Artificial Generation of Power-Law Graphs: A Historical Survey

Ross Rheingans-Yoo
Harvard School of Engineering and Applied Sciences, Cambridge, MA (USA)
rry@eecs.harvard.edu

Abstract

We provide an overview of the theory and practice of algorithmically generating graphs with power-law properties, summarize the history of their development, describe the present state of the art, and indicate promising directions for future inquiry. In particular, we highlight the practical and theoretical differences between pre-2004 generalizations of the Barabási–Albert model and more-recently-introduced methods applying Kronecker exponentiation to produce scale-free subgraph structures. We conclude with a minor (but, to our knowledge, novel) result in the probabilistic analysis of nonlinear Barabási–Albert processes.

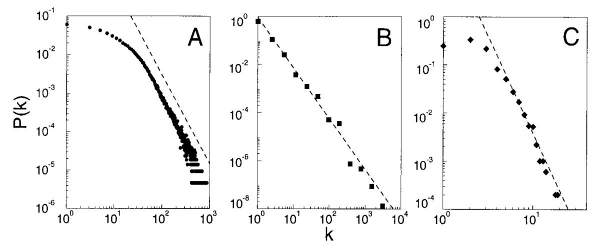

Figure 1. Figure reproduced from [13]: "The distribution function of connectivities for various large networks. **(A)** Actor collaboration graph with $N = 212,250$ vertices and average connectivity $\langle k \rangle = 28.78$. **(B)** [World Wide Web], $N = 325,729$, $\langle k \rangle = 5.46$. **(C)** Power grid data, $N = 4941$, $\langle k \rangle = 2.67$. The dashed lines have slopes (A) $\gamma_{\text{actor}} = 2.3$, (B) $\gamma_{\text{www}} = 2.1$, and (C) $\gamma_{\text{power}} = 4$." Note that the tail behavior approximated by a $y = x^\gamma$ power law (appearing on the log-log scale as a straigt line) diverges from the exponential $y = c^x$ behavior (which would appear as an inverted exponential) predicted for structureless random graphs.

I. Introduction

Since Erdös's early results in combinatorics, graphs have been an important conceptual tool in the study of computation. While the early study of graphs made gains through the use of probabilistic analysis on simple artificial models, the past few decades have given computer scientists the tools to analyze massive quantities of graph data on real-world networks, collected from the fields of sociology, bibliology[1], biology, civil engineering, and computer engineering (in the World Wide Web and Internet itself). The dawn of the age of empirical graph analytics revealed that many real-world networks exhibit distinctive structure—a small number of nodes serve as endpoints for a disproportionate majority of edges, noncentral nodes are clustered far more tightly than would be expected at random, and such phenomena are reflected fractally as we examine smaller and smaller subnetwork scales.

This realization, and the models proposed to characterize the range of unexpected structures emergent in real-world networks, has shed light on the subtle processes that underlie the growth of networks—from friendships, paper citations, and movie-acting roles to power lines, webpages, and axonal connections—as we continue to seek generative models that are capable of procedurally giving rise to the macro-level phenomena we observe in the world. While no patently correct model has yet emerged, the challenges and successes of the past fifteen years have led to theoretical revelations in computer science, as well as domain insights in other fields.

The current state of the art[2] is more or less able to artificially simulate graphs exhibiting an array of properties resembling real-world networks, but

[1] *i.e.* the study of writing and citation

[2] *i.e.* in the generation of artificial *power-law graphs* or *scale-free networks*; we'll provide a formal definition later

the truly exciting part of the state of research is that modern graph models are increasingly able to capture, quantify, and reproduce complex multi-agent interactions that give rise to self-organizing structures that it are computationally intractible to analyze by traditional dynamic analysis. Moreover, recent theoretical insights show considerable promise in their ability to reduce the simulation of network graphs to what appear to be essential parameters, hopefully shedding insight into the structure of organization itself.

In this survey, we seek to present the technical reader with a condensed history of the development of the theory of power-law graphs, the history of the models that have been proposed to generate them, and an understanding of the technical and mathematical foundations of what is now known about such graphs and their generation.

II. Related Work

Much has already been written about what is so far known about power-law graphs and scale-free networks, and much has already been eloquently summarized by eminent scholars of the field. In 2002, pioneers of the field Albert-László Barabási and Réka Albert published the survey "Statistical mechanics of complex networks" [15], a short yet comprehensive survey of practical graph theory and graph structure, describing the theory of small-world networks, scale-free networks, and what was then known about networks evolving over time, summarizing much earlier results by Erdös and Rényi ([1]), Gilbert ([2]) and Milgram ([3]), as well as contemporary results by Watts and Strogatz ([4]), Kleinberg *et al.* ([22]), Kumar *et al.* ([24]), and Newman (various).

Newman's own 2003 survey, "The structure and function of complex networks" ([8]) provides an early summary of Barabási and Albert's contemporary models of network growth, as well as further analysis of procedural properties of organized networks. Mitzenmacher's "A Brief History of Generative Models for Power Law and Lognormal Distributions" ([9]) not only surveys the array of generative processes then developed, but examines in depth their fundamental statistical assumptions and mechanisms. Chakrabarti and Faloutsos, writing in 2006, would present "Graph Mining: Laws, Generators, and Algorithms" ([26]), formalizing several 'laws' of power-law graphs by combining points of view from physics, mathematics, sociology, and computer science.

No more recent survey that we are aware of, however, has attempted to explore the breadth of the field, owing in no small part to the explosion of popularity (and attendant complexity) it enjoyed in connection with the work of Chakrabarti and Faloutsos themselves. This work, then, represents an effort to summarize in one sweep the early results traced out by previous surveyors, the recent powerful results offered by matrix-based generators, their mathematical relation, and their historical context.

III. General Historical Overview

We identify five main phases of research in this area, discussed in greater detail below:

- Early graph models
- The Barabási–Albert model, and variants
- Community-structured models
- R-MAT and Kronecker-Exponentiation models
- Recent results and open questions in graph topology.

While the study of each was initiated approximately in the order presented, the second and the third—and the third and the fourth—overlapped significantly in the timeframes during which they were actively researched, under different authors—or in some cases, the same ones—working on different paradigms and models in parallel.

By necessity, we will avoid acknowledging every paper germane to the topic, and instead select a few to provide the reader with a general impression of the shape of the field and its historical development. The interested reader is directed to the Further Reference addendum at the end of this paper, as well as the bibliographies of the papers listed in Previous Surveys, for further reading in particular topics. What follows is a brief overview of the history which is covered in depth in §IV.

A. Early Graph Models

The celebrated Erdös-Rényi model ([1]), and the Gilbert approximation ([2]), have proved wildly useful in probabilistic analysis of graphs and a variety of other combinatorial objects, largely because of their simplicity and statistical elegance. Early attempts to use similarly simple methods to generate graphs with other properties (high clustering coefficient in [4], specific degree distribution in [12]) were met with mixed success, either because they failed to capture relevant complexity, or because they proved infeasible for serious analysis. Nevertheless, as a body of literature suggested that a variety of real-world networks exhibited power-law behavior not produced in the ER or Gilbert models (*inter alia* [6]; [7]; [11]; [23]), the search for a probabilistic model of such graphs was a topic of significant interest.

B. Barabási–Albert, Variants

One highly influential model, proposed by Barabási and Albert in [13] and expanded in [14], grows a graph by preferentially attaching edges in sequence in a manner dependent on the instantaneous degree of vertices, in a 'rich-get-richer' fashion. Numerous variations of the preference parameters were suggested for specific contexts ([7]; [17]; [18]; [20] for the Internet topology; [19] for certain subnetworks of the Web graph), though the parameter seemingly most useful (namely, exponent of the preference in response to node degree) would prove challenging to analyze.

