_{OUT} = 1 + 1/c_2$ and C_{OUT} is a positive constant.

In the statement above, the $o(t)$ notation refers to $t \to \infty$ with i fixed, while $a(i) \sim b(i)$ means $a(i)/b(i) \to 1$ as $i \to \infty$.

In addition, the joint distribution of in- and out-degrees is studied; formulae are given for the limiting fraction $r_{i,j}$ of vertices with in-degree i and out-degree j. As these are not very simple, we shall not reproduce them here.

For the rest of the article we return to the LCD model, turning our attention to properties other than the degree sequence.

12 Clustering coefficient and small subgraphs

Following Watts and Strogatz, one of the basic properties of the scale-free random graphs considered in many papers is the *clustering coefficient* C. As we have seen in section 4, this coefficient describes 'what proportion of the acquaintances of a vertex know each other'. Formally, given a simple graph G (without loops and multiple edges), and a vertex v (with at least two neighbours, say), the *local clustering coefficient* at v is given by

$$C_v(G) = \frac{\text{number of edges between neighbours of } v}{\binom{d_G(v)}{2}}, \tag{6}$$

where $d_G(v)$ is the degree of v in G, so the denominator is the maximum possible number of edges between neighbours of v, and $0 \le C_v(G) \le 1$. There are then two possible definitions for the clustering coefficient $C = C(G)$ of the whole graph. Perhaps the most often stated, is '$C(G)$ is the average of $C_v(G)$', i.e., taking the vertex set to be $[n]$,

$$C(G) = C_1(G) = \sum_{v=1}^{n} C_v(G)/n. \tag{7}$$

(Again we commit a slight abuse of notation, as 1 is also a vertex of the graph.) This kind of 'average of an average' is often not very informative; the more natural alternative is to weight by the denominator of (6), giving

$$C(G) = C_2(G) = \left(\sum_{v=1}^{n} \binom{d_G(v)}{2} C_v(G) \right) \Big/ \sum_{v=1}^{n} \binom{d_G(v)}{2}. \tag{8}$$

This second definition is easily seen to have the following natural equivalent form:

$$C_2(G) = \frac{\text{no. of pairs } ab,\ ac \text{ of adjacent edges for which } bc \text{ is an edge}}{\text{no. of pairs } ab,\ ac \text{ of adjacent edges}},$$

which has the advantage that it makes sense when some degrees are less than 2. In turn we can re-write the equation above as

$$C_2(G) = \frac{3 \times \text{number of triangles}}{\text{number of pairs of adjacent edges}}. \tag{9}$$

In this form there is no problem applying the definition to multigraphs.

In some papers it is not clear which of the two definitions above is intended; when it is clear, sometimes C_1 is used, and sometimes C_2. It is not often pointed out that these definitions are different: for an extreme example, take G to be a double star, where vertices 1 and 2 are joined to each other and to all other vertices, and there are no other edges. Then $C_v(G)$ is 1 for $v \geq 3$ and $2/(n-1)$ for $v = 1, 2$. It follows that $C_1(G) = 1 - o(1)$, while $C_2(G) \sim 2/n$. In more balanced graphs the definitions will give more similar values, but they will still differ by at least a constant factor much of the time.

For this section we shall use $C_2(G)$ as the definition of the clustering coefficient $C(G)$, and we shall prove the following result.

Theorem 12. *Let $m \geq 1$ be fixed. The expected value of the clustering coefficient $C(G_m^{(n)}) = C_2(G_m^{(n)})$ satisfies*

$$\mathbb{E}(C(G_m^{(n)})) \sim \frac{m-1}{8} \frac{(\log n)^2}{n}$$

as $n \to \infty$.

To prove Theorem 12 we will count the number of triangles in $G_m^{(n)}$. More generally, we describe a method for counting subgraphs isomorphic to any small fixed graph.

When $m = 1$, the Barabási-Albert or LCD model is very simple, giving either a tree or a forest with loops according to the precise definition chosen. Although this model is less interesting than the general case, it has the advantage that its small subgraphs can be analyzed precisely (see section 15). Because of the exact choice made in the definition of $G_m^{(n)}$, these results then carry over to this full model.

Let S be a graph on $\{1, 2, \ldots, n\}$ with loops allowed. Orient each edge ij of S with $i \leq j$ from j to i. Let us write $V^+(S)$ for the set of vertices of S from which edges leave, and $V^-(S)$ for those vertices at which edges arrive. (These sets are, of course, not in general disjoint.) For $i \in V(S)$ let $d_S^{in}(i)$ be the in-degree of i in S and let $d_S^{out}(i)$ be the out-degree. (A loop at i contributes 1 to each of $d_S^{in}(i)$ and $d_S^{out}(i)$.) Finally, let $C_S(t)$ be the number of edges of S 'crossing' t, i.e., the number of edges ij of S with $i \leq t$ and $j \geq t$.

Note that S is a *fixed* graph, not an isomorphism class of graphs; there are $\binom{n}{3}$ different graphs S which are triangles, for example. When we say S is a subgraph of $G_1^{(n)}$, or write $S \subset G_1^{(n)}$, we shall mean that exactly the edges of S occur in $G_1^{(n)}$, not that $G_1^{(n)}$ has a subgraph isomorphic to S. Suppose that $d_S^{out}(i) \leq 1$ for every i, so S is a possible subgraph of $G_1^{(n)}$.

Theorem 13. *Let S be a possible subgraph of $G_1^{(n)}$. With the notation above, the probability p_S that $S \subset G_1^{(n)}$ satisfies*

$$p_S = \prod_{i \in V^-(S)} d_S^{in}(i)! \prod_{i \in V^+(S)} \frac{1}{2i-1} \prod_{t \notin V^+(S)} \left(1 + \frac{C_S(t)}{2t-1}\right). \tag{10}$$

Furthermore,

$$p_S = \prod_{i\in V^-(S)} d_S^{in}(i)! \prod_{ij\in E(S)} \frac{1}{2\sqrt{ij}} \exp\left(O\left(\sum_{i\in V(S)} C_S(i)^2/i\right)\right). \tag{11}$$

The result is simple to prove once one finds the correct quantity to calculate inductively; the details are given for a closely related model in section 15. Note that the second product in (11) gives essentially what one would expect if edges were present in $G_1^{(n)}$ independently of one another. The first product (and the final factor) show that they are not.

We now pass to $G_m^{(n)}$. Rather than writing down a cumbersome general formula, let us consider the case of triangles.

Theorem 14. *Let $m \geq 1$ be fixed. The expected number of triangles in $G_m^{(n)}$ is given by*

$$(1+o(1))\frac{m(m-1)(m+1)}{48}(\log n)^3$$

as $n \to \infty$.

