

Models for the Compressible Web*

Flavio Chierichetti[†] Ravi Kumar[‡] Silvio Lattanzi[§] Alessandro Panconesi[†]
Prabhakar Raghavan[‡]

Abstract

Graphs resulting from human behavior (the web graph, friendship graphs, etc.) have hitherto been viewed as a monolithic class of graphs with similar characteristics; for instance, their degree distributions are markedly heavy-tailed. In this paper we take our understanding of behavioral graphs a step further by showing that an intriguing empirical property of web graphs — their compressibility — cannot be exhibited by well-known graph models for the web and for social networks. We then develop a more nuanced model for web graphs and show that it does exhibit compressibility, in addition to previously modeled web graph properties.

1 Overview

There are three main reasons for modeling and analyzing graphs arising from the Web and from social networks: (1) they model social and behavioral phenomena whose graph-theoretic analysis has led to significant societal impact (witness the role of link analysis in web search); (2) from an empirical standpoint, these networks are several orders of magnitude larger than those studied hitherto (search companies are now working on crawls of 100 billion pages and beyond); (3) from a theoretical standpoint, stochastic processes built from independent random events — the classical basis of the design and analysis of computing artifacts — are no longer appropriate. The characteristics of such *behavioral* graphs (viz., graphs arising from human behavior) demand the design and analysis of new stochastic processes in which elementary events are highly dependent. This in turn demands new analysis and insights that are likely to be of utility in many other applications of probability and statistics.

In such analysis, there has been a tendency to lump together behavioral graphs arising from a variety of contexts, to be studied using a common set of models and tools. It has been observed [3, 9, 23] for instance that the directed graphs arising from such diverse phenomena as the web graph (pages are nodes and hyperlinks are edges), citation graphs, friendship graphs, and email traffic graphs all exhibit *power laws* in their degree distributions: the fraction of nodes with indegree $k > 0$ is proportional to $1/k^\alpha$ typically for some $\alpha > 1$; random graphs generated by classic Erdős–Rényi models cannot exhibit such power laws. To explain the power law degree distributions seen in behavioral graphs, several models have been developed for generating random graphs [2, 3, 7, 8, 11, 18, 22, 26] in which dependent events combine to deliver the observed power laws.

* A preliminary version of this paper appeared in the 50th IEEE Symposium on Foundations of Computer Science [12].

[†]Dipartimento di Informatica, Sapienza University of Rome. Email: {flavio, ale}@di.uniroma1.it

[‡]Google, Mountain View, CA. Email: {ravi.k53, pragh.prof}@gmail.com. This work was done when the author was at Yahoo! Labs.

[§]Google, New York, NY. Email: silviol@google.com. This work was done when the author was at Dipartimento di Informatica, Sapienza University.

While the degree distribution is a fundamental but local property of such graphs, an important global property is their compressibility — the number of bits needed to store each edge in the graph. Compressibility determines the ability to efficiently store and manipulate these massive graphs [19, 36, 32]. An intriguing set of papers by Boldi, Santini, and Vigna [4, 5, 6] shows that the web graph is highly compressible: it can be stored such that each edge requires only a small constant number — between one and three — of bits on average; a more recent experimental study confirms these findings [10]. These empirical results suggest the intriguing possibility that the Web can be described with only $O(1)$ bits per edge on average. Two properties are at the heart of the compression algorithm of Boldi and Vigna [5]. First, once web pages are sorted lexicographically by URL, the set of outlinks of a page exhibits locality; this can plausibly be attributed to the fact that nearby pages are likely to come from the same web site’s template. Second, the distribution of the lengths of edges follows a power law with exponent > 1 (the length of an edge is the distance of its endpoints in the ordering); this turns out to be crucial for high compressibility. This prompts the natural question: *Can we model the compressibility of the web graph, in particular mirroring the properties of locality and edge length distribution, while maintaining other well-known properties such as power law degree distribution?*

Main results. Our first set of results in this paper is to show that the best known models for the web graph cannot account for compressibility, in the sense that they require $\Omega(\log n)$ bits storage per edge on average. This holds even when these graphs are represented just in terms of their topology (i.e., with all labels stripped away). Specifically, we show that the preferential attachment model [3, 7], the ACL model [2], the copying model [22], the Kronecker product model [25], and Kleinberg’s model for navigability¹ on social networks [20], all have large entropy in the above sense.

This set of theorems, together with the empirical evidence of [5, 10] discussed earlier, provide a strong motivation for our second result: a new model for the web graph that has constant entropy per edge, while preserving crucial properties of previous models such as the power law distribution of indegrees, a large number of communities (i.e., bipartite cliques), small diameter, and a high clustering coefficient. In this model, nodes lie on the line and when a new node arrives it selects an existing node uniformly at random, placing itself on the line to the immediate left of the chosen node. An edge from the new to the chosen node is added, and moreover all outgoing edges of the chosen node but one are copied (these edges are chosen uniformly at random).

Although it might be difficult to distinguish experimentally between graphs that require only $O(1)$ bits per edge and those requiring, say, $\epsilon \log n$ bits, in order to assess the entropy of the real web graph, the point is that the compressibility of our model relies upon important structural properties of real web graphs that previous models, in view of our lower bounds, provably cannot have. The crucial property that we are able to reproduce is the power law distribution of edge lengths. Intuitively, this gives a long-get-longer effect: a long edge is likely to receive the new node (which selects its position uniformly at random) under its protective wing, and the longer it gets, the more likely it is to attract new nodes. Using this, we show that the graphs generated by our model are compressible to $O(1)$ bits per edge using the same algorithm of [5]. This holds even if the line embedding is not known (unlike the algorithm in [5]). Thus the graphs generated by our model are compressible in a strong sense. Both algorithms take linear time.

The graphs generated by our model also exhibit other salient properties known for the real web graph: power law indegree distribution, high clustering coefficient, small undirected diameter, and a large number of bipartite cliques. We also show how to extend our basic model to generate graphs with power law outdegree distributions.

¹Since navigability is a crucial property of real-life social networks (cf. [16, 27, 34]), it is tempting to conjecture that social networks are incompressible; see, for instance, [13].

Technical contributions and guided tour. In Section 3 we prove that several well-known web graph models are not compressible, i.e., they need $\Omega(\log n)$ bits per edge, i.e., the maximum possible. We remark that this incompressibility result holds even after the labels of nodes and orientations of edges are removed.

Sections 4 presents our new model and Sections 5, 6, and 8 establish its basic properties; in Section 8.4 we extend our model so to obtain a power law outdegree distribution.

Although our new model might at first sight closely resembles a prior *copying* model of [22], it differs in fundamental respects. First, our new model successfully admits the global property of compressibility which the copying model *provably* does not. Second, while the analysis of the distribution of the indegrees is rather standard, the crucial property that edge lengths are distributed according to a power law requires an entirely novel analysis. In particular, the proof requires a very delicate understanding of the structural properties of the graphs generated by our model in order to establish concentration of measure. Section 7 addresses the compressibility of our model, where we also provide an efficient algorithm to compress graphs generated by our model.