C. Community-Structured Models

One failing of these models, however, is that they failed to generate graphs with high clustering coefficient, which is observed empirically in natural networks. To address this failing, a series of papers by Kumar, Kleinberg, Leskovec, Faloutsos, Raghavan, Rajagopalan, and Tomkins ([21]; [22]; [23]; [24]; [25]) used various methods to induce community structure on variants of the Barabási–Albert process, with varying degrees of success.

D. R-MAT, Kronecker Exponentiation

In a significant departure from Barabási–Albert models and sequential preferential attachment in general, Chakrabarti, Zhan, and Faloutsos presented in [27] a model based on Kronecker exponentiation, a combinatorial product of recursive probabilities. This model proved both fruitful for analysis ([28]), easily generalizeable in the generative structure ([29]), able to be fit to graphs by efficient algorithms ([30]), and naturally descriptive of self-similar and scale-free community structure ([31]). It is now considered the modern standard in power-law graph generation, and is used for benchmark loads such as the *Graph500* test for supercomputing clusters ([32]).

E. Open Questions in Topology

Among the open questions still in the field of random graphs, however, is what the fundamental parameters of structured graphs actually *are*. Gupta, Roughgarden, and Seshadhri, in [34], provide some analysis of the *triangle density* of graphs as a metric peer to the edge density studied by Erdös, Rényi, and Gilbert, and we speculate that a return to such fundamentals, combined with the promise of generative models conducive to probabilistic analysis, represents a promising direction for future inquiry.

IV. Survey: Generative Models

Methods for randomly generating graphs with certain properties have been of interest since the introduction of randomness to the study of computation ([1]; [2]). In this section, we describe the progression of models proposed for this purpose in the past twenty-five years as a chronological history of the development of successive models; in §V, we compare several of them more directly and systematically.

A. Precursor: Erdös–Rényi

One of the first model of random graphs proposed for serious computational use was the two-parameter Erdös-Rényi random model (hereafter ER), as described in [1]:

> Our aim is to study the probable structure of a random graph $\Gamma_{n,N}$ which has n given labelled vertices $P_1, P_2, \ldots, P_n$ and N edges; we suppose that these N edges are chosen at random among the $\binom{n}{2}$ possible edges, so that all $\binom{\binom{n}{2}}{N} = C_{n,N}$ possible choices are supposed to be equiprobable.

We note, as an aside, that this model can be approximated by the Gilbert model $\widetilde{G}_{n,p}$ (presented in

[2]) in which each edge is included with probability $p := N/\binom{n}{2}$, in the sense that:

- Conditioned on $\left|\widetilde{E}_{n,p}\right| = N$, the models are equivalently distributed: $\Gamma_{n,N} \sim \widetilde{G}_{n,p}$.
- Any event that occurs with small probability in the $\Gamma_{n,N}$ ER model necessarily occurs with small probability in the corresponding ($\widetilde{G}_{n,p}$) Gilbert model.

In most cases, we will make little distinction between *exact-[edge]count* and *expected-count* models thus related—for more, see any introductory text in computational randomness, such as [36].

For decades (indeed, until the ‘small-world’ network model proposed by Watts and Strogatz in [4]) these were the only practical graph models of significant interest to computer scientists. Despite their seeming simplicity, they remain elegant tools for proving certain graph-theoretic computational propositions via the probabilistic method. Unfortunately, when compared to real-world graphs observed in various scientific domains, ER graphs fail to exhibit certain statistical properties commonly observed, and it was this observation that sparked the first interest in graph-generative models which exhibited (in expectation, or with high probability) properties and structure more complex than total edge-count.

B. Precursor: Small-World

A testament to the enduring mathematical utility of the ER model, however, it was 38 years before an alternate model was published; in [4], Watts and Strogatz observe that real-world networks often exhibit clustering behavior, or ‘cliquishness’, in excess of that observed in corresponding ER graphs, alongside ‘small-world’ properties, or $O(\log n)$ diameter, as earlier posited by Milgram ([3]). By interpolating between the canonical contemporary model for clustered networks—the regular ring lattice—and the ER model, they were able to simultaneously bring the low connectedness of the former and the nonlocality of the latter roughly in line with the values observed in real-world networks.

The Watts–Strogatz model (hereafter WS) produces a random graph from the parameterization (n, k, p) as follows:

- Begin with a regular ring lattice, *i.e.* a graph on n vertices labeled $0, \ldots, n-1$ modulo n with each connected to the $k/2$ immediately to its left and right[3] (wrapping into a circle).
- Each edge of the ring lattice is, with probability p, replaced by a non-duplicate edge chosen uniformly at random among all possible edges (potentially connecting vertices more the $k/2$ indices apart).

An alternative, equidistributed process is to, kn times, add a local edge (*i.e.* between two vertices no more than $k/2$ apart) with probability $1-p$, and a uniformly random edge with probability p. An expected-count process, then, is to add each nonlocal edge with probability $2pk/(n-1)$ and each local edge with probability $(1-p) + 2pk/(n-1)$.

It is apparent (in either the exact- or expected-count case) that in the $p \to 0$ limit, the WS model limits to a regular ring lattice, and in the $p \to 1$ limit, to the ER (or Gilbert) model. What Watts and Strogatz demonstrate in [4], though, is that for small positive valuse of p, the clustering coefficient[4] of the resulting graph remains large (as in a regular ring lattice) while the diameter[5] asymptotically approaches a logarithmic rate (as in an ER graph). As they note, this highly-clustered, low-diameter statistical profile is characteristic of several observed real-world networks (*e.g.* actors appearing together in films, connections between power grid stations, and the neural connectome of the worm *caenorhabditis elegans*), and their proposed model was one of the first that achieved both, and so initiated the study of local and global connectivity properties of networks by use of artificially-generated samples.

C. Power-Law Tail Behavior

This model produces graphs in which vertex degrees are Poisson-distributed—in particular, ex-

[3]Here, as thoughout this survey, we assume implicitly (and hereafter without mention) that such divisions are either integral, can be made so by an appropriate constraint on the free parameters, or can be rounded while introducing insignificant error.

[4]*i.e.* the average percentage of edges between vertices among a vertex’s neighborhood.

[5]*i.e.* the maximum over vertices of the length of the shortest path connecting them, referred to in [4] as the “characteristic path length”.

symbol	(context)	domain	meaning	introduced by
G		(V, E)	graph	Erdös/Rényi, [1]
V		(set)	set of vertices	*ibid.*
E		(set)	ser of edges	*ibid.*
n		$\mathbb{Z}; \gg 1$	# vertices $= \lvert V \rvert$	*ibid.*
k	(WS)	$\mathbb{Z}; \ll n$	# "local" connections, before rewiring	Watts/Strogatz, [4]
p	(WS)	$\mathbb{R}[0, 1]$	**pr.** replacing a local link with a long-distance	*ibid.*
γ		$\mathbb{R}(2, \infty)$	power-law parameter (*i.e.* $\Pr(\kappa(v) = x) = x^{\gamma}$)	
m_0	(BA*)	$\mathbb{Z}; \ll n$	# initial nodes	Barabási/Albert, [13]
m	(BA*)	$\mathbb{Z}; < m_0$	# edges added per time-step	*ibid.*
k, k_v		$\mathbb{Z}[0, n-1]$	degree of a vertex (v)	
$\Pi(k)$	(BA*)	$\mathbb{R}[0, 1]$	**pr.** attaching an edge to a vertex of degree k	*ibid.*
p	(BA*)	$\mathbb{R}[0, 1]$	per-timestep **pr.** adding links to existing vertices	Albert/Barabási, [14]
q	(BA*)	$\mathbb{R}[0, 1-p]$	per-timestep **pr.** re-wiring exiting links	*ibid.*
β	(GLP)	$\mathbb{R}[-\infty, m)$	preference-function offset (penalty)	Bu/Towsley, [20]
α	(BA+U)	$\mathbb{R}[0, 1]$	per-edge **pr.** adding uniformly (vs. preferentially)	Pennock *et al.*, [19]
w		(vertex)	'prototype' node for a new-growing one	Kumar *et al.*, [24]

Figure 2. Commonly used symbols, in order of appearance. Where noted, different authors may use the same variable with different meanings.

hibiting exponential tail behavior[6] in distribution of vertex degree. As [6] and [7] found, however, many real-world networks (*e.g.* the ones identified by Watts and Strogatz, the Internet topology, and others) have empirical *power-law*-distributed vertex degree distributions (with a parameter usually written γ), and so alternative models are necessary to generate families of artificial graphs faithfully resembling them. [7], in particular, demonstrates that such distributions in degree, cliquishness, and centrality arise from general, domain-nonspecific phenomena, namely:

- preferential connectivity
- incremental growth
- spatial distribution
- locality of connections,

which argues that we should expect to observe power-law distributions among graph statistics in any paradigm where such factors have significant influence.