Proof. Recall that that $G_m^{(n)}$ is obtained from $G_1^{(mn)}$ by identifying the vertices in groups of m. Let a, b, c with $1 \leq a < b < c \leq n$ be given. Then abc is a triangle in $G_m^{(n)}$ if and only if there are integers $m(a-1) < i, i' \leq ma$, $m(b-1) < j, j' \leq mb$, $m(c-1) < k, k' \leq mc$ such that the graph S with edges ij', jk' and $i'k$ is a subgraph of $G_1^{(mn)}$. Now for this S, provided $d_S^{out}(v) \leq 1$ for all v, we have from (11) that

$$p_S = \eta_1 \prod_{x\in V^-(S)} d_S^{in}(x)! \prod_{xy\in E(S)} \frac{1}{2\sqrt{xy}} = \eta_2 \prod_{x\in V^-(S)} d_S^{in}(x)! \frac{1}{8m^3abc},$$

where the 'correction factors' η_1, η_2 are bounded, and tend to 1 if $a \to \infty$. Given $1 \leq a < b < c \leq n$, what are the possible choices for i, i', j, j', k, k'? Note first that k, k' must be distinct, giving $m(m-1)$ choices, as if $k = k'$ then $d_S^{out}(k) = 2$. There are m^2 choices for j, j'. Finally we may have $i = i'$, in which case $d_S^{in}(i) = 2$ and $d_S^{in}(v) \leq 1$ for every other v, or $i \neq i'$ in which case $d_S^{in}(v) \leq 1$ for all v. There are $m(m-1)m^2m = m^4(m-1)$ choices with $i = i'$ and $m(m-1)m^2m(m-1) = m^4(m-1)^2$ choices with $i \neq i'$. Hence the expected number of triangles with vertices a, b, c in $G_m^{(n)}$ (recall that $G_m^{(n)}$ is a multigraph, so may contain several triangles with the same vertex set) is given by

$$\eta_3\left(m^4(m-1)2\frac{1}{8m^3abc} + m^4(m-1)^2 1\frac{1}{8m^3abc}\right) = \eta_3\frac{m(m-1)(m+1)}{8abc},$$

where η_3 is bounded and tends to 1 as $a \to \infty$. Summing over a, b and c with $1 \leq a < b < c \leq n$ we see that the main contribution is from terms with $a \to \infty$, and the expected number of triangles in $G_m^{(n)}$ is given by

$$(1+o(1)) \sum_{1\le a<b<c\le n} \frac{m(m-1)(m+1)}{8abc} \sim \frac{m(m-1)(m+1)}{48}(\log n)^3,$$

proving Theorem 14. □

One can use the same method to prove the following more general result.

Theorem 15. *Let $l \ge 3$ be fixed. Then the expected number of l-cycles in $G_m^{(n)}$ is of the form*

$$(1+o(1))C_{m,l}(\log n)^l$$

as $n \to \infty$ with $m \ge 2$ fixed, where $C_{m,l}$ is a positive constant. Furthermore, as $m \to \infty$ we have $C_{m,l} = \Theta(m^l)$.

Finding the exact constants in the result above becomes harder as l increases; for example, given $1 \le a < b < c < d < e \le n$ there are 12 ways to arrange a 5-cycle with these vertices. In 8 of these arrangements there are two vertices with two edges coming in from the right. In the other four there is only one such vertex. When passing to $G_1^{(mn)}$ there may thus be 0,1 or 2 vertices with in-degree 2.

Note that Theorems 14 and 15 contradict the heuristic value of $n^{l/4}$ for the number of l-cycles given by Farkas, Derényi, Barabási and Vicsek [31] on the basis of eigenvalue distribution.

Let us finish this section by returning to the clustering coefficient, calculated according to (9). Having estimated the number of triangles, we only need to know the number of pairs of adjacent edges. Let us write $P_2 = P_2(G_m^{(n)})$ for the number of pairs of adjacent edges ab, ac in $G_m^{(n)}$.

Theorem 16. *Let $m \ge 1$ and $\epsilon > 0$ be fixed. Then*

$$(1-\epsilon)\frac{m(m+1)}{2} n \log n \le P_2(G_m^{(n)}) \le (1+\epsilon)\frac{m(m+1)}{2} n \log n$$

holds **whp** *as $n \to \infty$.*

Proof. The result is easy to prove using the methods above, so we give only a brief sketch.

There are three types of contribution to P_2: we may have $b, c \le a$, $b \le a < c$ (equivalent to $c \le a < b$) or $a < b, c$. Since all out-degrees in $G_m^{(n)}$ are at most m, there can only be $O(n)$ pairs of the first two types. Hence, using the methods above (skipping the details),

$$\mathbb{E}(P_2(G_m^{(n)})) = O(n) + (1+o(1)) \sum_{1\le a<b<c} \frac{2m^3 + m^3(m-1)}{4m^2 a\sqrt{bc}}$$

$$\sim \frac{m(m+1)}{2} n \log n.$$

Standard martingale methods can be used to show that P_2 is concentrated within $O(n)$ of its mean, completing the proof. □

Note that both the number of triangles and the number of pairs of adjacent edges are not quite what one might expect just by looking at the individual edge probabilities (or the degrees). One difference is an extra factor of $(m+1)/m$ appearing in both, from the correlation between the presence of edges ab and ac when $a < b < c$. The other is a factor $(m-1)/m$ only in the number of triangles, from the fixed out-degrees. These factors are often ignored, for example in [39].

Combined with Theorem 14 and the definition (9), the result above shows that the expected clustering coefficient C of $G_m^{(n)}$ is asymptotically

$$\frac{m-1}{8}\frac{(\log n)^2}{n},$$

proving Theorem 12. Note that the clustering coefficient is very different from the experimental value $C \sim n^{-0.75}$ given for $m = 2$ by Barabási and Albert [2], or the heuristic $C \sim n^{-0.25}$ that would follow from the claims of Farkas, Derényi, Barabási and Vicsek [31]. Klemm and Eguiluz [39] give an 'analytic' value of

$$C_1(G) \sim \frac{m}{8}\frac{(\log n)^2}{n}$$

for $C_1(G)$; this is off by a constant factor for two reasons. One is that the heuristic used ignores the factor $(m-1)/m$ mentioned above. The other is that, while the aim is clearly to calculate $C_1(G)$, a heuristic used is to replace the top and bottom of (6) by their expectations. This introduces an error which, for $G_m^{(n)}$, one can check is a roughly constant factor; it turns out that this error is roughly the ratio between $C_2(G)$ and $C_1(G)$, explaining the similarity between the formula above and the true form of $C_2(G)$ given in Theorem 12.

13 Pairings on $[0, 1]$ and the diameter of the LCD model

So far the results concerning the LCD model have been obtained directly either from (2) or from the discrete combinatorial interpretation in terms of pairings. In [16] another way of generating $G_m^{(n)}$ was introduced; this formulation, in terms of pairings of random real numbers, is useful for proving more complicated results, as it allows the re-introduction of independence to a significant extent.

Let $N = mn$. The idea is that to obtain our LCD with N chords, instead of pairing off fixed points $1, 2, \ldots 2N$, we shall pair off random points in the interval $[0, 1]$. In fact, taking iid uniformly random points $x_1, \ldots, x_{2N}$, we may as well pair x_{2i-1} with x_{2i} for all i; the randomness of the order in which the x_i appear when moving from 0 to 1 guarantees that the LCD obtained is the uniformly random LCD we require.

We now consider generating the pairing starting with the right endpoints. As in [16], we call a random variable with density function $2x$, $0 < x < 1$, an $M_2(0, 1)$ *random variable*. Let us write l_i, r_i for the left and right endpoints of the chord $x_{2i-1}x_{2i}$, so $\{l_i, r_i\} = \{x_{2i-1}, x_{2i}\}$ with $l_i < r_i$. Then $\Pr(r_i \leq t) =$

$\Pr(x_{2i-1}, x_{2i} \le t) = t^2$, so the r_i are iid $M_2(0,1)$ random variables. Also, given $r_1, \ldots, r_N$, the random variables $l_1, \ldots, l_N$ are independent with l_i uniformly distributed on $[0, r_i]$.