Related prior work. The observation of power law degree distributions in behavioral (and other) graphs has a long history [3, 23]; indeed, such distributions predate the modern interest in social networks through observations in linguistics [37] and sociology [31]; see the survey by Mitzenmacher [29]. Simon [31], Mandelbrot [28], Zipf [37], and others have provided a number of explanations for these distributions, attributing them to the dependencies between the interacting humans who collectively generate these statistics. These explanations have found new expression in the form of rich-get-richer and herd-mentality theories [3, 35]. Early rigorous analyses of such models include [2, 7, 14, 22]. Whereas Kumar et al. [22] and Borgs et al. [8] focused on modeling the web graph, the models of Aiello, Chung, and Lu (ACL) [2], Kleinberg [20], Lattanzi and Sivakumar [24], and Leskovec et al. [25] addressed social graphs in which people are nodes and the edges between them denote friendship. The ACL model is in fact known not to be a good representation of the web graph [23], but is a plausible model for human social networks. Kleinberg’s model of social networks focuses on their *navigability*: it is possible for a node to find a short route to a target using only local, myopic choices at each step of the route. The papers by Boldi, Santini, and Vigna [4, 5, 6] suggests that the web graph is highly compressible (see also [1, 10, 13, 32]).

2 Preliminaries

The graph models we study will either have a fixed number of nodes or will be evolving models in which nodes arrive in a discrete-time stochastic process; for many of them, the number of edges will be linear in the number of nodes. We analyze the space needed to store a graph randomly generated by the models under study; this can be viewed in terms of the entropy of the graph generation process. Note that a naive representation of a graph would require $\Omega(\log n)$ bits per edge; entropically, one can hope for no better for an Erdős–Rényi graph. We are particularly interested in the case when the amortized storage per edge can be reduced to a constant. As in the work of Boldi and Vigna [5, 6], we view the nodes as being arranged in a linear order. To prove compressibility we then study the distribution of edge *lengths* — the distance in this linear order between the end-points of an edge.

Background. Given a function $f : A_1 \times \dots \times A_n \rightarrow \mathbb{R}$, we say that f satisfies the *c-Lipschitz property* if, for any sequence $(a_1, \dots, a_n) \in A_1 \times \dots \times A_n$, and for any i and $a'_i \in A_i$,

$$|f(a_1, \dots, a_{i-1}, a_i, a_{i+1}, \dots, a_n) - f(a_1, \dots, a_{i-1}, a'_i, a_{i+1}, \dots, a_n)| \leq c.$$

In order to establish that certain events occur w.h.p., we will make use of the following concentration result known as the *method of bounded differences* (cf. [17]).

Theorem 1 (Method of bounded differences). *Let X_1, \dots, X_n be independent r.v.'s. Let f be a function on X_1, \dots, X_n satisfying the c -Lipschitz property. Then,*

$$\Pr[|f(X_1, \dots, X_n) - \mathbb{E}[f(X_1, \dots, X_n)]| > t] \leq 2e^{-t^2/(c^2n)}.$$

The *Gamma function* is defined as $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$. We use these properties of the Gamma function: (i) $\Gamma(x+1) = x\Gamma(x)$, (ii) $\Gamma(x)\Gamma(x+\frac{1}{2}) = \Gamma(2x)2^{1-2x}\sqrt{\pi}$, and (iii) for constants $a, b \in \mathbb{R}$, $\lim_{n \rightarrow \infty} \frac{\Gamma(n+a)}{\Gamma(n+b)}n^{b-a} = 1$. We also prove the following lemma about the Gamma function. We will use it in the compressibility analysis of our new model.

Lemma 2. *Let $a, b \in \mathbb{R}^+$ be such that $b \neq a+1$. For each $t \in \mathbb{Z}^+$, it holds that*

$$\sum_{i=1}^t \frac{\Gamma(i+a)}{\Gamma(i+b)} = \frac{1}{b-a-1} \cdot \left(\frac{\Gamma(a+1)}{\Gamma(b)} - \frac{\Gamma(t+a+1)}{\Gamma(t+b)} \right).$$

Proof. We start by giving an expression of $\frac{\Gamma(i+a)}{\Gamma(i+b)}$, for $i \geq 1$, which we will use to telescope the sum. Consider the following chain of equations:

$$\begin{aligned} b-a-1 &= (i+b-1) - (i+a) \\ \frac{\Gamma(i+a)}{\Gamma(i+b)} \cdot (b-a-1) &= \frac{\Gamma(i+a)}{\Gamma(i+b)} \cdot (i+b-1) - \frac{\Gamma(i+a)}{\Gamma(i+b)} \cdot (i+a) \\ \frac{\Gamma(i+a)}{\Gamma(i+b)} \cdot (b-a-1) &= \frac{\Gamma(i+a)}{\Gamma(i+b-1)} - \frac{\Gamma(i+a+1)}{\Gamma(i+b)} \\ \frac{\Gamma(i+a)}{\Gamma(i+b)} &= \frac{1}{b-a-1} \cdot \left(\frac{\Gamma(i+a)}{\Gamma(i+b-1)} - \frac{\Gamma(i+a+1)}{\Gamma(i+b)} \right) \end{aligned}$$

Then, by telescoping, we get:

$$\begin{aligned} \sum_{i=1}^t \frac{\Gamma(i+a)}{\Gamma(i+b)} &= \frac{\left(\frac{\Gamma(a+1)}{\Gamma(b)} - \frac{\Gamma(a+2)}{\Gamma(b+1)} \right) + \left(\frac{\Gamma(a+2)}{\Gamma(b+1)} - \frac{\Gamma(a+3)}{\Gamma(b+2)} \right) + \dots + \left(\frac{\Gamma(a+t)}{\Gamma(b+t-1)} - \frac{\Gamma(a+t+1)}{\Gamma(b+t)} \right)}{b-a-1} \\ &= \left(\frac{\Gamma(a+1)}{\Gamma(b)} - \frac{\Gamma(a+t+1)}{\Gamma(b+t)} \right) \cdot \frac{1}{b-a-1}, \end{aligned}$$

proving the claim. □

3 Incompressibility of the existing models

In this section we prove the inherent incompressibility of commonly-studied random graph models for social networks and the web. We show that on average $\Omega(\log n)$ bits per edge are necessary to store graphs generated by several well-known models for web/social networks, including the preferential attachment and the copying models. In our lower bounds, we show that the random graph produced by the models we consider are incompressible, *even after removing the labels of their nodes and orientations of their edges*. Given a labeled/directed graph and its unlabeled/undirected counterpart, the latter is more compressible than the former; in fact, the gap can be arbitrarily large. Thus the task of proving incompressibility of unlabeled/undirected versions of graphs generated by various models is made more challenging. (Note that

it is crucial to analyze the compressibility of unlabeled graphs — the experiments on web graph [5, 6] show how just the edges can be compressed using only ≈ 2 bits per edge.)

We now give some intuition on why one cannot preclude an incompressible directed/labeled graph from becoming very compressible after removing the labels and directions. Consider the following (non-graph) random process. Suppose we have two bins B_1 and B_2 and suppose we toss two independent fair coins c_1, c_2 . If c_1 is head (resp., tail), then we place a white (resp., black) ball in B_1 . Analogously, if c_2 is head (resp., tail), then we place a white (resp., black) ball in B_2 . Now, consider the r.v. X describing the status of the two distinguishable bins. It has four possible outcomes $((W, W), (W, B), (B, W), (B, B))$ and each of them is equally likely; thus $H(X) = 2$. Now, suppose we empty the bins B_1 and B_2 on a table, and let Y be the random variable describing the status of the table after the two balls are placed on it. Y has three possible outcomes $(\{W, W\}, \{W, B\}, \{B, B\})$ and its entropy is $H(Y) = \frac{3}{2} < 2 = H(X)$. Similarly, for n coins and n bins, we have $H(X_n) = n$ and $H(Y_n) = \Theta(\log n)$. Thus, we can get an exponential gap between the entropies of the labeled (i.e., each outcome can be matched to the coin toss that determined it) and the unlabeled processes. For a graph-based example, suppose we choose a labeled transitive tournament on n nodes u.a.r. There are $n!$ such graphs, each equally likely, so that the entropy would be $\log(n!) = \Theta(n \log n)$. On the other hand, there exists a single unlabeled transitive tournament, i.e., the entropy of the unlabeled version is zero.