However, Mitzenmacher, in [9], citing an earlier paper [10] by Downey, argued that many properties believed to be power-law, such as website traffic and filesize (as well as various phenomena throughout the rest of the natural sciences), may instead follow instead the related *lognormal* distribution, which can arise under very similar processes with only slightly different numeric parameters. While the asymptotic behaviors of the two distributions differ (the latter, crucially, exponential), Mitzenmacher goes to note that the former arises from the latter both in mixtures of lognormals and stopping-time-heterogenous lognormal processes, which may explain why aggregated phenomena typically exhibit power-law distributions.

To quote [9], though:

> From a more pragmatic point of view, it might be reasonable to use whichever distribution makes it easier to obtain results… The recent work [[11]] argues that for at least some network applications the difference in tails is not important[, but w]e believe that formalizing this idea is an important open question.

The field as a whole, however, seems to

[6] *Tail behavior* will be a recurring topic of inquiry in this survey; with respect to a particular positive quantity κ_v of a vertex v, it refers to the asymptotic behavior of $\Pr(\kappa_v = x)$ in the limit of large x, on models of suitably large $|V|$. In particular, a distinction is drawn between *exponential* behavior, which is $2^{-\Theta(x)}$, and *power-law* (or *scale-free*) behavior, $\Theta(x^{-\gamma})$ for some $\gamma > 2$.[7] It is of particular interest that the latter is exponentially more likely to produce vertices v with large κ_v.

[7] Requiring that the probabilities normalize to 1 (or more fundamentally, sum finitely) only imposes the condition $\gamma > 1$, but requiring that $\sum_{v \in V} \kappa_v < \infty$ requires $\gamma > 2$. Since we most often are interested in vertex degree as our κ, the requirement that our graph have finite $|E|$ imposes the stricter $\gamma > 2$ condition.

have largely focused on the more-mathematically-tractable power-law distribution models, and as such we defer the discussion of the lognormal distribution and its appropriateness in this field, referring the interested reader to [9] as an introduction to the topic.

D. Early Nongenerative Power-Law Models

Aiello, Chung, and Lu, in [12] provide a model (hereafter ACL) for producing (self-linked, multi-)graphs with arbitrary (finite-$\sum_v k_v$) degree distribution:

- Consider the set of (vertex, link number) pairs.
- Take a random matching on the set, and include one edge corresponding to each matched pair.

Though, as they note, their method is not equivalently distributed to uniform over all graphs with the desired degree-distribution, they apply a lemma from [5] to prove, similarly to the result of [36], that it approximates the latter in the sense that events that happen with high probability in one model happen with high probability in the other as well. And though the constructed self-linked multi-graph is not necessarily connected, we can achieve an approximation of the graph we wanted by taking the largest component and discarding self-links and duplicate links; [20] later demonstrated that this method produced probabilistically-close approximations.

Though their results further include a probabilistic analysis of the size of the connected components of such a graph (as their particular interest was in studying the Internet topology), which we will pass over here, the fact that their model is prescriptive, rather than generative, poses a difficulty for general probabilistic analysis. The expected-count version (which includes each edge (u, v) with probability $\frac{k_u \cdot k_v}{|E|-1}$) is for some purposes a little better, but, as we will see below, a *generative* model in which the degree distribution arises naturally from initial parameters is in many cases preferable.

E. BA Models: Sequential Preferential Attachment

While Aiello, Chung, and Lu's construction produced vertex-degrees in the proper distribution (effectively by constructive fiat), it had previously been noted by [7] that power-law distributions arise in real-world networks that grow over time from the influence of certain procedural pressures influencing where new connections attach. Several researchers asked if a procedural generative model which relied on these processes to drive its power-law growth might produce a more statistically faithful model of such networks, and Barabási and Albert, in [13], proposed a (m_0, m, T)-parametrized model for non-spatial systems which makes use of the first two phenomena mentioned in §§IV-C above to 'grow' a power-law graph from a small set of vertices by adding edges and vertices stochastically. Specifically, their construction grows according to the following stochastic process:

- Begin with m_0 isolated edges.
- At each time step $t \in [1, T]$, add an additional vertex connected to m already-present vertices, with the probability of it connecting to vertex i proportional to the degree of the latter:

$$\Pi(k) := k / \sum_{v \in V} k_v, \tag{1}$$

  where $Pi(k)$ is the probability that a vertex with degree k is chosen, and k_v is the degree of vertex v.

They prove that this process produces graphs with power-law degree distributions with parameter $\gamma = 3$, independent of m.[8] By comparing the results of their (vertex-growth and preferential-attachment) model with models that use only one or the other, they concluded that:

- Preferential attachment was required to induce a power-law degree-distribution.
- Vertex growth was required in order for the resulting model to be *stationary*—*i.e.* equidistributed regardless of t.

While the second finding is slightly contrived (it's easy to keep $t \ll \binom{n}{2}$, so that the distribution remains approximately t-insensitive), this Barabási-Albert (hereafter BA1) model would form the basis of the majority of models proposed during the early 2000s, largely because the process intuitively reflects our ideas about how networks grow, making

[8]This is actually problematic; [20] notes that *e.g.* the Internet topology exhibits the significantly smaller $\gamma = 2.18$.

	add-vertex	add-edges	rewire-edges	preference	mode	γ
BA1	1	—	—	$\propto k$	m	3
BA2	$1-p-q$	p	q	$\propto k$	$[0,m]$	$1+\frac{2m(1-q)+1-p-q}{m}$
GLP	$1-p$	p	—	$\propto k-\beta$	m	$\frac{2m-\beta(1-p)}{(1+p)m}$
BA+U	1	—	—	$\propto k+\frac{1-\alpha}{\alpha}2m$	$m(2-\alpha)$	$1+\alpha^{-1}$

Figure 3. Comparison of non-community-structured BA-variant models: BA1 (§§IV-E), BA2 (§§IV-E), GLP (§§IV-F), and BA+U (§§IV-G), in terms of the allowed operations, (proportional) preference in terms of degree, and certain distribution statistics.

it easier to invent and understand modifications to its construction rules.

A sequel paper by the same authors ([14]) generalizes the model (hereafter BA2) to five parameters (m_0, m, T, p, q) by introducing an additional two possibilities at each timestep:

- With probability p, add m edges between existing vertices, with one endpoint chosen uniformly and the other chosen preferentially.
- With probability q, 'rewire' m edges, chosen uniformly at random, by re-assigning one end preferentially.

(This means that a new node is added with its own m edges, as per BA1, only with probability $1-p-q$.)

Intuitively, this generalization makes sense, since real-world networks grow not only by the addition of new nodes, but by also the addition of links between old nodes (and, in some networks, the occasional removal of links). Analytically, these additional degrees of freedom give the model the freedom to fit γ other than 3, and vary edge density without necessarily increasing the minimum vertex-degree (which, in BA1, was given by m); as a proof-of-concept, the authors manage to fit the degree-distribution of the movie actors network extremely closely with $(m, p, q) = (1, 0.937, 0)$. (See Figure 4.)