To express the LCD we have defined as a pairing on $\{1, 2, \ldots, 2N\}$, we must sort all the x_i together. We do this by first sorting the r_i, and then considering between which r_i each l_i lies.

The construction for $G_1^{(mn)}$ is as follows: we start with $N = mn$ iid $M_2(0,1)$ random variables, $r_1, \ldots, r_N$. Sort these into increasing order, to obtain $R_1, \ldots, R_N$, setting $R_0 = 0$. Let $L_1', \ldots, L_N'$ be independent, with L_i' uniform on $[0, R_i]$. Then our LCD $\mathcal{L}$ is given by pairing L_i' and R_i. As the right endpoints $R_1, \ldots, R_N$ are already in order, if $R_{j-1} < L_i' < R_j$ then in the graph $G_1^{(mn)}$ obtained as $\phi(\mathcal{L})$ (see section 7), vertex i sends its out-going edge to vertex j. (Throughout we of course ignore the probability zero event that two endpoints are the same.)

For $G_m^{(n)}$ we must merge vertices in groups of m, so what will really matter is where the m^{th}, $2m^{\text{th}}$ etc. right endpoints lie. Simplifying very slightly, the construction is as follows: let the R_i be defined as above. For $1 \le i \le n$ set $W_i = R_{mi}$, taking $W_0 = 0$. To obtain exactly $G_m^{(n)}$ we should consider N independent random variables which we denote $L_{i,r}$, $1 \le i \le n$, $1 \le r \le m$, with $L_{i,r}$ uniform on $[0, R_{(m-1)i+r}]$. In fact it is often good enough to work only with the W_i, taking $L_{i,r}$ uniform on $[0, W_i] = [0, R_{mi}]$. The graph $G_m^{(n)}$ is obtained by taking edges from i to m (not necessarily distinct) vertices $t_{i,1}, \ldots, t_{i,m}$ obtained as follows: $t_{i,r}$ is the integer t for which $W_{t-1} < L_{i,r} < W_t$.

In summary, the following is an almost exact alternative description of $G_m^{(n)}$. (The modifications to make it exact are implicit in the paragraph above.)

Let random variables W_i be defined as above, and set $w_i = W_i - W_{i-1}$. Given the W_i, define independent random variables $t_{i,r}$, $1 \le i \le n$, $1 \le r \le m$, with

$$\Pr(t_{i,r} = j) = \begin{cases} w_j / W_i & j \le i, \\ 0 & j > i. \end{cases} \tag{12}$$

Then the graph formed by taking edges from i to $t_{i,r}$ has (essentially) the same distribution as $G_m^{(n)}$. The power of this approach is that we may condition on the W_i, assuming they have 'typical' properties. Then the $t_{i,r}$ are conditionally independent.

As a simple application of this approach, we observe that the maximum degree of $G_m^{(n)}$ has the following rather unpleasant description. Let $X_1, X_2, \ldots$ be the points of a Poisson process on $[0, \infty]$ with rate m, so, setting $X_0 = 0$, the variables $X_i - X_{i-1}$ are iid exponentials with mean $1/m$. Let $Y_i = \sqrt{X_{mi}}$, and let $D_m = \max\{Y_i - Y_{i-1}, 1 \le i < \infty\}$, noting that this maximum exists with probability one. Note that the distribution of D_m depends on m only. Let $\Delta(G_m^{(n)})$ denote the maximum degree of $G_m^{(n)}$.

Theorem 17. *Let $m \ge 1$ be fixed. Then $\Delta(G_m^{(n)})/(2m\sqrt{n})$ converges in distribution to D_m as $n \to \infty$.*

Proof. As before, here we can only give a sketch. Note that if U is a random variable which is uniform on $[0, 1]$, then $\sqrt{U}$ has a $M_2(0, 1)$ distribution, since $\Pr(\sqrt{U} \leq t) = \Pr(U \leq t^2) = t^2$. Loosely speaking, it follows that for large n, the distribution of the squares of the first few R_i is given by a Poisson process of rate mn on $[0, \epsilon]$, for ϵ sufficiently small, so we may take $R_i^2 = X_i/n$, or $R_i = \sqrt{X_i/n}$. Then the W_i are given by $Y_i/\sqrt{n}$, so $\max\{w_i\}$ is given by $D_m/\sqrt{n}$. Finally, *given the* W_i, the degree in $G_m^{(n)}$ of a particular early vertex j is concentrated about its expectation of $(2 + o(1))mnw_j$. □

A much more substantial result proved using this description of $G_m^{(n)}$ is the diameter formula in [16]. Before stating this result, let us pause for a moment to consider what we might expect the diameter to be. Computer experiments presented by Barabási, Albert and Jeong [3, 7] and heuristic arguments given by Newman, Strogatz and Watts [48] suggest that $G_m^{(n)}$ should have diameter of the form $A + B \log n$. At first sight, such a small diameter might seem surprising, but it is in line with the Watts-Strogatz small-world phenomenon described in section 4. What would we expect from the point of view of random graphs? Certainly *at most* $\Theta(\log n)$: as described in section 4, even a small amount of global randomness gives logarithmic diameter. In fact one might expect the diameter to be even smaller: the unbalanced degree distribution pushes up the number of small paths, and thus, perhaps, pushes the diameter down. As shown in [16], this is indeed the case, though it is not very easy to prove.

Theorem 18. *Fix an integer* $m \geq 2$ *and a positive real number* ϵ. *Then* **whp** $G_m^{(n)}$ *is connected and has diameter* $\operatorname{diam}(G_m^{(n)})$ *satisfying*

$$(1 - \epsilon) \log n / \log \log n \leq \operatorname{diam}(G_m^{(n)}) \leq (1 + \epsilon) \log n / \log \log n.$$

The lower bound is relatively straightforward, based on counting the expected number of paths between two fixed vertices using techniques similar to those in section 12. The upper bound, proved via a neighbourhood expansion argument, is much more complicated. Essential use is made of the independence introduced by conditioning on the W_i, but even with this there are many complications.

As pointed out in [16], and independently by Cohen and Havlin [23], there is a heuristic argument giving the correct diameter of $\log n / \log \log n$. (This is just the standard neighbourhood expansion argument without the details; it is important to take the whole degree sequence and not apply some form of cutoff.) However one must be careful with such heuristics. For example, they apply also to the case $m = 1$, where, as shown by Pittel [49] in the context of scale-free trees, the diameter is $\Theta(\log n)$.

14 Robustness and vulnerability

Another property of scale-free graphs and the real-world networks inspiring them which has received much attention is their 'robustness'.

Suppose we delete vertices independently at random from $G_m^{(n)}$, each with probability q. What is the structure of the remaining graph? Is it connected? Does it have a giant component? A precise form of the question is: fix $0 < q < 1$. Suppose vertices of $G_m^{(n)}$ are deleted independently at random with probability $q = 1 - p$. Let the graph resulting be denoted G_p. For which p is there a constant $c = c(p) > 0$ independent of n such that with high probability G_p has a component with at least cn vertices? What is the critical value of p below which no such constant exists?