3.1 Proving incompressibility

Let \mathcal{G}_n denote the set of all directed labeled graphs on n nodes. Let $P_n^\theta : \mathcal{G}_n \rightarrow [0, 1]$ denote the probability distribution on \mathcal{G}_n induced by the random graph model θ . In this paper we consider the preferential attachment model ($\theta = \text{pref}$), the ACL model ($\theta = \text{acl}$), the copying model ($\theta = \text{copy}$), the Kronecker multiplication model ($\theta = \text{krm}$), and Kleinberg’s small-world model ($\theta = \text{kl}$).

For a given θ , let $H(P_n^\theta)$ denote the Shannon entropy of the distribution P_n^θ , i.e., the average number of bits needed to represent a directed labeled random graph generated by θ . Our goal is to obtain lower bounds on the representation. Our main technical tool is the following simple lemma.

Lemma 3 (Min-entropy bound). *Let $\mathcal{G}_n^* \subseteq \mathcal{G}_n$, $P^+ \leq \sum_{G \in \mathcal{G}_n^*} P_n^\theta(G)$, and $P^* \geq \max_{G \in \mathcal{G}_n^*} P_n^\theta(G)$. Then, $H(P_n^\theta) \geq P^+ \cdot \log(1/P^*)$.*

Proof.

$$H(P_n^\theta) = \sum_{G \in \mathcal{G}_n} P_n^\theta(G) \log \frac{1}{P_n^\theta(G)} \geq \sum_{G \in \mathcal{G}_n^*} P_n^\theta(G) \log \frac{1}{P_n^\theta(G)} \geq \sum_{G \in \mathcal{G}_n^*} P_n^\theta(G) \log \frac{1}{P^*} \geq P^+ \cdot \log \frac{1}{P^*}. \quad \square$$

Thus, to obtain lower bounds on $H(P_n^\theta)$, we will upper bound $\max_{G \in \mathcal{G}_n^*} P_n^\theta(G)$ by P^* and lower bound $\sum_{G \in \mathcal{G}_n^*} P_n^\theta(G)$ by P^+ , for a suitably chosen $\mathcal{G}_n^* \subseteq \mathcal{G}_n$. For good lower bounds on $H(P_n^\theta)$, \mathcal{G}_n^* has to be chosen judiciously. Indeed, choosing a large \mathcal{G}_n^* (i.e., \mathcal{G}_n) might yield a large P^* ; choosing a small \mathcal{G}_n^* might result in a small P^+ . To get a strong lower bound, we need to guarantee that P^* is small, and yet P^+ is large.

Let \mathcal{H}_n denote the set of all undirected unlabeled graphs on n nodes. Let $\varphi : \mathcal{G}_n \rightarrow \mathcal{H}_n$ be the many-to-one map that discards node and edge labels and edge orientations. For a given model θ , let $Q_n^\theta : \mathcal{H}_n \rightarrow [0, 1]$ be the probability distribution such that $Q_n^\theta(H) = \sum_{\varphi(G)=H} P_n^\theta(G)$. Clearly, $H(Q_n^\theta) \leq H(P_n^\theta)$ and therefore, lower bounds on $H(Q_n^\theta)$ are stronger and harder to obtain.

In the following subsections we consider a number of Web Graph models, showing that each of them requires $\Omega(\log n)$ bits per link, i.e., they are incompressible. We consider, in this order, the preferential attachment model [7], the Aiello–Chung–Lu (ACL) model [2], the copying model [22], the Kronecker multiplication model [25] and Kleinberg’s small-world model [21].

3.2 Incompressibility of the preferential attachment model

Consider the preferential attachment model ($\text{pref}[k]$) defined in [7]. This model is parametrized by an integer $k \geq 1$. At time 1, the (undirected) graph consists of a single node x_1 with one self-loop. At time $t > 1$,

- (1) a new node x_t , labeled t , is added to the graph;
- (2) a random node y is chosen from the graph with probability proportional to its current degree (in this phase, the degree of x_t is taken to be 1);
- (3) the edge $x_t \rightarrow y$, labeled $t \bmod k$, is added to the graph;² and
- (4) if t is a multiple of k , nodes $t - k + 1, \dots, t$ are merged, preserving self-loops and multi-edges.

For $k = 1$, note that the graphs generated by the above model are forests. Since there are $2^{O(n)}$ unlabeled forests on n nodes (see e.g. [30]), whose edges can be directed in at most 2^n ways, $H(Q_n^{\text{pref}[k]}) = O(n)$, i.e., the graph without labels and edge orientations is compressible to $O(1)$ bits per edge. The more interesting case is when $k \geq 2$ for which we show an incompressibility bound.

We underscore the importance of a good choice of \mathcal{G}_n^* in applying Lemma 3. Consider the graph G having the first node of degree $k(n+1)$ and the other $n-1$ nodes of degree k . Clearly, $P_n^{\text{pref}[k]}(G) = \prod_{i=k+1}^{nk} \frac{k-1+i}{2i-1} \geq 2^{-nk}$. Thus, choosing a set \mathcal{G}_n^* containing G , would force us to have $P^* \geq 2^{-nk}$ so that the entropy bound given by Lemma 3 would only be $H(P_n^{\text{pref}[k]}) \geq nk = \Theta(n)$. (A similar issue would be encountered in the unlabeled case as well.) A careful choice of \mathcal{G}_n^* , however, yields a better lower bound.

Theorem 4. $H(Q_n^{\text{pref}[k]}) = \Omega(n \log n)$, for $k \geq 2$.

Proof. Let G be a graph generated by $\text{pref}[k]$. Let $\deg_t(x_i)$, for $i \leq t$, be the degree of the i th inserted node at time t in G . By [15, Lemma 6], with probability $1 - O(n^{-3})$, for each $1 \leq t \leq n$, each node x_i , $1 \leq i \leq t$, will have degree $\deg_t(x_i) < (\sqrt{t/i}) \log^3 n$ in G .

Let $t^* = \lceil \sqrt[3]{n} \rceil$ and let ξ be the event: “ $\exists t \geq t^*, \sum_{i=1}^{t^*} \deg_t(x_i) \geq n^{3/4}$.” At time n , the sum of the degrees of nodes x_1, \dots, x_{t^*} can be bounded from above by

$$\sum_{i=1}^{t^*} \deg_n(x_i) \leq \sum_{i=1}^{t^*} \sqrt{\frac{n}{i}} \log^3 n = \sqrt{n} \log^3 n \sum_{i=1}^{t^*} i^{-1/2} < O(n^{3/4}),$$

w.h.p. . Indeed, $\Pr[\xi] = O(n^{-3})$.

Now define $t^+ = \lceil \epsilon n \rceil$, for some small enough $\epsilon > 0$; let n be large enough such that $t^* < t^+$. We call a node added after time t^+ *good* if it is not connected to any of the first t^* nodes. To bound the number of good nodes from below, we condition on ξ , and we upper bound the number of bad nodes. Using a union bound, the probability that node x_t for $t \geq t^*$ is bad can be upper bounded by $k \cdot n^{3/4} / (\epsilon n) = O(n^{-1/4})$.