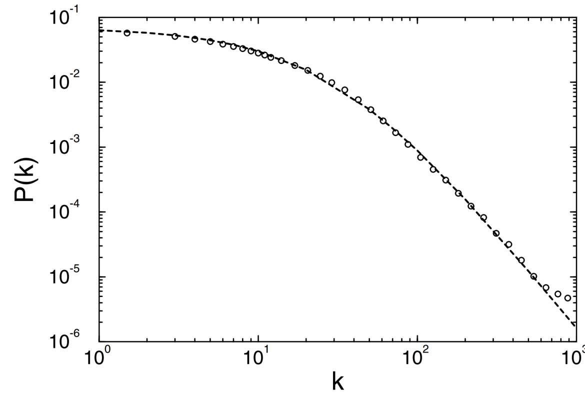

Figure 4. Figure reproduced from [14]: "The connectivity distribution of movie actors (circles) based on the Internet Movie Database containing 212 250 actors and 3 045 787 links. . . The data were logarithmically binned. The dashed line corresponds to the two parameter fit . . . with $p = 0.937$ and $m = 1$."

We defer the discussion of expected-count versions of sequential preferential attachment models (or "general BA models") to §V.

F. Digression: Internet Topology Generators

One distinguishing feature of networks that we have so far neglected to discuss is the possibility of spatial locality (in some low-dimensional embedding—$d = 2$ or $d = 3$ for physical cases). Importantly, a special subclass of graph generator, such as appears in [16], [17], and [18], aims to generate simulated sample topologies of the Internet at various agent-scales, usually for the purpose of network system simulation, and makes heavy use of the $d = 2$ spatial embedding exhibited. While they borrow ideas from the graph models previously discussed in order to induce power-law structure (in various statistics) on the generated graphs, their power-law phenomena are conflated with spatial-embedding one, which makes them (1) more accurate at generating sample Internet topologies and (2) less appropriate for generating general non-planar power-law graphs.

As a demonstration of the importance of considering spatial embedding in topology simulation, Bu and Towsley compare several graph-generation models in [20], namely the previously-discussed ACL, BA1, BA2, and the Inet-2.0 generator described in [17]—as well as their own model, which they dub *Generalized Linear Preference* (hereafter

GLP). It is a generalization of BA2 (fixing $q = 0$), in which the preferential attachment probability $\Pi(k)$ depends on degree *affine*-linearly:

$$\Pi(k) := (k - \beta) / \sum_{v \in V} (k_v - \beta). \tag{2}$$

Note that, though the authors are clearly interested in Internet-topology generation, their proposed model is spatially-agnostic. Nevertheless, their results, described in much more detail in [20], indicate that they perform comparably to Inet-2.0 on many metrics. (However, Inet-3.0, as described in [18], further improved the Inet model on many metrics.)

In any case, graph generators which make use of spatial embedding are largely outside the scope of this survey, due to their specialized nature; the reader interested in their development is directed to [18] and [20] and their bibliographies for a characteristic introduction.

G. Partially Uniform Preferential Attachment

A recurring theme in papers proposing a generalization of BA is that they:

- Note that the BA model is over-simplified, and fails to reflect [a certain property] empirically observed in real networks.
- Note that a general $\Pi(k) = k^\alpha / \sum_{v \in V} k_v^\alpha$ model of attachment would be preferable…
- …but that such a model proves too difficult to analyze, so they propose an alternate, simpler model instead.

Pennock *et al.*, in [19] follow this trend, noting that BA models are unable to describe the positive-mode, “winners don’t take all” distribution observed in certain subnetworks of the WWW graph, such as science faculty personal webpages, and university homepages.

By way of remedy, they propose another generalization of BA1 (hereafter BA+U) that is almost entirely complementary to GLP: every preferential endpoint-attachment is performed as per usual only with probability α; with probability $1 - \alpha$, the endpoint in question is instead drawn uniformly. This can either be understood as a mixture of preferential-attachment and random-attachment models, or as an affine-linear preference model with *positive* (rather than GLP’s negative) offset:

$$\begin{aligned}
\Pi(k) &\propto \alpha \frac{k}{\sum_{v \in V} k_v} + (1 - \alpha) \frac{1}{|V|} \\
&= \frac{\alpha k}{2\,|E|} + \frac{1 - \alpha}{|V|} \\
&\approx \frac{\alpha k}{2\,|V|\,m} + \frac{1 - \alpha}{|V|} \\
&\propto k + \frac{(1 - \alpha) m}{\alpha},
\end{aligned} \tag{3}$$

though the authors neglect to mention this fact.

Noting that α parametrizes both the mode ($= 2m(1-\alpha)$) and γ ($= 1+\alpha^{-1}$), Pennock *et al.* fit the (m, α) model to several subnetworks of the WWW network, in all five presented cases able to fit both power-law exponent and modal distribution in the low-connectivity region with a single α. While not all subnetworks exhibit empirical α significantly less than 1 (and the Web as a whole does not)[9], their experiments demonstrate (again) that in certain subnetworks, it may be useful to model attachment preference more generally than as strictly linear. Taken together with Bu and Towsley’s analysis in [20], this indicates that a general offset may be a useful extension to BA models across a variety of contexts.[10]

H. Community Structure in BA Variants

Largely orthogonal to progress in more accurately fitting distribution tail-behavior was an interest in the community structure of such graphs, particularly focused on the Web graph. In [21], Kumar *et al.* noted that the Web graph exhibited *self-similar* clustering structure, often described as “communities within communities”, and several BA-inspired models were introduced to attempt

[9] [19] uses inlinks to five different categories of webpages for their analysis: company homepages, newspaper homepages, all pages, university homepages, and science faculty personal pages. The first three exhibit high-α behavior, while the latter two fit to $\alpha = 0.612, 0.602$, respectively

[10] Mitzenmacher’s later survey on lognormal distributions, [9], will formalize the intuition that power-law distributions emerge as the marginal mixture of lognormal distributions, suggesting that the subnetworks may be, in fact, lognormal, and that the largely power-law networks may simply be too aggregated to observe their subnetworkwise-lognormal structure.

to replicate this structure procedurally. In [22], Kleinberg *et al.* propose a squential-attachment model with implicit preference through a 'copying' mechanism whereby new nodes preferentially adopt connections similar to an existing node in the graph. Each new node:

- Chooses a 'prototype' w uniformly among existing nodes, link to w.
- For every link v_i in the prototype, include it in the new node with probability $1 - \alpha$, and include a uniformly-chosen link instead with probability α.

The linear-in-degree preference of the BA model, then, is implicit, as a high-degree node is referenced and available for copying in many possible places.

The same authors generalized the model in [24] to an epoch-based 'exponential growth' process, in which only the oldest fraction of nodes is 'visible', and therefore, available as prototypes. By varying the strength of the copying preference (vs. a α-probability to choose a uniform vertex), the power-law parameter γ can be fit, by the same argument as in Bu and Towsley's BA+U model from [20].

A further generalization, by Leskovec, Kleinberg, and Faloutsos in [25], includes an additional probability that a growing node not only links to a neighbor of its prototype, but also adopts said neighbor as an *additional* prototype, recursing. Further, the number of links adopted varies stochastically, so the full process is, for every new node v,

- Choose a prototype (in [25], an 'ambassador') w uniformly; link to w.
- Select a geometrically-distributed number of links to follow $X \sim \text{Geom}\left(\frac{1}{1-p}\right)$; add each as a prototype and recurse. (A further variant chases in- and out-links separately.) Ignore links once visited once.