As noted in [16], it is easy to see from the neighbourhood expansion argument used there that there is no critical p. Once the neighbourhoods of a given vertex reach a certain size (which happens with some positive probability), they continue expanding, and the vertex is almost certainly joined to the vertex surviving in G_p with lowest index. However, it turns out that this 'giant' component becomes very small as p approaches zero. To estimate its size we can use the pairing model to relate the structure of $G_m^{(n)}$ to a certain scale-free branching process; here we shall only give an outline of the argument, referring the reader to [18] for the rather technical details.

Theorem 19. *Let $m \geq 2$ and $0 < p < 1$ be fixed, and let G_p be obtained from $G_m^{(n)}$ by deleting vertices independently with probability $1 - p$. Then as $n \to \infty$,* **whp** *the largest component of G_p has order $(c(p, m) + o(1))n$ for some constant $c(p, m) > 0$. Furthermore, as $p \to 0$ with m fixed, $c(p, m) = \exp(1/O(p))$.*

Proof. Recall from the pairing model that each vertex i sends out m edges, to targets $t_{i,1}, \ldots, t_{i,m}$, where the $t_{i,r}$ are independent and for $j \leq i$ we have, essentially,

$$\Pr(t_{i,r} = j) = w_j / W_i$$

for random quantities w_j and $W_i = \sum_{j=0}^{i-1} w_i$ defined earlier. To a good approximation (in a sense we shall not make precise here), as long as $i \to \infty$ we can replace W_i by the value $\sqrt{i/n}$ around which it is concentrated. Also, w_i is the 'waiting time' from W_{i-1} for m samples from a distribution with density $2mnx$, $0 < x < 1$. So to a good approximation the w_i are given by

$$w_i = \frac{Z_i}{2mnW_i} = \frac{Z_i}{2m\sqrt{in}},$$

where the Z_i are iid, each with a distribution Z the sum of m independent exponential random variables with parameter 1, so Z has density

$$f_Z(x) = \frac{x^{m-1}e^{-x}}{(m-1)!}. \tag{13}$$

To this degree of approximation we have

$$\Pr(t_{i,r} = j) = \frac{w_j}{W_i} \sim \frac{Z_j}{2m\sqrt{jn}}\sqrt{\frac{n}{i}} = \frac{Z_j}{2m\sqrt{ij}}. \tag{14}$$

Re-scaling, let us write $i = \alpha n$, and consider the probability that $t_{i,r} = j$ for some j with $j \in [\beta n, (\beta + \mathrm{d}\beta)n]$ and $Z_j \in [y, y + \mathrm{d}y]$. Since there are $n\mathrm{d}\beta$ vertices j in this range, of which a fraction $f_Z(y)\mathrm{d}y$ will have Z_j in the required interval, this probability is approximately

$$n\mathrm{d}\beta f_Z(y)\mathrm{d}y \frac{y}{2m\sqrt{\alpha n \beta n}} = \frac{y f_Z(y)}{2m\sqrt{\alpha\beta}} \mathrm{d}y\mathrm{d}\beta, \tag{15}$$

provided $\beta < \alpha$. Similarly, let us fix a vertex i with $i = \alpha n$ and $Z_i = x$ and consider the probability that there is a vertex $j > i$ with $j \in [\beta n, (\beta + \mathrm{d}\beta)n]$ and $Z_j \in [y, y + \mathrm{d}y]$ sending an edge to i. Again there are $n\mathrm{d}\beta$ vertices j in the right range, a fraction $f_Z(y)\mathrm{d}y$ of which have Z_j in the right range, so using (14) again this probability is approximately

$$mn\mathrm{d}\beta f_Z(y)\mathrm{d}y \frac{x}{2m\sqrt{\alpha n \beta n}} = \frac{x f_Z(y)}{2\sqrt{\alpha\beta}} \mathrm{d}y\mathrm{d}\beta. \tag{16}$$

(Here the initial factor of m comes from each vertex sending out m edges independently).

Motivated by the above let us define a birth process as follows: in generation $t \geq 0$ there will be a finite number $N(t)$ of 'vertices'. Each vertex v has three numbers associated with it: $\alpha(v) \in (0, 1)$, corresponding to the α in $i = \alpha n$ above, $x(v)$, corresponding to Z_i above, and an integer $l(v)$ which will be either m or $m-1$. This tells us the number of 'left-children' of v: as we work outwards from an initial vertex of $G_m^{(n)}$ finding all vertices at distance 1, then distance 2, etc., there are two ways we can reach a new vertex w from an old vertex w'; from the right (so $w < w'$ and $t_{w',r} = w$ for some r) or from the left. In the next step there will be m or $m - 1$ new left-children of w (vertices $t_{w,s}$, $1 \leq s \leq m$) respectively.

A vertex v in generation t with $\alpha(v) = \alpha$ and $x(v) = x$ gives rise to provisional offspring in generation $t + 1$ as follows: v gives rise independently to exactly $l(v)$ provisional left-children $w_1, \ldots, w_{l(v)}$. For each i we have $l(w_i) = m$, and the values $\beta_i = \alpha(w_i)$ and $y_i = x(w_i)$ are chosen according to the density (15), with $0 < \beta_i \leq \alpha$. Also, v gives rise to a Poisson number of provisional right-children w, each with $l(w) = m - 1$, with the chance of v giving rise to a provisional right-child w having $\alpha(w) \in [\beta, \beta + \mathrm{d}\beta]$ and $x(w) \in [x, x + \mathrm{d}x]$ given by (16), for $\alpha \leq \beta < 1$.

To obtain the next generation, we take all the provisional children of the current generation, and keep each with probability p, independently of the others.

Let us write $\mathcal{N} = \mathcal{N}_p^{l,\alpha,x}$ for the process defined above, starting with a single vertex v having $l(v) = l$, $\alpha(v) = \alpha$ and $x(v) = x$. Note that the definition of $\mathcal{N}$ does not involve n. Let us write $c(v)$ for the total number of descendants of v in all generations, which may be infinite. There is a certain 'survival probability' $s(p, l, \alpha, x) = \Pr(c(v) = \infty)$. By elementary probability theory, we have

$$s(p, l, \alpha, x) = \lim_{t \to \infty} \Pr(c(v) \geq t),$$

and hence

$$s(p,l,\alpha,x) = \Pr(c(v) \geq (\log n)^{10}) + o(1), \tag{17}$$

say, as $n \to \infty$. From the remarks of the last few paragraphs, when n is large the process $\mathcal{N}_1^{m,\alpha,x}$ gives a good approximation to the initial growth (up to $(\log n)^{10}$ vertices, say) of the neighbourhoods of a vertex i in $G_m^{(n)}$ with $i = \alpha n$ and $Z_i = x$. Once the neighbourhoods reach size $(\log n)^{10}$, with high probability i is in the giant component. Using (17), it follows that for α bounded away from 0 and 1 and for any fixed x, the probability that such a vertex lies in the giant component of G_p is given by $s(p,m,\alpha,x) + o(1)$.