Let ξ' be the event: “at least $(1 - 2\epsilon)n$ nodes are good”; by stochastic dominance, the event ξ' happens w.h.p. . In our application of Lemma 3, we will choose $\mathcal{G}_n^* \subseteq \mathcal{G}_n$ to be the set of graphs satisfying $\xi \cap \xi'$. Thus, $P^+ = \Pr[\xi \cap \xi'] = 1 - o(1)$. Moreover,

$$\max_{G \in \mathcal{G}_n^*} P_n^{\text{pref}[k]}(G) \leq \left(\frac{\sqrt{\frac{n}{\sqrt[3]{n}}} \log^3 n}{kn} \right)^{(1-2\epsilon)kn} \leq \left(O(n^{-2/3+\epsilon}) \right)^{2(1-2\epsilon)n} \leq n^{-\frac{4}{3}n + \frac{14}{3}\epsilon n} = \rho.$$

(Notice that, by applying Lemma 3 at this point, we already have that $H(P_n^{\text{pref}[k]}) \geq \Omega(n \log n)$.)

Now, we proceed to lower bound $H(Q_n^{\text{pref}[k]})$ using an upper bound on $|\varphi^{-1}(H)|$ for $H \in \mathcal{H}_n^{\text{pref}[k]}$. Given some such H , it is possible to determine for each of its edges which endpoint was responsible for adding

²In the original PA model, edges are both undirected and unlabeled: we direct and label them for simplicity of exposition. The entropy lower bound will hold for the undirected and unlabeled version of these graphs.

the edge to the graph. This task is easy for edges incident to any node of degree k , for that node will have necessarily added all k edges to the graph. So, we can remove all degree k nodes from the graph and repeat this process until the graph becomes empty.

Thus, H could have been produced from at most $n! \cdot (k!)^n$ labeled graphs, since there are at most $n!$ ways of labeling the nodes, and $k!$ ways of labeling each of the “outgoing” edges of each node. In other words, $|\varphi^{-1}(H)| \leq n! \cdot (k!)^n \leq n^n k^{kn}$. Then, choosing $\mathcal{H}_n^* \subseteq \mathcal{H}_n$ to be the set of unlabeled graphs obtained by removing labels from \mathcal{G}_n^* , $\mathcal{H}_n^* = \{\varphi(G) \mid G \in \mathcal{G}_n^*\}$, we obtain $P^+ = 1 - o(1)$, and

$$\max_{H \in \mathcal{H}_n^*} Q_n^{\text{pref}[k]}(H) \leq \rho \cdot n^n \cdot k^{kn} = n^{-\Omega(n)} k^{kn} = P^*.$$

Finally, an application of Lemma 3 gives $H(Q_n^{\text{pref}[k]}) \geq P^+ \cdot \log \frac{1}{P^*} = \Omega(n \log n)$, completing the proof. \square

3.3 Incompressibility of the ACL model

We recall the ACL model (model A in [2]). This model ($\text{acl}[\alpha]$) is parametrized by some $\alpha \in (0, 1)$. At time 1, the graph consists of a single node. At time $t + 1$, a coin is tossed: with probability $1 - \alpha$ a new node is added to the graph, otherwise with probability α an edge from x to y is added to the graph, where node x is chosen with probability proportional to its outdegree while node y is chosen randomly with probability proportional to its indegree.

We assume that $\alpha > 1/2$. This is because the edge density of the graph generated by the model is $\alpha/(1 - \alpha)$ w.h.p. . If $\alpha < 1/2$ there are many more nodes than edges, an uninteresting case both in theory and in practice.

Theorem 5. $H(Q_n^{\text{acl}[\alpha]}) = \Omega(n \log n)$, provided that $\alpha > 1/2$.³

Proof. Assuming $\alpha > 1/2$, let \mathcal{G}'_n be the set of all time-labeled graphs, that can be generated by $\text{acl}[\alpha]$ model in n time steps, where the label represents the time when a node or an edge was added to the graph. Let \mathcal{H}'_n be the set of all undirected and unlabeled graphs that can be obtained by removing the orientations and (time-)labels from the graphs in \mathcal{G}'_n .⁴

Let $P_n^{\text{acl}[\alpha]} : \mathcal{G}'_n \rightarrow [0, 1]$ denote the probability distribution induced on \mathcal{G}'_n by the model $\text{acl}[\alpha]$. We define the following two events.

ξ : the number of edges is $\alpha n \pm o(n)$, while the number of nodes is $(1 - \alpha)n \pm o(n)$, and

ξ' : the number of edges going from a node of $O(1)$ outdegree to a node of $O(1)$ indegree is at least $(\alpha - \epsilon)n$, for some $\epsilon > 0$ to be fixed later.

The proof follows from two lemmas whose proof is deferred to after this theorem for clarity of exposition. Lemma 6 establishes that $\Pr[\xi \wedge \xi'] = 1 - o(1)$. Furthermore, using the notation of Lemma 3, Lemma 7 states that $P^* := \max_{G' \in \mathcal{G}'_n} P_n^{\text{acl}[\alpha]}(G') \leq n^{-(2\alpha - \epsilon)n}$

Let $\mathcal{G}'_n^* \subseteq \mathcal{G}'_n$ be the subset of \mathcal{G}'_n containing the graphs satisfying $\xi \wedge \xi'$. Then, recalling again the notation of Lemma 3, it follows from Lemma 6 that $P^+ = 1 - o(1)$.

To complete the proof, let $\varphi' : \mathcal{G}'_n \rightarrow \mathcal{H}'_n$ be the map that removes edge and node labels from the graphs of \mathcal{G}'_n . As before, $Q_n^{\text{acl}[\alpha]}(H') = \sum_{\varphi'(G')=H'} P_n^{\text{acl}[\alpha]}(G')$. Note that for each H' we have that $|\varphi'(G')| \leq n!$

³Here we do not use the probability distribution Q on the graphs of n nodes — in the $\text{acl}[\alpha]$ model the number of nodes is a r.v. $Q_n^{\text{acl}[\alpha]}$ denotes the probability distribution on the graphs that can be generated by the $\text{acl}[\alpha]$ model in n steps.

⁴Note that here it would be unnatural to consider the previously defined class \mathcal{G}_n , as the number of nodes in the $\text{acl}[\alpha]$ model is an r.v. The same holds for \mathcal{H}_n .

(as each element of the graph has one label out of the set $\{1, \dots, n\}$). Thus,

$$\max_{G' \in \mathcal{G}_n^{l*}} P_n^{\text{acl}[\alpha]}(G) \leq n! \cdot n^{-(2\alpha-\epsilon)n} \leq n^{-(2\alpha-\epsilon)n+n} = n^{(1-2\alpha+\epsilon)n}.$$

The proof can be concluded by applying Lemma 3. \square

We now prove the two statements used in the preceding proof.

Lemma 6. $\Pr[\xi \wedge \xi'] = 1 - o(1)$.

Proof. By the Chernoff bound, $\Pr[\xi] = 1 - o(1)$. Thus it suffices to show that $\Pr[\xi'] = 1 - o(1)$.