This model, the authors show, not only produces community structure phenomena, but also two empirical observations seen in real networks but not earlier BA models, *densification over time*, and *shrinking diameter over time*, and they show that this *forest fire model* can be tuned to match the densification and diameter-shrinking properties of (1) citations on the arXiv, (2) citations in registered US patents, (3), the Internet topology, and (4) the arXiv affiliation graph (a bipartite expansion of the co-authorship network). A competing model introduced in the same paper, the *community-guided attachment model* differs significantly, attempting to fit nodes into a tree structure which defines distances, and thereby, connection probabilities, but is found to be less successful at fitting the community structure of real networks.

I. R-MAT and Kronecker Exponentiation

A persistent problem with sequential-attachment models in general, however, is that the sequential nature of their generative processes pose persistent difficulties for probabilistic analysis, especially in the face of a growing pile of suggested ad-hoc modifications. Further, there was slow-developing but serious concern that community-structured attachment models, while a natural fit for BA processes, failed to capture the scale-free[11] meta-community structure observed in several real networks, and worse, the procedural clunkiness of sequential attachment made rigorous analysis of the community-structure problem extremely difficult.

In 2004, Chakrabarti, Zhan, and Faloutsos proposed instead a new model that broke from the sequential-attachment paradigm, R-MAT[12], laid out in [27]. In something of a return to the spirit of the ER and WS models (or, as an alternative interpretation, a timeless generalization of the unsuccessful community-guided attachment model discussed in [25]), it assigned a time-independent probability to each edge being added at an arbitrary timestep, and constructed a graph by drawing a number of edges from the distribution so produced. In their formulation, each edge is chosen by the following recursive process on a $2^k \times 2^k$ adjacency matrix:

- Choose a quadrant from $\{(0,0),(0,1),(1,0),(1,1)\}$ with probabilities $Q_{0,0}, Q_{0,1}, Q_{1,0}, Q_{1,1}$, respectively.
- Descend into that quadrant; recurse.
- When a single cell is reached, mark that cell.

As usual, we can either allow self-links and multiple edges, or ignore and re-draw them without affecting the gross probabilistic properties of the

[11] *i.e.* observed "community" structure at the zoomed-out scale where communities-of-nodes were considered as single nodes, as well as at the level where communities-of-communities were single nodes, *etc.*

[12] abbrev. **R**ecursive **MAT**rix

graph. (More precisely, discarding edges either during or after the fact is equivalent to choosing from only allowed edges with probabilities proportional to the originals, up to an exact-/expected-count approximation.)

This process, however, is clearly equivalent to simply taking draws from a single flat matrix of probabilities $\mathcal{Q}$ defined by

$$\begin{aligned}\mathcal{Q}_{i,j} &= \prod_{\ell=1}^{k} \sum_{\alpha,\beta=0,0}^{1,1} Q_{\alpha,\beta} \cdot 1_{\langle i_{(\ell)}=\alpha \rangle} \cdot 1_{\langle j_{(\ell)}=\beta \rangle} \\ &= \prod_{\alpha,\beta=0,0}^{1,1} Q_{\alpha,\beta}^{(i,j)_{\alpha,\beta}} \\ &= Q_{0,0}^{(i,j)_{0,0}} Q_{0,1}^{(i,j)_{0,1}} Q_{1,0}^{(i,j)_{1,0}} Q_{1,1}^{(i,j)_{1,1}}, \end{aligned} \tag{4}$$

where $i_{(\ell)}$ is the ℓth-most-significant bit of $i < 2^k$ and

$$(i,j)_{\alpha,\beta} := \# \left\{ \ell \in [1,k] \;\middle|\; (i_{(\ell)} = \alpha) \text{ and } (j_{(\ell)} = \beta) \right\}, \tag{5}$$

as noted in [28]. As discussed later in §V, this flattening of attachment structure greatly simplifies the probabilistic analysis of such a system (for example, we can once again consider the expected-count model by taking $|E|\,\mathcal{Q}$ as the probabilities of each edge occurring), and the ability to capture fractal community structure in a non-sequential model is a fundamental contribution to the field.

In later papers [29], [30], the authors and their collaborators generalized the R-MAT model's 2×2 recursive-rule matrix to an arbitrary $n \times n$, repeatedly applied to itself per *Kronecker multiplication*:

$$A \otimes B := \begin{pmatrix} a_{1,1}B & a_{1,2}B & \dots & a_{1,n}B \\ a_{2,1}B & a_{2,2}B & \dots & a_{2,n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1}B & a_{n,2}B & \dots & a_{n,n}B \end{pmatrix} \tag{6}$$

$$A^{\otimes k} := \underbrace{A \otimes A \otimes \ldots \otimes A}_{k \text{ times}}. \tag{7}$$

(With this notation, we can note simply that $\mathcal{Q} := Q^{\otimes k}$.)

This generalization allows them to better match the density-over-time and diameter-over-time behaviors of the arXiv, patents, and Internet-topology datasets introduced in [25]. In [30], Leskovec and Faloutsos introduce an efficient ($O(n)$) algorithm for fitting Kronecker graph iniators (*i.e.* exponent base matrices) to observed graphs, fitting many network parameters and parameter-over-time behaviors closely. In [31], Leskovec further explores the manner in which Kronecker products are a natural fit for modeling self-similar community structures—namely, because community-within-community structure is captured by the recursive nature of the Kronecker product on a 'community prototype' connectivity matrix.

J. Statistical Methods and Triangle Density

Kronecker models represent the state of the art in power-law graph generation, and the availability for approximative algorithms to fit them to empirical graphs represents an enormous achievement of the graphs community. Since the Kronecker initiator (*i.e.* the base matrix) encodes information about clustering, graph width, degree-distribution skew, and tree-decomposition depth ([30]), we can use Kronecker-matrix fitting to approximate many relevant features of power-law graphs, in much the same way as the ER and Gilbert models capture combinatorial information about uniform graphs in their parameter models. Further, well-fitting matrix generators that assign high posterior likelihood to provided graph examples allow us to make statistical claims about those graphs, by appealing to expectation properties of the Kronecker-generated graphs which we can study analytically. While the theory of Kronecker graphs is still new enough that there are few results in this area, we expect that it represents a promising topic of future research.

Even more fundamentally, the study of fundamental graph properties has led Gupta, Roughgarden, and Seshandhri to study triangle-density of graphs:

$$\tau(G) = \frac{3\#\{\text{triangles in } G_k\}}{\#\{\text{wedges in } G_k\}} \tag{8}$$

as the natural expansion of the two model parameters proposed by Erdös and Rényi—vertices and edges. Making use of new probabilistic methods for approximating triangle-counts ([33]), they demonstrate several topological properties which are sensitive to the density of triangles. As with Kronecker initiators, this is exciting as a potential topic for future research, as triangles are a natural candidate for a fundamental graph parameter allowing us to describe clustering behavior, and further link clustering to other combinatorial properties, in the style of Erdös.

V. A Comparison of Matrices

To allow a more systematic comparison of these procedural generative methods, we can describe each as a series of (potentially dependent) random variables $E_1, \ldots, E_{|E|}$. Further, we can treat each as a vector of probabilities (*i.e.* of choosing each edge), writing the matrices

$$(\mathcal{E}_t)_{i,j} := \Pr\left(E_t = (i,j) \mid E_1, \ldots, E_{t-1}\right) \tag{9}$$

for $t \in [1, |E|]$. For computational convenience, we will normalize these so that probabilities sum to 1, and will (as in many of the models above) simply discard self-links or duplicate links when they occur. We'll refer to the distribution allowing self-links and duplicate links (which are later discarded) as the *with-replacement* distribution.[13]

In §§V-A–§§V-D below, we will determine the edge-probability matrices of the models discussed above (passing, when appropriate, to the with-replacement model as an approximation when appropriate); in §§V-E, we will provide an analytic interpretation of these matrices, and describe the distributions over vertex degree that they induce in both the exact-count and expected-count case, as well as an elegant exponential process model.