Fix $0 < p < 1$; from now on we suppress dependence on p. Let us write $L(\alpha)$ for the chance that a particular potential left-child w of a vertex v with $\alpha(v) = \alpha$ and $x(v) = x$ itself survives, and has $c(w) = \infty$. Note that this probability does not depend on x. Also, let us write $r(\alpha, x)$ for the chance that *some* potential right-child w of v survives and has $c(w) = \infty$. Since $c(v)$ is finite if and only if $c(w)$ is finite for all surviving children w of v, we have

$$s(p,l,\alpha,x) = 1 - (1 - L(\alpha))^l (1 - r(\alpha,x)). \tag{18}$$

Since the density (16) is proportional to x, it is easy to see that $r(\alpha,x)$ has the form

$$r(\alpha,x) = 1 - e^{-xR(\alpha)}$$

for some function R depending on α only. (R is the negative of the log of the chance that all right-children of v die out, if $\alpha(v) = \alpha$ and $x(v) = 1$.) Using again that $c(v) = \infty$ if and only if at least one child w of v survives and has $c(w) = \infty$, it is easy to see that

$$L(\alpha) = p \int_{\beta=0}^{\alpha} \int_{y=0}^{\infty} \left(1 - (1 - L(\beta)^m) e^{-yR(\beta)}\right) \frac{y f_Z(y)}{2m\sqrt{\alpha\beta}} \mathrm{d}y \mathrm{d}\beta.$$

Indeed, given v, the last factor gives (from (15)) the chance that the particular potential left-child w of v we are considering has $\alpha(w) = \beta$ and $x(w) = y$. From (18) the factor in brackets is the chance that $c(w) = \infty$. The first factor p is the chance that w itself is actually born in the first place. From the form (13) of $f_Z(y)$ it turns out that the y integral is easy to do, giving

$$L(\alpha) = \frac{p}{2\sqrt{\alpha}} \int_{\beta=0}^{\alpha} \frac{1}{\sqrt{\beta}} \left(1 - \frac{(1 - L(\beta)^m)}{(1 + R(\beta))^{m+1}}\right) \mathrm{d}\beta. \tag{19}$$

Similarly, one can check that

$$R(\alpha) = \frac{p}{2\sqrt{\alpha}} \int_{\beta=\alpha}^{1} \frac{1}{\sqrt{\beta}} \left(1 - \frac{(1 - L(\beta)^{m-1})}{(1 + R(\beta))^{m}}\right) \mathrm{d}\beta. \tag{20}$$

The pair of equations above may have more than one solution (in particular, both L and R identically zero is a solution); standard probability theory tells us that the functions L and R are the (unique) maximal solution to (19),(20).

From the equations above it is easy to deduce Theorem 19 (as well as more detailed results). Using monotonicity of the functionals on the right hand side one can find upper and lower bounds on $L(\alpha)$ and $R(\alpha)$ and hence on the expected size of the giant component of G_p. Concentration follows immediately since the proof shows that all vertices are in very small components (order $O((\log n)^{10})$, say) or the giant component, and the number of vertices in very small components can be shown to be concentrated by standard martingale methods. □

The situation when the graph $G_m^{(n)}$ is deliberately attacked is rather different. Given the actual (random) graph, determining the 'best' attack is not an easy problem, and is in any case not realistic. Here we consider the natural, simple approach of deleting the earliest vertices, up to some cutoff cn. This time there is a value $c < 1$ beyond which there is no giant component in what remains.

Theorem 20. *Let G_c be obtained from $G_m^{(n)}$ by deleting all vertices with index less than cn, where $0 < c < 1$ is constant, and let $c_m = (m-1)/(m+1)$. If $c < c_m$ then* **whp** *G_c has a component with $\Theta(n)$ vertices. If $c > c_m$ then* **whp** *G_c has no such component.*

Proof. We just give an outline, using the methods above. Considering the quantities analogous to $L(\alpha)$ and $R(\alpha)$ but defined for G_c, we obtain equations identical to (19) and (20) except that now $p = 1$, and the lower limit in the integral in (19) is c rather than 0. Near the critical probability, the functions L and R will be small, and hence close to the solution of the linearized form of the equations. It is easy to solve these linearized equations; a non-zero solution exists if and only if $c = c_m$, and one can deduce the result. □

15 The case $m = 1$: plane-oriented recursive trees

A simple special case of the BA model that has been considered in several papers is the $m = 1$ case, where each vertex sends a single edge to an earlier vertex, giving rise to a tree. In this context the LCD model is not the most natural interpretation of the BA description, as it gives rise to a forest with loops. It turns out that essentially the $m = 1$ case of the BA model had been considered more than a decade before [5].

For $m = 1$, the generally accepted interpretation of the imprecise description in [5] is to start with one vertex, the *root* which has an extra 'virtual edge' coming in to it from nowhere, so the degree of the root at the start counts as 1. Thus at time t, when there are t vertices, although there are $t-1$ edges in the tree, the effective sum of the degrees is $2(t-1)+1$. As has occasionally been pointed out, this precise version of the $m = 1$ model is not at all new; it is exactly the standard model for random plane-oriented recursive trees. A tree on a labelled vertex set $V = \{1, 2, \ldots, t\}$ is *recursive* if each vertex other than 1 is joined to exactly one earlier vertex. In other words, the tree can be grown by adding the vertices in numerical order, joining each new vertex to some old vertex. Uniform random recursive trees, grown one vertex at a time by joining the new vertex to an old vertex chosen uniformly at random, have been studied for some time; see, for example, the survey [44].

A *plane-oriented* tree is one with a cyclic order at each vertex, induced, for example, by drawing the tree in the plane. When a new vertex v is added to a plane-oriented recursive tree T and joined to an existing vertex w, the number of different plane-oriented recursive trees that may result is given by the degree d of w in T, as there d different ways in which the new edge can meet the vertex w. In fact, as in the BA model, the standard definition treats the first vertex, the root, differently, effectively imagining an edge from the root going off to infinity. In this way branches of plane-oriented recursive trees are again plane-oriented recursive trees. Plane oriented recursive trees were introduced by Szymański [54] in 1987 (although with a slightly different treatment of the root) and have been studied since in several papers, including [8, 43, 45, 54].

For this section by T_n we shall mean the random plane-oriented recursive tree with n vertices, with vertex set $\{1, 2, \ldots n\}$, or, equivalently, the Barabási-Albert scale-free random tree given by (a precise version of) the $m = 1$ case of the model introduced in [5]. Formally, T_1 consists of a single vertex 1 with no edges. For $n \geq 2$, given T_{n-1}, the tree T_n is constructed by adding a new vertex n and joining it to and old vertex v, $1 \leq v \leq n-1$, with

$$\Pr(v = j) = \frac{d_{n-1}(j)}{2n-3}$$

for $j \geq 2$ and

$$\Pr(v = 1) = \frac{d_{n-1}(1) + 1}{2n-3},$$

where $d_{n-1}(j)$ is the degree of the vertex j in the tree T_{n-1}. As the tree grows out from the root we shall say that the new vertex added as above is a *child* of v, and that v is its *parent*.

Note that definition of T_n is very similar to that of $G_1^{(n)}$ as given in section 7.

Since T_n has been around for some time, it is not surprising that various properties are known already. For example, the 'load scaling' considered in [53] and [35] was already determined in a more precise form in [45] (see also [17]). In contrast, while [53] claims that the distribution of shortest path lengths has also been established, this is not the case as far as we know; it is certainly not in the references cited.

Here we shall show how to calculate exactly the probability that a certain subgraph is present in T_n. A form of this result for $G_1^{(n)}$ was used in section 12 to study clustering and small subgraphs of $G_m^{(n)}$. From this result one can also obtain the distribution of shortest paths.