Let X_i^t (Y_i^t) be the r.v. denoting the number of nodes having indegree (outdegree) i at time t . In [2] it is shown that

$$\frac{\mathbb{E}[X_i^t]}{t} = \frac{\mathbb{E}[Y_i^t]}{t} = \frac{1-\alpha}{\alpha} \Gamma\left(1 + \frac{1}{\alpha}\right) \frac{\Gamma(i)}{\Gamma\left(i + 1 + \frac{1}{\alpha}\right)} \pm O\left(\frac{1}{t}\right),$$

and that

$$\begin{aligned} \Pr\left[|X_i^t - \mathbb{E}[X_i^t]| > \sqrt{2t} \log n + 2\right] &< \exp(-\log^2 n), \\ \Pr\left[|Y_i^t - \mathbb{E}[Y_i^t]| > \sqrt{2t} \log n + 2\right] &< \exp(-\log^2 n). \end{aligned}$$

Note that, by union bound, each one of the r.v.'s X_i^t, Y_i^t can be shown to deviate from their mean by at most the stated error term w.h.p. .

Let j be an integer to be fixed later. An edge is *good* if it goes from a node of outdegree less than j to a node of indegree less than j . Let us denote by Z_j^t the number of good edges at time t . Note that $Z_j^{t-1} + 1 \geq Z_j^t \geq Z_j^{t-1} - 2j$. This is because at most one edge is added in a single step and adding an edge changes the degree of at most 2 nodes. Thus, the number of good edges can decrease by at most $2j$ in a single step, i.e., Z_j^t satisfies the $(2j)$ -Lipschitz condition.

Then,

$$\mathbb{E}[Z_j^t] = \mathbb{E}[Z_j^{t-1}] + \Pr[Z_j^t = Z_j^{t-1} + 1] - \sum_{i=1}^{2j} i \Pr[Z_j^t = Z_j^{t-1} - i].$$

In order to increase the number of good edges, a node of indegree less than j and a node of outdegree less than j must be chosen as the ending and the starting point of the new edge.

$$\Pr[Z_j^t = Z_j^{t-1} + 1] = \alpha \frac{\left(\sum_{i=1}^{j-1} i X_i^{t-1}\right) \left(\sum_{i=1}^{j-1} i Y_i^{t-1}\right)}{(t-1)^2}.$$

For the number of good edges to decrease, either the origin of the new edge has outdegree j , or the destination of the new edge has indegree j . Thus,

$$\Pr[Z_j^t < Z_j^{t-1}] \leq \frac{j X_j^{t-1}}{t-1} + \frac{j Y_j^{t-1}}{t-1}.$$

By calculations,

$$\sum_{i=1}^{2j} i \Pr[Z_j^t = Z_j^{t-1} - i] \leq 2j \sum_{i=1}^{2j} \Pr[Z_j^t = Z_j^{t-1} - i] \leq 2j^2 \frac{X_j^{t-1} + Y_j^{t-1}}{t-1}.$$

Thus,

$$\mathbb{E}[Z_j^t] \geq \mathbb{E}[Z_j^{t-1}] + \alpha \frac{\mathbb{E}\left[\left(\sum_{i=1}^{j-1} i X_i^{t-1}\right) \left(\sum_{i=1}^{j-1} i Y_i^{t-1}\right)\right]}{(t-1)^2} - 2j^2 \frac{\mathbb{E}[X_j^{t-1}] + \mathbb{E}[Y_j^{t-1}]}{t-1}.$$

Recall that $\frac{1}{2} < \alpha < 1$. We have that, with probability $1 - o(1)$, for all $\log^3 n \leq t \leq n$ and $1 \leq i \leq j-1$, we have

$$iX_i^t = iY_i^t = \frac{1-\alpha}{\alpha} \Gamma\left(1 + \frac{1}{\alpha}\right) \frac{\Gamma(i+1)}{\Gamma\left(i+1 + \frac{1}{\alpha}\right)} t \pm o\left(\frac{1}{\sqrt{\log n}}\right).$$

Thus w.h.p., for all $t \geq \log^3 n$,

$$\sum_{i=1}^{j-1} i \frac{X_i^t}{t} = \sum_{i=1}^{j-1} i \frac{Y_i^t}{t} = 1 - \frac{\Gamma(j+1)\Gamma\left(1 + \frac{1}{\alpha}\right)}{\Gamma\left(j + \frac{1}{\alpha}\right)} \pm o\left(\frac{j}{\sqrt{\log n}}\right).$$

As j is a constant, the error term is $o(1)$. Then,

$$\mathbb{E}[Z_j^t] \geq \mathbb{E}[Z_j^{t-1}] + \alpha \left(1 - \frac{\Gamma(j+1)\Gamma\left(1 + \frac{1}{\alpha}\right)}{\Gamma\left(j + \frac{1}{\alpha}\right)}\right)^2 - 4 \frac{1-\alpha}{\alpha} \Gamma\left(1 + \frac{1}{\alpha}\right) \frac{j\Gamma(j+1)}{\Gamma\left(j + 1 + \frac{1}{\alpha}\right)} \pm o\left(\frac{j}{\sqrt{\log n}}\right).$$

Note that, as j grows, both $\frac{\Gamma(j+1)\Gamma\left(1 + \frac{1}{\alpha}\right)}{\Gamma\left(j + \frac{1}{\alpha}\right)}$ and $\frac{j\Gamma(j+1)}{\Gamma\left(j + 1 + \frac{1}{\alpha}\right)}$ tend to 0. That is, for each ϵ_1 , there exists a $j = j(\epsilon_1)$ such that

$$\mathbb{E}[Z_j^t] \geq \mathbb{E}[Z_j^{t-1}] + (1 - \epsilon_2)\alpha. \quad (1)$$

For each j , and for each t , we will define a B_j^t in such a way that, w.h.p., $B_j^t \leq \mathbb{E}[Z_j^t]$. Let $B_j^t = 0$ for $t \leq \lceil \log^3 n \rceil$ so that the base case is true. Define $B_j^t = (t - \lceil \log^3 n \rceil)(1 - \epsilon_2)\alpha$. This definition satisfies $B_j^t \leq \mathbb{E}[Z_j^t]$ for all t — this can be shown by induction on (1). Recall that all these hold w.h.p. . As we have already argued, the r.v. Z_k^t satisfies the $(2j)$ -Lipschitz condition, i.e., using [2, Lemma 1], $Z_j^t = \mathbb{E}[Z_j^t] \pm o(t) \geq t(1 - \epsilon_2)\alpha$, for every $t \geq \lceil \log^3 n \rceil$, w.h.p. .

In particular for any $\epsilon_2 > 0$ there exists a $j = j(\epsilon)$ s.t. $Z_j^n \geq n(1 - \epsilon_2)\alpha$, w.h.p. . \square

Lemma 7. *Conditioned on $\xi \wedge \xi'$, $\max_{G' \in \mathcal{G}'_n} P_n^{\text{acl}[\alpha]}(G') \leq n^{-(2\alpha - \epsilon)n}$.*

Proof. Since we condition on ξ' , there are at least $n(1 - \epsilon_2)\alpha$ good edges. These good edges are labeled with their order of arrival. For $i \geq \epsilon_3 \alpha n$, the probability that the i th arrived edge is good is at most $\frac{j^2}{i^2} \leq \frac{j^2}{(\epsilon_3 \alpha n)^2}$.

The probability that all the edges with label at least $\epsilon_3 \alpha n$ are good is at most

$$\left(\frac{j}{\epsilon_3 \alpha n}\right)^{2(1 - \epsilon_3)\alpha n} \leq \left(\frac{j}{\epsilon_3 \alpha}\right)^{2(1 - \epsilon_3)\alpha n} n^{-2(1 - \epsilon_3)\alpha n} \leq n^{-(2\alpha - \epsilon)n}.$$

Thus, the maximum probability of generating a graph in \mathcal{G}'_n , conditioned on $\xi \wedge \xi'$, is at most $n^{-(2\alpha - \epsilon)n}$.