A. Erdös–Rényi, Watts–Strogatz

As a first example, the edge-probability matrices for the ER model are simple:

$$\mathcal{E}_t^{\mathrm{ER}} = \mathcal{E}^{\mathrm{ER}} = \frac{1}{|V|^2} \cdot J_{|V|}, \tag{10}$$

where J_k is the $k \times k$ matrix composed of all ones.

The WS model, then, has matrices

$$\mathcal{E}_t^{\mathrm{WS}} = \mathcal{E}^{\mathrm{WS}} = p\mathcal{E}^{\mathrm{ER}} + (1-p)\mathcal{E}^{\mathrm{Ring}} \tag{11}$$

where

$$\left(\mathcal{E}^{\mathrm{Ring}}\right)_{i,j} := \begin{cases} 1/nk & \text{if } |(i-j) \pmod n| \leq k/2 \\ 0 & \text{else.} \end{cases} \tag{12}$$

These three are notable in that the with-replacement distribution is time-invariant. It follows that, in the Poisson–balls-in-bins approximation of [36][14], we can treat the entire process as independently including each edge with a probability dependent on only the indices of its endpoints, a convenience of the sort that made the original ER model so useful, while capturing slightly more structure than simple Erdös graphs.

B. Aiello–Chung–Lu

The edge-probability matrices of the ACL model are, unfortunately, time-dependent, complicating our analysis slightly. They are, however, quite simple when considered individually:

- $\mathcal{E}_t^{\mathrm{ACL}}$ is rectangular, *i.e.*

$$\forall t, \exists \{\epsilon_{t,i}\}_{i=1}^{|V|} : \forall i,j, \left(\mathcal{E}_t^{\mathrm{ACL}}\right)_{i,j} = \epsilon_{t,i} \cdot \epsilon_{t,j}. \tag{13}$$

- Each ϵ_i is proportional to the current degree of vertex i, so that

$$\begin{aligned} \left(\mathcal{E}_t^{\mathrm{ACL}}\right)_{i,j} &= \epsilon_{t,i} \cdot \epsilon_{t,j} \\ &= \frac{\ell_i^{(t)}}{\sum_{i'=1}^{|V|} \ell_{i'}^{(t)}} \cdot \frac{\ell_j^{(t)}}{\sum_{i'=1}^{|V|} \ell_{i'}^{(t)}} \\ &= \left(\sum_{i'=1}^{|V|} \ell_{i'}^{(t)}\right)^{-2} \cdot \ell_i^{(t)} \ell_j^{(t)} \\ &\propto \ell_i^{(t)} \cdot \ell_j^{(t)}, \end{aligned} \tag{14}$$

  where $\ell_i^{(t)} := k_i - k_i^{(t)}$, the target degree of vertex i minus the number of edges connected to it so far.

We see these by observing that the endpoints of each edge are chosen independently respectively (in the random-matching step), taking each vertex with probability proportional to the number of attachment sites still available. The time-dependence is introduced by the change in the $\ell_{t,i}$ as edge-connections get 'used up'.

However, if we are willing to pass to the with-replacement model as an approximation, then the

[13] If we wished to enforce an exact edge-count, we could redraw each immediately until we reached an acceptable edge, but this complicates our analysis somewhat, so we will often disregard it.

[14] As referenced, the source actually only proves this claim for balls-in-bins processes with equiprobable bins, though the additivity of Poisson processes gives an easy extension of the proof to the general case here.

analysis simplifies significantly, and the matrices become time-independent:

$$\left(\mathcal{E}_t^{\text{ACL}}\right)_{i,j} = \left(\mathcal{E}^{\text{ACL}}\right)_{i,j} = k_i \cdot k_j. \tag{15}$$

C. Barabási–Albert and Friends

Even the $m = 1$ with-replacement matrices of the BA models are truly time-dependent, complicating our analysis. The simplest, BA1, is given by:

$$\left(\mathcal{E}_t^{\text{BA1}}\right)_{i,j} = \delta(i, m_0 + t) \cdot \frac{mk_j}{\sum_{v \in V} k_v}. \tag{16}$$

Similarly, BA+U has matrices

$$\mathcal{E}_t^{\text{BA+U}} = \alpha \mathcal{E}_t^{\text{BA1}} + (1 - \alpha)\mathcal{E}_t^{\text{U1}}, \tag{17}$$

where $\mathcal{E}_t^{\text{U1}}$ is the matrix corresponding to the process "add a new vertex attached to m existing vertices uniformly at random":

$$\left(\mathcal{E}_t^{\text{U1}}\right)_{i,j} = \delta(i, m_0 + t) \cdot \frac{m \cdot 1_{\langle j < m_0 + t \rangle}}{m_0 + t - 1}, \tag{18}$$

and GLP is identical, up to sign and parameter:

$$\mathcal{E}_t^{\text{GLP}} = (1 + \beta)\mathcal{E}_t^{\text{BA1}} - \beta \mathcal{E}_t^{\text{U1}} \tag{19}$$

BA2 is further complicated by the inclusion of an edge-removal operation, and the various community-attachment models are fully dependent in edge-attachments for any nontrivial setting of the community-influence parameters. To conclude, none of the BA-variant models decompose nicely into time-independent with-replacement matrices, and as such, we exclude them from this section's probabilistic analysis—for an alternate probabilistic result, however, see Appendix A.

D. R-MAT

As before, R-MAT provides a return to simplicity—recall from §§IV-I that its edge-probability matrix $\mathcal{E}^{\text{RMAT}}$ is given by

$$\mathcal{E}^{\text{RMAT}} = \mathcal{Q} = Q^{\otimes k}, \tag{20}$$

and that the matrix for a general Kronecker graph is similarly given for Q a general $n \times n$ matrix.

E. Interpretation of Edge-Probability Matrices

What use are these matrices? Well, they provide a single, simple point of comparison for the nonsequential/time-independent models, which are defined completely by their $\mathcal{E}$. The dependence relations in the sequence $\mathcal{E}_1, \mathcal{E}_2, \ldots$ shed light on the ways in which the sequential models exhibit feedback from epoch to epoch (though we've glossed over the more complicated ones here, the reader may find them interesting to determine as an exercise). But perhaps most usefully, they give rise to an interesting related model:

Let every matrix entry $A_{i,j}$ represent a bin which receives exactly one ball at a time exponentially distributed with parameter $1/A_{i,j}$. Then:

- The expected-count model is equidistributed to the distribution of bins which are occupied by time $t = 1$,
- The exact-count $|E| = m$ model is equidistributed to the distribution of bins which are occupied by the time the mth ball arrives,

and the with-replacement models are likewise equidistributed to the outcomes of a process by which bins receive balls repeatedly, with successive arrival times independent and exponentially distributed. Equivalently, the count of each bin is distributed as a Poisson process (either multi-linked, or clamped in value to $[0, 1]$).

This is a useful model. We present two statistical analyses it affords us: first, we can apply the analysis of balls-in-bins processes on matrices presented in [28], most interestingly:

Theorem V.1. *Given a set of m urns with probabilities $\vec{q} := \{q_1, \ldots, q_m\}$ with $\sum_{i=1}^m q_i = 1$, let X be the random variable corresponding to the number of empty urns after tossing r balls inot these urns. Then if $r, m \to \infty$ with $r/m \to C_1$ where $0 < C_1 < \infty$ and $m \cdot q_i \leq C_2 < \infty$ for each i, then the probability distribution of X is asymptotically normal.*

Since the conditions are satisfied for each of the matrices produce above, this implies that for any of them, the degree in any particular vertex is normal in the limit. Note that for the Kronecker-graph models, this means that any individual vertex has normal (*i.e.* thin-tailed) degree distribution—and that the overall power-law distribution over degree arises from the power-law-distributed *means*, rather than fat tails in individual vertex-degrees. If vertex indices are marginally identical, on the other hand, the identicality of their distributions imply that the overall degree-distribution must be normal in likelihood, as a (not particularly surprising) corollary:

Corollary V.1. *Any graph-generation model with an edge-probability matrix with rows homogeneous up to permutation of entries will exhibit asymptotically exponential degree distribution.*

It follows that any matrix model for generating power-law graphs must have rows heterogeneous up to permutation of entries.