Based on the work in [16] it is possible to write down a formula for the probability that a particular subgraph is present in T_n. The key is to consider a certain sequence of expectations that can be calculated inductively. As the formulae are a little complicated, we start with some motivation. Note that throughout this section, as in section 12, we are asking whether a particular subgraph S is present in T_n, not whether some subgraph *isomorphic* to S is present. Thus for example S may consist of the edges $\{2,3\}$, $\{3,7\}$ and $\{2,8\}$. Then $\Pr(S \subset T_n)$ is taken to mean the probability that in T_n we have exactly these edges (so the parent of 3 is 2, that of 7 is 3 and that of 8 is 2), not the probability that T_n contains a path of 3 edges.

Let us write f_t for the parent of vertex t. Then, given T_t, the probability that f_{t+1} is i is proportional to the degree of i in T_t; more precisely, from the definition of T_n we have $\Pr(f_{t+1} = i \mid T_t) = \frac{d_t(i)}{2t-1}$, where $d_t(i)$ is the degree of i in T_t. Taking expectations, we see that $\Pr(f_{t+1} = i) = \mathbb{E}(d_t(i))/(2t-1)$, so we would like to know the expectations $\mathbb{E}(d_t(i))$. These are easy to calculate: $\mathbb{E}(d_i(i)) = 1$ for all $i \geq 1$, and in going from T_t to T_{t+1} the degree of i increases by one if $f_{t+1} = i$, which happens with probability $d_t(i)/(2t-1)$. Thus

$$\mathbb{E}(d_{t+1}(i)) = \mathbb{E}(d_t(i)) + \frac{\mathbb{E}(d_t(i))}{2t-1} = \frac{2t}{2t-1}\,\mathbb{E}(d_t(i)),$$

giving $\mathbb{E}(d_t(i)) = \prod_{s=i}^{t-1} \frac{2s}{2s-1}$.

More generally, suppose that S is a fixed graph consisting of a set S' of edges with both ends numbered less than j, say, and one more edge from k to i, with $i \leq j < k$. Given that $S' \subset T_j$ we will have $S \subset T_k$ if and only if $f_k = i$, and the probability of the latter event depends on the degree of i, as before. Thus we would like to calculate $\mathbb{E}(d_t(i) \mid S' \subset T_t)$. Once we have this expectation for $t = j$, its values for $t = j+1, \ldots, k-1$ can be calculated inductively as before. The problem occurs when two edges end at the same vertex: if $S' = S'' \cup \{ij\}$, say, then while we have $\Pr(S' \subset T_j) = \mathbb{E}(d_{j-1}(i) \mid S'' \subset T_{j-1})/(2j-3)$, the event that $f_j = i$ is more likely when $d_{j-1}(i)$ is large, so $\mathbb{E}(d_j(i) \mid S' \subset T_j)$ is not related in a simple way to quantities we have already calculated.

The key turns out to be to consider rising factorials: $[d]_r = d(d+1)\cdots(d+r)$. We keep track of an expectation involving $[d_t(i)]_r$ for each vertex i of the subgraph S which will have r edges coming into it in the future (times later than t).

From now on, fix a graph S which is possible as a subgraph of T_n for large n. (So S has no loops, and for each vertex i, there is at most one edge in S from i to an earlier vertex.) For $t \geq i$ let $R_t(i)$ be the number of $j > t$ such that ij is an edge of S, i.e., the number of edges in S coming in to i after time t. Let S_t consist of those edges ij of S for which $i, j \leq t$, let

$$X_t = \prod_{ij \in E(S_t)} \mathbb{I}_{ij \in E(T_t)} \prod_{i \in V(S), i \leq t} [d_t(i)]_{R_t(i)},$$

and set $\lambda_t = \mathbb{E}(X_t)$. Here $\mathbb{I}_{\mathcal{A}}$ is the indicator function of the event $\mathcal{A}$. Note that $\lambda_0 = 1$, while for t large (at least the largest vertex in S) we have $X_t = \mathbb{I}_{S \subset T_t}$, so $\lambda_t = \Pr(S \subset T_t)$, the quantity we wish to calculate.

Lemma 21. *For $t \geq 0$ we have*

$$\lambda_{t+1} = R_{t+1}(t+1)! \frac{1}{2t-1} \lambda_t \tag{21}$$

if there is an edge $\{k, t+1\}$ in S with $k \leq t$, and

$$\lambda_{t+1} = R_{t+1}(t+1)! \left(1 + \frac{C_S(t+1)}{2t-1}\right) \lambda_t \tag{22}$$

otherwise, where $C_S(t+1)$ is the number of edges $ij \in E(S)$ with $i \leq t$ and $j > t$.

Proof. Since in T_{t+1} the degree of $t+1$ is always exactly 1, we can write X_{t+1} as

$$X_{t+1} = R_{t+1}(t+1)! Y, \tag{23}$$

where

$$Y = \prod_{ij \in E(S_{t+1})} \mathbb{I}_{ij \in E(T_{t+1})} \prod_{i \in V(S), i \leq t} [d_{t+1}(i)]_{R_{t+1}(i)}.$$

Suppose first that S does not contain an edge $\{k, t+1\}$ with $k \leq t$. Then $S_{t+1} = S_t$, and for each $i \leq t$ we have $R_{t+1}(i) = R_t(i)$. Also, as each edge ij of $S_{t+1} = S_t$ has $i, j \leq t$, for each such edge we have $\mathbb{I}_{ij \in E(T_{t+1})} = \mathbb{I}_{ij \in E(T_t)}$. Thus, in this case,

$$Y = \prod_{ij \in E(S_t)} \mathbb{I}_{ij \in E(T_t)} \prod_{i \in V(S), i \leq t} [d_{t+1}(i)]_{R_t(i)}.$$

Note that this is exactly the formula for X_t except that $d_t(i)$ has been replaced by $d_{t+1}(i)$. Now let us fix T_t, and hence X_t, and consider the random choice of f_{t+1}, the vertex that the next new vertex joins to. We will have $Y = X_t$ unless f_{t+1} is a vertex of S, as all relevant degrees will be the same in T_{t+1} as in T_t. What happens if $f_{t+1} = j$ for some $j \in V(S)$? Then $d_{t+1}(j) = d_t(j) + 1$, so

$$[d_{t+1}(j)]_{R_t(j)} = [d_t(j) + 1]_{R_t(j)} = \frac{d_t(j) + R_t(j)}{d_t(j)} [d_t(j)]_{R_t(j)},$$

and, as all other degrees stay the same, $Y - X_t = X_t R_t(j)/d_t(j)$. Now for each $j \in V(S)$, $j \leq t$, the probability that $f_{t+1} = j$ is just $d_t(j)/(2t-1)$. Thus the expected difference $Y - X_t$ is given by

$$\mathbb{E}(Y - X_t) = \sum_{j \in V(S), j \leq t} \frac{d_t(j)}{2t-1} X_t R_t(j)/d_t(j) = X_t \frac{C_S(t+1)}{2t-1}.$$

Taking expectations of both sides gives $\mathbb{E}(Y) = (1 + C_S(t+1)/(2t-1))\, \mathbb{E}(X_t)$, which together with (23) proves (22).