\square

3.4 Incompressibility of the copying model

We now turn our attention to the (linear growth) copying model ($\text{copy}[\alpha, k]$) of Kumar et al. [22]. This model is parametrized by an integer $k \geq 1$ and an $\alpha \in (0, 1)$. Here, k represents the outdegree of the nodes and α determines the ‘‘copying rate’’ of the graph. At time $t = 1$, the graph consists of a single node with k self-loops. At time $t > 1$,

- (1) a new node x_t is added to the graph;
- (2) a node x is chosen uniformly at random among x_1, \dots, x_{t-1} ; and
- (3) for each $i = 1, \dots, k$, an α -biased coin is flipped: with probability α , the i th outlink of x_t is chosen uniformly at random from x_1, \dots, x_{t-1} and with probability $1 - \alpha$, the i th outlink of x_t will be equal to the i th outlink of x , i.e., the i th outlink will be ‘‘copied’’.

Theorem 8. $H(Q_n^{\text{copy}[\alpha, k]}) = \Omega(n \log n)$, for $k > 2/\alpha$.

Proof. We start by noting that the copying model with outdegree k can be completely described by k independent versions of the copying model with outdegree 1. We use $\text{copy}[\alpha, k]$ to denote the copying model with k outlinks, $\mathcal{G}_{n,k}$ for the set of labeled graphs on n nodes⁵ that can be generated by $\text{copy}[\alpha, k]$, and $\mathcal{H}_{n,k}$ for the set of unlabeled graphs that can be obtained by removing labels and orientations from the graphs in $\mathcal{G}_{n,k}$.

We start with the case $k = 1$. Let $E[X_i^t]$ be the expected indegree at time t of the node inserted at time $i \leq t$. Then,

$$E[X_i^t] = \begin{cases} 0 & t = k \\ E[X_i^{t-1}] \left(1 + \frac{1-\alpha}{t-1}\right) + \frac{\alpha}{t-1} & t > i. \end{cases}$$

Note that $E[X_i^t] = \frac{\alpha \Gamma(t+1-\alpha) \Gamma(i)}{(1-\alpha) \Gamma(i+1-\alpha) \Gamma(t)} - \frac{\alpha}{1-\alpha}$. We now show that X_i^t satisfies a $O(1)$ -Lipschitz condition, with the constant depending on i and α .

Let Y_j^t denote the number of edges ‘‘copied’’, directly or indirectly, from the j th edge until time $t \geq j$. More precisely, let $S_j^j = \{e_j\}$ be the singleton set containing the j th added edge. The set S_j^t , $t > j$, will be defined as follows: if the t th edge e_t was copied from some edge in S_j^{t-1} , then $S_j^t = \{e_t\} \cup S_j^{t-1}$, otherwise $S_j^t = S_j^{t-1}$. With this notation, $Y_j^t = |S_j^t|$.

We now use the following concentration bound [17].

Theorem 9 (Method of average bounded differences). *Suppose f is some function of (possibly dependent) r.v.’s X_1, \dots, X_n . Suppose that, for each $i = 1, \dots, n$, there exists a c_i such that, for all pairs x_i, x'_i of possible values of X_i , and for any assignment $X_1 = x_1, \dots, X_{i-1} = x_{i-1}$, it holds that $|E - E'| \leq c_i$, where*

$$E = E[f(X_1, \dots, X_n) \mid X_i = x_i, X_{i-1} = x_{i-1}, \dots, X_1 = x_1],$$

$$E' = E[f(X_1, \dots, X_n) \mid X_i = x'_i, X_{i-1} = x_{i-1}, \dots, X_1 = x_1].$$

Let $c = \sum_{i=1}^n c_i^2$. Then,

$$\Pr[|f(X_1, \dots, X_n) - E[f(X_1, \dots, X_n)]| > t] \leq 2 \exp\left(-\frac{t^2}{2c}\right).$$

Let j be fixed. Our goal is to bound c_j in such a way that Theorem 9 can be applied. Observe that $E[Y_j^t] = E[Y_j^{t-1}] \cdot \left(1 + \frac{1-\alpha}{t-1}\right)$, for $t > j$, and $Y_j^j = 1$. Then, it follows that $E[Y_j^t] = \frac{\Gamma(t+1-\alpha) \Gamma(j)}{\Gamma(t) \Gamma(j+1-\alpha)}$.

⁵Nodes are labeled with $1, \dots, n$ and, for each node, its outlinks are labeled with $1, \dots, k$.

Suppose we want to bound the degree of the i th node x_i . Then, we are interested in bounding the maximum expected change c_j in the degree X_i^t of x_i over the possible choices of the j th edge, for $j = i + 1, \dots, n$. We have $c_j \leq 2 \mathbb{E} [Y_j^n]$.

Let us consider $c = \sum_{j=i+1}^n c_j^2$. We have

$$\begin{aligned} c &\leq \sum_{j=i+1}^n (2 \mathbb{E} [Y_j^n])^2 \\ &\leq 4 \left(\frac{\Gamma(n+1-\alpha)}{\Gamma(n)} \right)^2 \cdot \sum_{j=i+1}^n \left(\frac{\Gamma(j)}{\Gamma(j+1-\alpha)} \right)^2 \\ &\leq a \cdot n^{2-2\alpha} \sum_{j=i+1}^n \frac{1}{j^{2-2\alpha}}, \end{aligned}$$

for some large enough constant $a > 0$. Thus we obtain,

$$c \leq \begin{cases} a \cdot n^{2-2\alpha} \cdot \frac{i^{2\alpha-1} - n^{2\alpha-1}}{1-2\alpha} & \alpha < \frac{1}{2} \\ a \cdot n \cdot (\log n + 1) & \alpha = \frac{1}{2} \\ a \cdot n^{2-2\alpha} \cdot \frac{n^{2\alpha-1} - i^{2\alpha-1} + 1}{2\alpha-1} & \alpha > \frac{1}{2} \end{cases}$$

Let us fix $i = \lceil \epsilon n \rceil$. Then,

$$c \leq \begin{cases} a \cdot n^{2-2\alpha} \cdot n^{2\alpha-1} \frac{1-\epsilon^{1-2\alpha}}{\epsilon^{1-2\alpha}} = a \cdot n \cdot \frac{1-\epsilon^{1-2\alpha}}{\epsilon^{1-2\alpha}} & \alpha < \frac{1}{2} \\ a \cdot n \cdot (\log n + 1) & \alpha = \frac{1}{2} \\ a \cdot n^{2-2\alpha} \cdot n^{2\alpha-1} \frac{1}{2\alpha-1} + n^{2-2\alpha} = a \cdot n \cdot \frac{1}{2\alpha-1} + o(n) & \alpha > \frac{1}{2} \end{cases}$$

Thus, $c \leq O(n \log n)$. Applying Theorem 9, we get

$$\Pr \left[|X_i^t - \mathbb{E} [X_i^t]| \geq 2\sqrt{c \log n} \right] \leq 2 \exp \left(-\frac{4c \log n}{2c} \right) = 2 \exp(-2 \log n) = \frac{2}{n^2}.$$

By the union bound, with probability $1 - O(1/n)$, each node $i = \lceil \epsilon n \rceil, \lceil \epsilon n \rceil + 1, \dots, n$ will have degree upper bounded by $O(\sqrt{n \log n})$ (note that the expected degree of these nodes is constant). Conditioning (as in the proof of Theorem 4) on this event we obtain $\mathcal{G}_{n,1}^* \subseteq \mathcal{G}_{n,1}$, $P^+ = 1 - o(1)$, and for $k = 1$,

$$\max_{G \in \mathcal{G}_{n,1}^*} P_n^{\text{copy}[\alpha,1]}(G) \leq \left(O \left(\sqrt{\frac{\log n}{n}} \right) \right)^{\alpha n}.$$

Now let us consider $\text{copy}[\alpha, k]$, with $k > 1$. Since the outlinks are chosen independently, it holds that

$$\max_{G \in \mathcal{G}_{n,k}^*} P_n^{\text{copy}[\alpha,k]}(G) \leq \left(O \left(\sqrt{\frac{\log n}{n}} \right) \right)^{k\alpha n}.$$

For constant $k > 2/\alpha$, this upper bound is less than $n^{-(1+\epsilon)n}$ for some constant $\epsilon > 0$.