A second application of 'nice' matrices is the following identities on the non-deduplicated multigraphs G_k they produce after k edge-draws:

$$\sum \mathcal{E}^2 = \mathbb{E}\# \left\{\text{wedges in } G_2\right\} \tag{21}$$

$$\binom{k}{2} \sum \mathcal{E}^2 = \mathbb{E}\# \left\{\text{wedges in } G_k\right\} \tag{22}$$

$$\frac{1}{3} \left\langle \mathcal{E}, \mathcal{E}^2 \right\rangle = \mathbb{E}\# \left\{\text{triangles in } G_3\right\} \tag{23}$$

$$\frac{1}{3} \binom{k}{3} \left\langle \mathcal{E}, \mathcal{E}^2 \right\rangle = \mathbb{E}\# \left\{\text{triangles in } G_k\right\}, \tag{24}$$

with sums taken over the elements of the matrix, and $\langle \bullet, \bullet \rangle$ representing the elementwise sumproduct over elements. For graph processes with efficiently-computable matrices (as are all of the ones discussed), this gives a computationally-tractable formula for the expected triangle density τ:

$$\begin{aligned}
\mathbb{E}\tau(G_k) &= \frac{3\mathbb{E}\# \left\{\text{triangles in } G_k\right\}}{\mathbb{E}\# \left\{\text{wedges in } G_k\right\}} \\
&= \frac{\binom{k}{3} \left\langle \mathcal{E}, \mathcal{E}^2 \right\rangle}{\binom{k}{2} \sum \mathcal{E}^2} \qquad (25) \\
&= \frac{(k-2) \left\langle \mathcal{E}, \mathcal{E}^2 \right\rangle}{3 \sum \mathcal{E}^2}. \qquad (26)
\end{aligned}$$

VI. Conclusion

The story of the study of power-law graphs and scale-free networks is, in the first, a story of the discovery and exploration of unexpected self-organizing structure in networks all around us; and in the second, a story of trial, error, insight, and ongoing effort to develop a rigorous and elegant theory of network structure and generative process that is both general enough to explain the diversity of properties exhibited in natural networks, and restrictive enough to guarantee that such structures do appear from local, stochastic processes. The field has progressed from haphazard and prescriptive models, through a period of seeming endless variations on a single sequential-procedural theme, before embracing a once-more fundamentally combinatorial approach that shows great promise, and hints at deeper structure yet.

The edges of the field are sites of active research, converging to a coherent, fundamental theory of the structure of organized networks.

Appendix A
Long-Run Statistics of Nonlinear Preferential Attachment

The BA family of sequential preferential attachment consists solely of (affine-)linear attachment preferences, but since there is generally-expressed interest in general (*i.e.* nonlinear) preference models, we provide a little analysis of the case where preferences are proportional to k^ρ, for $\rho > 1$.

Definition A.1. *A n-bin ρ-feedback process is an n-bin balls-in-bins stochastic process in which, at each time-step, a ball is added to bin i (which already has x_i balls) with probability proportional to x_i^ρ.*

In [37], Mitzenmacher and Upfal present the following theorem

Theorem A.1. *Under any starting conditions for a 2-bin ρ-feedback process, if $\rho > 1$, then with probability 1 there exists a number c such that one bin gets no more than c balls.*[15]

We extend the theorem to n bins:

Theorem A.2. *Under any starting conditions for an n-bin ρ-feedback process, if $\rho > 1$, then with probability 1 there exists numbers $c_1, \ldots, c_n$ such only one bin i receives more than c_i balls.*

[15] To quote further from [37]: "Note the careful wording of the theorem. We are not saying that there is some fixed c such that one bin gets no more than c balls. Instead, we are saying that, with probability 1, at some point one bin stops receiving balls.

Proof. We follow Mitzenmacher and Upfal's proof; consider the process where all n bins receive balls independently, at continuous times, letting $T_{i,z}$ be the time between when bin i receives its zth ball and when it receives its $(z+1)$st. In particular, consider the case where $T_{i,z} \sim \text{Expo}(z^\rho)$; then the distribution of balls after k have arrived is equivalently distributed to the distribution of a n-bin ρ-feedback process, as argued in [37]. Then, defining the *saturation time* F_i of bin i by

$$F_i := \sum_{z=1}^{\infty} T_z, \tag{27}$$

we find that, with probability 1, the saturation time is bounded when $\rho > 1$:

$$\mathbb{E}F_i = \mathbb{E}\sum_{z=1}^{\infty} T_{i,z} = \sum_{z=1}^{\infty} \mathbb{E}T_{i,z} = \sum_{z=1}^{\infty} \frac{1}{z^\rho} < \infty \tag{28}$$

$$\implies \Pr(F_i \not< \infty) = 0. \tag{29}$$

With probability 1, the F_i are distinct and there is some unique j such that $F_j = \min\{F_i \mid i \in [1,n]\}$. Writing $E_{i,\ell} := \sum_{z=1}^{\ell} T_{i,z}$, it follows that $\forall i \neq j, \exists k_i : E_{i,c_i} < F_j < E_{i,c_i+1}$, and so, for all sufficiently large m, we have $E_{i,c_i} < E_{j,m} < E_{i,c_i+1}$, which is to say, that after each bin i has received c_i balls, all the rest go to bin j. □

We generalize further:

Definition A.2. *An (n, m, ρ)-feedback process is an n-bin balls-in-bins stochastic process in which, at each time-step, m balls are added to distinct bins i chosen without replacement with probability proportional to x_i^ρ.*

Theorem A.3. *Under any starting conditions for an (n, m, ρ)-feedback process, if $\rho > 1$, then with probability 1 there exist numbers $c_1, \ldots, c_n$ such only m bins i receive more than c_i balls.*

Proof. We again appeal to the model where each bin independently receives balls with successive exponentially-distributed arrival times, except instead of giving a ball to the single bin with the bin with the next successive arrival time at each timestep, we give a ball to the m distinct bins with the next-least arrival times. As before, we note that some m bins reach their saturation points before the other $n - m$ reach some finite ball-counts c_i, and thereafter receive all of the balls. □

The application to nonlinear BA models is as follows: We model BA$^\rho$ as an (n, m, ρ)-feedback process, in which the number of balls received corresponds to the number of links added. While it is no longer necessarily true that m bins receive all of the balls after a certain point, due to the arrival of new bins, it is nevertheless the case that all but m bins will stop receiving balls at some point, by extension of the proof.

While these results do not apply to linear BA models (recall that we required $\rho > 1$), it does have implications for the study of nonlinear BA models, as well as network-growth systems that are believed to follow such attachment-models: With high probability, everyone but a finite set of winners loses, in the long run.

Acknowledgements

The author would like to acknowledge the invaluable assistance of Daniel Margo, student at Harvard, who provided numerous explanations, clarifications, and elucidations along the way; as well as Michael Mitzenmacher, of Harvard, whose introductory course in randomized algorithms was foundational to the exploration of ideas here, and whose co-authored text provided an excellent introduction to the theory of random graphs.

References

[1] P. Erdös, A. Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.* **5**, 17–60, 1960.

[2] E. Gilbert. Random Graphs. *Ann. Math. Stat.* **30**, 1141–1144, 1959.

[3] S. Milgram. The small-world problem. *Psych. Today* **2** 60–67, 1967.

[4] D. Watts, S. Strogatz. Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442, 1998.