We now turn to (21). Suppose that $\{k, t+1\}$ is an edge of S with $k \leq t$. Given T_t, we have $Y = 0$ unless $f_{t+1} = k$, which happens with probability

$d_t(k)/(2t-1)$. Supposing that f_{t+1} is equal to k, what is Y? In this case $S_{t+1} = S_t \cup \{k, t+1\}$, and, since we have $\{k, t+1\} \in E(T_{t+1})$, we have

$$\prod_{ij \in E(S_{t+1})} \mathbb{I}_{ij \in E(T_{t+1})} = \prod_{ij \in E(S_t)} \mathbb{I}_{ij \in E(T_t)}.$$

For $i \leq t$, $i \neq k$ we have $d_{t+1}(i) = d_t(i)$, while $d_{t+1}(k) = d_t(k) + 1$. Also, $R_{t+1}(i) = R_t(i)$ for $i \neq k$, while $R_{t+1}(k) = R_t(k) - 1$. Thus

$$\begin{aligned} \prod_{i \in V(S), i \leq t} [d_{t+1}(i)]_{R_{t+1}(i)} &= [d_t(k)+1]_{R_t(k)-1} \prod_{i \in V(S), i \leq t, i \neq k} [d_t(i)]_{R_t(i)} \\ &= \frac{1}{d_t(k)} \prod_{i \in V(S), i \leq t} [d_t(i)]_{R_t(i)}. \end{aligned}$$

Thus, in this case, if $f_{t+1} = k$ then we have $Y = X_t/d_t(k)$. Since this event has probability $d_t(k)/(2t-1)$, we have $\mathbb{E}(Y \mid T_t) = X_t/(2t-1)$. Taking the expectation of both sides gives $\mathbb{E}(Y) = \mathbb{E}(X_t)/(2t-1) = \lambda_t/(2t-1)$. Together with (23) this proves (21). $\square$

Lemma 21 has the following immediate consequence. For a possible subgraph S of T_n, orient each edge $ij \in E(S)$ with $i < j$ from j to i. As in section 12, we write $V^+(S)$ for the set vertices of S from which edges leave, and $V^-(S)$ for those vertices at which edges arrive. (These sets are not in general disjoint.) For $i \in V^-(S)$ let $d_S^{in}(i)$ be the in-degree of i in S (so $d_S^{in}(i) = R_i(i)$), let $d_S^{out}(i)$ be the out-degree, and $d_S(i)$ the total degree of i in S.

Corollary 22. *Let S be a possible subgraph of T_n. With the notation above, the probability p_S that $S \subset T_n$ satisfies*

$$p_S = \prod_{i \in V^-(S)} d_S^{in}(i)! \prod_{i \in V^+(S)} \frac{1}{2i-3} \prod_{t \notin V^+(S)} \left(1 + \frac{C_S(t)}{2t-3}\right). \tag{24}$$

Furthermore,

$$p_S = \prod_{i \in V^-(S)} d_S^{in}(i)! \prod_{ij \in E(S)} \frac{1}{2\sqrt{ij}} \exp\left(O\left(\sum_{i \in V(S)} C_S(i)^2/i\right)\right). \tag{25}$$

Proof. The first statement follows immediately from Lemma 21; replace t by $t-1$ in (21) and (22), and write $p_S = \lambda_n = \lambda_n/\lambda_0$ as the product of the factors appearing in these equations.

The second statement follows by simple approximations: for all $x \geq 0$ we have $\log(1+x) = x + O(x^2)$. Thus for $c \geq 0$,

$$\log\left(\prod_{t=i+1}^{j-1} \left(1 + \frac{c}{2t-3}\right)\right) = \sum_{t=i+1}^{j-1} \left(\frac{c}{2t} + O(c^2/t^2)\right) = \frac{c}{2}\log\left(\frac{j}{i}\right) + O(c^2/i).$$

Writing the vertices of S in order as $v_1, \ldots, v_l$, let $c_k = C_S(v_k+1)$ be the number of edges in S from $\{v_{k+1}, \ldots, v_l\}$ to $\{v_1, \ldots, v_k\}$. Then for $v_k < t < v_{k+1}$ we have $C_S(t) = c_k$, so

$$p_S = \prod_{i \in V^-(S)} d_S^{in}(i)! \prod_{i \in V^+(S)} \frac{1}{2i} \prod_{k=1}^{l-1} (v_{k+1}/v_k)^{c_k/2} \exp\left(O\left(\sum_{i \in V(S)} C_S(i)^2/i\right)\right).$$

(One can easily check that the error from replacing $2i-3$ by $2i$ in the second product is absorbed by the final error term.) Now the final exponent of v_k is just

$$-d_S^{out}(v_k) + c_{k-1}/2 - c_k/2.$$

Since $c_k = c_{k-1} - d_S^{out}(v_k) + d_S^{in}(v_k)$, this is just $-d_S(v_k)/2$. Finally, there is one factor of two in the denominator for each edge, so (25) follows. $\square$

Essentially this result, but stated for the related model $G_1^{(n)}$, was used in section 12 to find the clustering coefficient of $G_m^{(n)}$. We finish this section with a direct application of Corollary 22 to T_n from [17]. We write $E_k = E_k(n)$ for the expected number of (shortest) paths in T_n of length k, so $\sum_{k=1}^{\infty} E_k(n) = \binom{n}{2}$.

Theorem 23. *Suppose that $k = k(n)$ satisfies $k/\log n \to \alpha$, where $0 < \alpha < e$. Then*

$$E_k = \Theta(n^{1+\alpha \log(e/\alpha)}/\sqrt{\log n}), \tag{26}$$

as $n \to \infty$. Furthermore, if $k = \log n + x\sqrt{\log n}$ where $x = x(n) = o(\log n)$, then

$$E_k \sim \frac{n^2}{2} \frac{1}{\sqrt{2\pi \log n}} e^{-x^2/2} \tag{27}$$

as $n \to \infty$.

Note that the second statement says that the distribution of path lengths is asymptotically normal with mean and variance $\log n$.

16 Conclusion

Most but not all of the rigorous results concerning models of large-scale real-world networks we have reviewed confirm the computer experiments and heuristic calculations performed on these models. In many cases, the results are surprisingly difficult to prove, and need techniques not used in the theory of classical random graphs. However, much remains to be done.

There is a great need for models of real-life networks that incorporate many of the important features of these systems, but can still be analyzed rigorously. The models defined and analyzed in this article are too simple for many applications, but there is also a danger in constructing models which take into account too many features of the real-life networks. Beyond a certain point, a complicated model hardly does more than describe the particular network that inspired it.

If the right balance can be struck, well constructed models and their careful analysis should give a sound understanding of growing networks that can be used to answer practical questions about their current state, as well as to predict their future development.

References

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

[2] R. Albert and A.-L. Barabási, Statistical mechanics of complex networks, *Rev. Mod. Phys.* **74** (2002), 47–97.

[3] R. Albert, H. Jeong and A.-L. Barabási, Diameter of the world-wide web, *Nature* **401** (1999), 130–131.

[4] Austin, T.L., Fagen, R.E., Penney, W.F., and Riordan, J., The number of components of random linear graphs, *Ann. Math. Statist.* **30** (1959), 747–754.

[5] A.-L. Barabási and R. Albert, Emergence of scaling in random networks, *Science* **286** (1999), 509–512.

[6] A.-L. Barabási, R. Albert and H. Jeong, Mean-field theory for scale-free random networks, *Physica A* **272** (1999), 173-187.

[7] A.-L. Barabási, R. Albert and H. Jeong, Scale-free characteristics of random networks: the topology of the world-wide web, *Physica A* **281** (2000), 69–77.