To show a lower bound on $\mathbb{H}(Q_n^{\text{copy}[\alpha,k]})$, we once again upper bound $|\varphi^{-1}(H)|$, for $H \in \mathcal{H}_{n,k}$. We proceed as in the proof of Theorem 4. Given H , for each node v , it is possible to determine which edges incident to v were outlinks in all the G 's such that $\varphi(G) = H$ (this can be done by induction, noting that a node of degree k in H would have indegree 0 in G). As there are exactly k labels for the outlinks of each node, and the number of nodes is n , we have that, for each $H \in \mathcal{H}_{n,k}$, $|\varphi^{-1}(H)| \leq n! \cdot (k!)^n$. The proof can be concluded as in Theorem 4. \square

3.5 Incompressibility of the Kronecker multiplication model

We now turn our attention to the Kronecker multiplication model (krm) of Leskovec et al. [25].

Given two matrices $A \in \mathbb{R}^{n \times n}$ and $B \in \mathbb{R}^{m \times m}$, their Kronecker product $A \otimes B$ is an $nm \times nm$ matrix

$$A \otimes B = \begin{pmatrix} a_{1,1}B & a_{1,2}B & \cdots & a_{1,n}B \\ a_{2,1}B & a_{2,2}B & \cdots & a_{2,n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1}B & a_{n,2}B & \cdots & a_{n,n}B \end{pmatrix},$$

where $A = \{a_{i,j}\}$ and $a_{i,j}B$ is the usual scalar product.

The Kronecker multiplication model is parametrized by a square matrix $M \in [0, 1]^{\ell \times \ell}$, and a number s of multiplication “steps”. The graph will consist of ℓ^s nodes. The edges are generated as follows. For each pair of distinct nodes (i, j) in the graph an edge going from i to j will be added independently with probability $M_{i,j}^{[s]}$, where $M_{i,j}^{[s]} = \underbrace{M \otimes M \otimes \cdots \otimes M}_{s \text{ times}}$.

It is clear that for some choices of the matrix M , the graph *will* be compressible. Indeed, if M has only 0/1 values then the random graph has zero entropy, as its construction is completely deterministic. On the other hand, we show here that there exists some M that makes the graph incompressible. Indeed, even some 2×2 matrix M would generate graphs requiring at least $\Omega(\log n)$ bits per edge. (Note that a 1×1 matrix can only produce graphs containing a single node.)

Theorem 10. *Let $\ell \geq 2$ and $1/\ell < \alpha < 1$. Then, w.h.p., $H(Q_n^{\text{krm}[M,s]}) = \Omega(m \log n)$, where $n = \ell^s$, $M = \alpha \cdot J_\ell$, and m is the number of edges.*

Proof. Consider the original directed version of the graph. Note that $M^{[s]} = \alpha^s \cdot J_{\ell^s}$. Thus the events “the edge $i \rightarrow j$ is added to the graph” are i.i.d. trials, each having probability of success α^s .

In the undirected and simple version of the graph, the events “the edge $\{i, j\}$ is added to the graph”, for $i \neq j$, are again i.i.d. trials, each of probability $\beta = 1 - (1 - \alpha^s)^2 = \Theta(\alpha^s)$. Thus we obtain an Erdős–Rényi $G_{n,p}$ graph with $n = \ell^s$ and $p = \Theta(\alpha^s)$. By a Chernoff bound, $m = \Theta(n^2 p)$, w.h.p. . Now,

$$m = \Theta(n^2 p) = \Theta(n \cdot (\ell \alpha)^s) = \Theta\left(n \cdot \left(\ell^{1+\log_\ell \alpha}\right)^s\right) = \Theta\left(n \cdot (\ell^s)^{1-\log_\ell \frac{1}{\alpha}}\right) = \Theta\left(n^{2-\log_\ell \frac{1}{\alpha}}\right).$$

Since $\alpha > \ell^{-1}$ by hypothesis, we have $\log_\ell \frac{1}{\alpha} < 1$. Thus, $m = \Theta(n^2 p)$ is a polynomial in n of degree > 1 .

Recall that, for Lemma 3 to apply, we need to find a subset $\mathcal{H}_n^* \subseteq \mathcal{H}_n$ having large total probability P^+ and such that each graph in \mathcal{H}_n^* occurs with probability at most P^* , with P^* small. The condition $\{m = \Theta(n^2 p)\}$ determines our \mathcal{H}_n^* , giving us $P^+ = 1 - o(1)$. To upper bound P^* , note that each labeled version of each graph in \mathcal{H}_n^* has probability at most $p^{\Theta(n^2 p)} \leq 2^{-\Theta(s \cdot n^2 p)}$. There are at most $n! \leq 2^{O(n \log n)}$ many labeled versions of each fixed graph in \mathcal{H}_n^* . Thus,

$$P^* \leq 2^{O(n \log n) - \Theta(s \cdot n^2 p)} = 2^{-\Theta(s \cdot n^2 p)}.$$

By Lemma 3, we have that $H(Q_n^{\text{krm}}) \geq P^+ \log(1/P^*) = \Omega(s \cdot n^2 p)$. Noting that $s = \Theta(\log n)$ and $m = \Theta(n^2 p)$ concludes the proof. \square

3.6 Incompressibility of Kleinberg’s small-world model

Let us recall Kleinberg’s small-world model (kl) [20, 21] on the d -dimensional mesh, with nodes $1, \dots, n$, for any constant $d \geq 1$. A directed labeled random graph is generated by the following stochastic process.

Each node x independently chooses a node y with probability inversely proportional to the distance in the mesh between x and y ; node x then adds the directed edge $x \rightarrow y$. These are the so-called *long-range* edges. In addition, the node x has (fixed) directed edges to its neighbors in the mesh (that is, the nodes at ℓ_1 -distance 1 from x in the mesh).

An important difference between Kleinberg’s small-world model and other models considered in this paper lies in their degree distribution. In Kleinberg’s model the degree of a node is $O(\log n)$ w.h.p. while the other models have a power law degree distribution, and thus nodes of polynomial degree w.h.p. .

For simplicity, we start by proving the following weaker result. After the proof, we will comment on how one can obtain the stronger incompressibility of $\Omega(n \log n)$.

Lemma 11. $H(Q_n^{\text{kl}}) = \Omega(n \log \log n)$.

Proof. In Kleinberg’s model, the probability of choosing a long range link at distance k from a given node is $\Theta\left(\frac{1}{k \log n}\right)$. Hence, for every node x , the maximum probability of choosing a particular long-range edge $x \rightarrow y$ is at most $c_1/\log n$, for some constant c_1 . Since each node chooses edges independently, the maximum probability of generating any labeled n -node graph is $O((c_1/\log n)^n)$ so that $\max_{G \in \mathcal{G}_n} P_n^{\text{kl}}(G) \leq (c_1/\log n)^n$, for some constant c_1 . By Lemma 3, it follows that $H(P_n^{\text{kl}}) = \Omega(n \log \log n)$.