[5] M. Molloy, B. Reed. The Size of the Giant Component of a Random Graph with a Given Degree Sequence. *Comb. Prob. Comput.* **7**, 295–305, 1998.

[6] M. Faloutsos, P. Faloutsos, C. Faloutsos. On Power-Law Relationships of the Internet Topology. *ACM SIGCOMM* **29**, 251–262, 1999.

[7] A. Medina, I. Matta, J. Byers. On the Origin of Power Laws in Internet Topologies. *ACM SIGCOMM* **30**, 18–28, 2000.

[8] M. Newman. The structure and function of complex networks. *SIAM Review* **45**, 167–256, 2003.

[9] M. Mitzenmacher. A Brief History of Generative Models for Power Law and Lognormal Distributions. *I.net Math.* **1**, 226–251, 2004.

[10] A. Downey. The Structural Causes of File Size Distributions. *Proc. MASCOTS '01*, 2001.

[11] W. Gong, Y. Liu, V. Misra, D. Towsley. On the Tails of Web File-size Distributions. *Proc. 39th Ann. Allerton Conf. Comm. Control Comp.* 192–201, 2001.

[12] W. Aiello, F. Chung, L. Lu. A Random Graph Model for Massive Graphs. *Proc. STOC '00*, 171–180, 2000.

[13] A. Barabási, R. Albert. Emergence of Scaling in Random Networks. *Science* **286**, 509–512, 1999.

[14] R. Albert, A. Barabási. Topology of Evolving Networks: Local Events and Universality. *Phys. Rev. Lett.* **85**, 5234–5237, 2000.

[15] R. Albert, A. Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97, 2002.

[16] A. Medina, I. Matta, J. Byers. BRITE: A Flexible Generator of Internet Topologies. *Bost. U. Tech. Rep.* BU-CS-TR-2000-005, 2000.

[17] C. Jin, Q. Chen, S. Jamin. Inet: Internet Topology Generator. *U. Mich. Tech. Rep.* CSE-TR-433-00, 2000.

[18] J. Winick, S. Jamin. Inet-3.0: Internet Topology Generator. *U. Mich. Tech. Rep.* CSE-TR-456-02, 2002.

[19] D. Pennock, G. Flake, S. Lawrence, E. Glover, C. Giles. Winners don't take all: Characterizing the competition for links on the web. *PNAS* **99**, 5207–5211, 2002.

[20] T. Bu, D. Towsley. On Distinguishing between Internet Power Law Topology Generators. *INFOCOM* **2**, 638–647, 2002.

[21] R. Kumar, P. Raghavan, S. Rajagopalan, A. Tomkins. Trawling the web for emerging cyber-communities. *Proc. 8th WWW Conf.* 403–416, 1999.

[22] J. Kleinberg, R. Kumar, P. Raghavan, S. Rajagopalan, A. Tomkins. The web as a graph: measurements, models, and methods. *Proc. Intl. Conf. Comb. Comp. '99* 1–18, 1999.

[23] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, J. Wiener. Graph structure on the web: experiments and models. *Proc. 9th WWW Conf.* 309–320, 2000.

[24] R. Kumar, R. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, E. Upfal. Stochastic models for the web graph. *Proc. FOCS '00* 57, 2000.

[25] J. Leskovec, J. Kleinberg, C. Faloutsos. Graphs over Time: Densification Laws, Shrinking Diameters, and Possible Explanations. *Proc. KDD '05* 177–187, 2005.

[26] D. Chakrabarti, C. Faloutsos. Graph Mining: Laws, Generators, and Algorithms. *ACM Comp. Surv.* **38**, 69–123, 2006.

[27] D. Chakrabarti, Y. Zhan, C. Faloutsos. R-MAT: A Recursive Model for Graph Mining. *Proc. SDM '04* 2004.

[28] C. Groër, B. Sullivan, S. Poole. A Mathematical Analysis of the R-MAT Random Graph Generator. *Networks* **58**, 159–170, 2011.

[29] J. Leskovec, D. Chakrabarti, J. Kleinberg, C. Faloutsos. Realistic, mathematically tractable graph generation and evolution, using Kronecker multiplication. *Proc. PKDD '05* 133–145, 2005.

[30] J. Leskovec, C. Faloutsos. Scalable modeling of real graphs using Kronecker multiplication. *Proc. ICML '07* 497–504, 2007.

[31] J. Leskovec. Networks, Communities, and Kronecker Products. *Proc. CNIKM '09*, 2009.

[32] D. Bader, J. Feo, J. Gilbert, J. Kepner, D. Koester, E. Loh, K. Madduri, B. Mann, T. Meuse. HPCS Scalable Synthetic Compact Applications #2 Graph Analysis. *SSCA Tech. Rep.*, http://www.graphanalysis.org/, 2006. ret. 1 May 2015.

[33] C. Seshadhri, A. Pinar, T. G. Kolda. Fast triangle counting through wedge sampling. *Proc. SDM*, 2013.

[34] R. Gupta, T. Roughgarden, C. Seshadhri. Decompositions of Triangle-Dense Graphs. *Proc. ITCS '14*, 471–482, 2014.

[35] M. Mitzenmacher, E. Upfal. *Probability and Computing: Randomized Algorthms and Probabilistic Analysis*, Cam. U. Press 2005.

[36] *Ibid.* §5.6, pp.112–119.

[37] *Ibid.* §§8.3.2, pp.199–201.

FURTHER REFERENCE

[1] R. Albert, H. Jeong, A. Barabási. Diameter of the World Wide Web. *Nature* **401**, 130–131, 1999.

[2] R. Kumar, P. Raghavan, S. Rajagopalan, A. Tomkins. Extracting large-scale knowledge bases from the web. *Proc. VLDB '99* 639–650, 1999.

[3] J. Kleinberg. The small-world phenomenon: An algorithmic perspective. *Corn. Tech. Rep.* TR-99-1776, 1999.

[4] R. Albert, H. Jeong, A. Barabási. Error and attack tolerance of complex networks. *Nature* **406**, 378–382, 2000.

[5] M. Newman, S. Strogatz, D. Watts. Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E* **64**, 1–17, 2001.

[6] W. Aiello, F. Chung. Random Evolution in Massive Graphs. *Proc. FOCS '01*, 510–519, 2001.

[7] S. Tauro, C. Palmer, G. Siganos, M. Faloutsos. A simple conceptual model for the Internet topology. *Glob. I.net*, IEEE Comp. Soc. Press, 2001.

[8] E. Drinea, M. Enachescu, M. Mitzenmacher. Variations on Random Graph Models for the Web. *Harv. Tech. Rep.* TR-06-01, 2001.

[9] P. Krapivsky, S. Redner. Organization of Growing Random Networks. *arXiv:* cond-mat/0011094. 2001.

[10] X. Zhu, J. Yu, J. Doyle. Heavy Tails, Generalized Coding, and Optimal Web Layout. *Proc. INFOCOM '01* 1617–1626, 2001.

[11] C. Cooper, A. Frieze. On a General Model of Undirected Web Graphs. *Proc. 9th Ann. Eur. Sym. Alg.* 500–511, 2001.

[12] C. Cooper, A. Frieze. A general model of web graphs. *Rand. Struct. Alg.* **22**, 311–335, 2003.

[13] M. Newman. Power laws, pareto distributions and Zipf's law. *Contemp. Phys.* **46**, 323–351, 2005.

PREVIOUS SURVEYS

[1] R. Albert, A. Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97, 2002.

[2] M. Newman. The structure and function of complex networks. *SIAM Review* **45**, 167–256, 2003.

[3] M. Mitzenmacher. A Brief History of Generative Models for Power Law and Lognormal Distributions. *I.net Math.* **1**, 226–251, 2004.

[4] D. Chakrabarti, C. Faloutsos. Graph Mining: Laws, Generators, and Algorithms. *ACM Comp. Surv.* **38**, 69–123, 2006.