[8] F. Bergeron, P. Flajolet and B. Salvy, Varieties of increasing trees, in Proc. 17th Colloq. on Trees in Algebra and Programming held in Rennes, *Lecture Notes in Computer Science* **581**, Springer Verlag, Berlin, 1992, pp 24–48.

[9] Bollobás, B., The diameter of random graphs, *Trans. Amer. Math. Soc.* **267** (1981), 41–52.

[10] Bollobás, B., The evolution of random graphs. *Trans. Amer. Math. Soc.* **286** (1984), 257–274.

[11] B. Bollobás, *Random Graphs, Second Edition,* Cambridge studies in advanced mathematics, vol. 73, Cambridge University Press, Cambridge, 2001, xviii + 498 pp.

[12] Bollobás, B., Borgs, C., Chayes, T., and Riordan, O.M., Directed scale-free graphs, preprint.

[13] B. Bollobás and F.R.K. Chung, The diameter of a cycle plus a random matching, *SIAM J. Discrete Math.* **1** (1988), 328–333.

[14] Bollobás, B. and Fernandez de la Vega, W., The diameter of random regular graphs. *Combinatorica* **2** (1982), 125–134.

[15] B. Bollobás and O.M. Riordan, Linearized chord diagrams and an upper bound for Vassiliev invariants, *J. Knot Theory Ramifications* **9** (2000), 847–853.

[16] Bollobás, B. and Riordan, O.M., The diameter of a scale-free random graph, to appear in *Combinatorica*. (Preprint available from http://www.dpmms.cam.ac.uk/∼omr10/.)

[17] Bollobás, B. and Riordan, O.M., On shortest paths and load-scaling in scale-free trees, preprint.

[18] Bollobás, B. and Riordan, O.M., Robustness and vulnerability of scale-free random graph, in preparation.

[19] B. Bollobás, O. Riordan, J. Spencer and G. Tusnády, The degree sequence of a scale-free random graph process, *Random Structures and Algorithms* **18** (2001), 279–290.

[20] Buckley, P.G. and Osthus, D., Popularity based random graph models leading to a scale-free degree sequence, preprint available from http://www.informatik.hu-berlin.de/∼osthus/.

[21] Burtin, Ju.D., Asymptotic estimates of the diameter and the independence and domination numbers of a random graph, *Dokl. Akad. Nauk SSSR* **209** (1973), 765–768, translated in *Soviet Math. Dokl.* **14** (1973), 497–501.

[22] Burtin, Ju.D., Extremal metric characteristics of a random graph. I, *Teor. Verojatnost. i Primenen.* **19** (1974), 740–754.

[23] Cohen, R., and Havlin, S., Ultra small world in scale-free networks, http://www.arxiv.org/abs/cond-mat?0205476

[24] C. Cooper and A. Frieze, A general model of web graphs, preprint.

[25] E. Drinea, M. Enachescu and M. Mitzenmacher, Variations on random graph models for the web, technical report, Harvard University, Department of Computer Science, 2001.

[26] S.N. Dorogovtsev, J.F.F. Mendes, A.N. Samukhin, Structure of growing networks with preferential linking *Phys. Rev. Lett.* **85** (2000), 4633.

[27] S.N. Dorogovtsev and J.F.F. Mendes, Evolution of networks, *Adv. Phys.* **51** (2002), 1079.

[28] Erdős, P., and Rényi, A., On random graphs I., *Publicationes Mathematicae Debrecen* **5** (1959), 290–297.

[29] Erdős, P., and Rényi, A., On the evolution of random graphs, *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **5** (1960), 17–61.

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

[31] I.J. Farkas, I. Derényi, A.-L. Barabási and T. Vicsek, Spectra of "real-world" graphs: Beyond the semicircle law, *Phys. Rev. E* **64** (2001), 026704.

[32] Gilbert, E.N., Enumeration of labelled graphs, *Canad. J. Math.* **8** (1956), 405–411.

[33] Gilbert, E.N., Random graphs, *Annals of Mathematical Statistics* **30** (1959), 1141–1144.

[34] Gilbert, N., A simulation of the structure of academic science, *Sociological Research Online* **2** (1997).

[35] K.-I. Goh, B. Kahng and D. Kim, Universal behavior of load distribution in scale-free networks, *Physical Review Letters* **87** (2001), 278701.

[36] P. Holme and B.J. Kim, Growing scale-free networks with tunable clustering, *Phys. Rev. E* **65** (2002), 026107.

[37] Janson, S., Knuth, D., E., Łuczak, T., and Pittel, B., The birth of the giant component, *Random Structures and Algorithms* **3** (1993), 233–358.

[38] Janson, S., Łuczak, T., and Ruciński, A. (2000). *Random Graphs.* John Wiley and Sons, New York, xi+ 333pp.

[39] K. Klemm and V.M. Eguiluz, *Phys. Rev. E* **65** (2002), 057102.

[40] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins and E. Upfal, Stochastic models for the web graph, FOCS 2000.

[41] Lotka, A.J., The frequency distribution of scientific productivity, *J. of the Washington Acad. of Sci.* **16** (1926), 317.

[42] Łuczak, T., Component behavior near the critical point of the random graph process, *Random Structures and Algorithms* **1** (1990), 287–310.

[43] H.M. Mahmoud, Distances in random plane-oriented recursive trees, *J. Comput. and Applied Math.* **41** (1992), 237–245.

[44] H.M. Mahmoud and R.T. Smythe, A survey of recursive trees, *Th. of Probability and Math. Statistics* **51** (1995), 1–27.

[45] H.M. Mahmoud, R.T. Smythe and J. Szymański, On the structure of random plane-oriented recursive trees and their branches, *Random Struct. Alg.* **4** (1993), 151–176.

[46] McKay, B.D., and Wormald, N.C., The degree sequence of a random graph. I. The models, *Random Structures and Algorithms* **11** (1997), 97–117.

[47] Milgram, S., The small world phenomenon, *Psychol. Today* **2** (1967), 60–67.

[48] M.E.J. Newman, S.H. Strogatz and D.J. Watts, Random graphs with arbitrary degree distribution and their applications, *Physical Review E* **64** (2001), 026118.

[49] B. Pittel, Note on the heights of random recursive trees and random m-ary search trees, *Random Struct. Alg.* **5** (1994), 337–347.

[50] Riddell, R.J.Jr., and Uhlenbeck, G.E., On the theory of virial development of the equation of state of monoatomic gases, *J. Chem. Phys.* **21** (1953), 2056–2064.

[51] Simon, H., On a class of skew distribution functions, *Biometrika* **42** (1955), 425–440.

[52] A. Stoimenow, Enumeration of chord diagrams and an upper bound for Vassiliev invariants, *J. Knot Theory Ramifications* **7** (1998), 93–114.

[53] G. Szabó, M. Alava and J. Kertész, Shortest paths and load scaling in scale-free trees, *Phys. Rev. E* **66** (2002), 026101.

[54] J. Szymański, On a nonuniform random recursive tree, *A.. Discrete Math.* **33** (1987), 297–306.

[55] D. J. Watts and S. H. Strogatz, Collective dynamics of 'small-world' networks, *Nature* **393** (1998), 440–442.

[56] Zipf, G.K., Human behavior and the principle of least effort, New York: Hafner, 1949.