To get a lower bound on $H(Q_n^{\text{kl}})$, we first obtain an upper bound on the number $\rho(H)$ of Hamiltonian paths in an undirected graph H with m edges (this upper bound holds for directed graphs too). Suppose that H has degree sequence $d_1 \geq \dots \geq d_n$, with $2m = \sum_{i=1}^n d_i$. Clearly, $\rho(H) \leq n \cdot \prod_{i=1}^n d_i$, where the leading n is for the different choices of the starting node. Applying the AM-GM inequality ($\sqrt[n]{\prod_{i=1}^n x_i} \leq \frac{1}{n} \sum_{i=1}^n x_i$, for non-negative x_i ’s), we have that $\rho(H) \leq n \cdot \prod_{i=1}^n d_i \leq n \cdot (2m/n)^n$.

Let $H \in \mathcal{H}_n$. By just considering all possible permutations of the node labels, we can see that $|\varphi^{-1}(H)| \leq n!$. However, not all permutations are valid. In particular, a valid permutation preserves adjacency, hence the number of valid permutations is upper bounded by the number of Hamiltonian paths in H . Since $m = O(n)$ in kl, by the above argument, $\rho(H) \leq c_2^n$, for some constant c_2 . Thus, $|\varphi^{-1}(H)| \leq c_2^n$. We have

$$Q_n^{\text{kl}}(H) = \sum_{\varphi(G)=H} P_n^{\text{kl}}(G) \leq |\varphi^{-1}(H)| \cdot (\max_{G \in \mathcal{G}_n} P_n^{\text{kl}}(G)) \leq c_2^n \left(\frac{c_1}{\log n}\right)^n = O\left(\frac{1}{\log n}\right)^n.$$

The proof follows from Lemma 3. □

The above lower bound can be improved as follows. First, we only consider graphs in which $\Omega(n)$ of the edges exist between nodes that are $n^{\Omega(1)}$ apart. By a Chernoff bound, a graph generated by Kleinberg’s model satisfies this property w.h.p. (i.e., the value P^+ of Lemma 3 is $\Omega(1)$). It can then be shown that the maximum probability of generating any such graph is at most $P^* = n^{-\Omega(n)}$. By applying Lemma 3, we obtain the following theorem,

Theorem 12. $H(Q_n^{\text{kl}}) = \Omega(n \log n)$.

Finally, we note that the similar incompressibility bounds can be obtained for the rank-based friendship model [27].

4 The new web graph model

The results in the previous section provide a strong motivation for a new web graph model that is highly compressible. This is what we do in this section. Our new model will exhibit all the “classical” properties

of previous models besides having low entropy — it can be stored using as few as $O(1)$ bits per link on average. Let $k \geq 2$ be a fixed positive integer. Our new model creates a directed simple graph (i.e., no self-loops or multi-edges) by the following process.

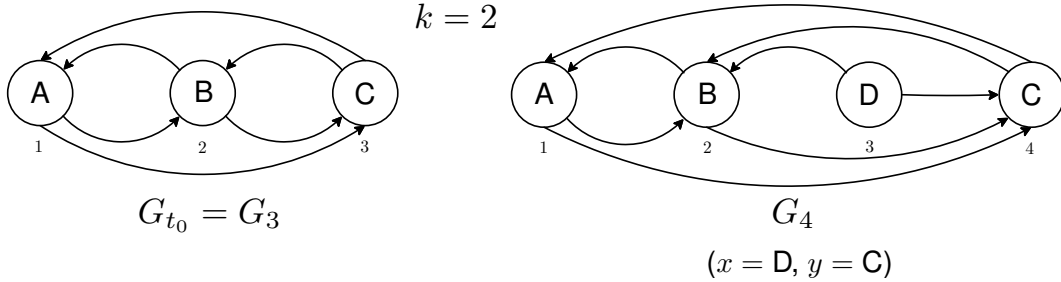


Figure 1: The new node $x = D$ chooses $y = C$ as its prototype. The edge $C \rightarrow B$ is copied and the new edge $D \rightarrow C$ is added for reference. Notice that all the edges incident to C in $G_{t_0} = G_3$ increase their length by 1 in $G_{t_0+1} = G_4$.

The process starts at time t_0 with a simple directed *seed graph* G_{t_0} whose nodes are arranged on a (discrete) line, or list. The graph G_{t_0} has t_0 nodes, each of outdegree k . Here, G_{t_0} could be, for instance, a complete directed graph with $t_0 = k + 1$ nodes.

At time $t > t_0$, an existing node y is chosen uniformly at random (u.a.r.) as a *prototype*:

- (1) a new node x is placed to the immediate left of y (so that y , and all the nodes on its right, are shifted one position to the right in the ordering),
- (2) a directed edge $x \rightarrow y$ is added to the graph, and
- (3) $k - 1$ edges are “copied” from y , i.e., $k - 1$ successors (i.e., out-neighbors) of y , say z_1, \dots, z_{k-1} , are chosen u.a.r. without replacement and the directed edges $x \rightarrow z_1, \dots, x \rightarrow z_{k-1}$ are added to the graph.

See Figure 1 for an illustration of our model.

We now give an intuitive justification of the process. Suppose that a website owner decides to add a new web page to her site; to do this, she could take one of the existing web pages from her site as a prototype, modify it as needed, add an edge to the prototype for reference — finally publishing the new page on her site. We claim that, by doing this, her two web pages will end up close in the ordering used for compressing snapshots of the web.

Indeed, consider the ordering used in [5] for compression. It is obtained in two steps. First, one maps each URL to a list of its substrings [top-level domain, second-level domain, ..., first-level directory of the server, second-level directory, ...] (e.g., one maps the URL `a.b.com/d/e` to the list “com,b,a,d,e”). Finally, the lists of strings are sorted lexicographically, obtaining an ordering of the web pages.

Obviously the positions of the owner’s new page, and of the prototype, will be close to each other since the two URLs share all the domain strings, and a large path of the server path string.

In our model, we can show the following:

- (1) The fraction of nodes of indegree i is asymptotic to $\Theta(i^{-2-\frac{1}{k-1}})$; this power law is often referred to as “the rich get richer.”

(2) The fraction of edges of length ℓ in the given embedding is asymptotic to $\Theta(\ell^{-1-\frac{1}{k}})$; analogously, we refer to this as “the long get longer” (the *length* of an edge $x \rightarrow y$ is the absolute difference between the positions of node x and y in the given embedding.)

To obtain their compression rate Boldi and Vigna [5] exploit the distribution of *gaps* in the web graph. Consider the embedding on the line that is obtained by sorting the web pages of a given snapshot lexicographically by URLs. If a web page $x = z_0$ has edges to z_1, \dots, z_j in this order, the gaps are given by $|z_{i-1} - z_i|$, $1 \leq i \leq j$. They observe that the gap distribution in real web graph snapshots follows a power law with exponent ≈ 1.3 . Our model can capture a similar distribution for the edge *lengths* (in our example the length of edge $x \rightarrow z_i$ is $|x - z_i|$), by choosing k appropriately. In fact, both the average edge length and the average gap in our model are small and intuitively, though not immediately, this leads to the compressibility result of Section 7. Qualitatively, a power law with exponent greater than 1 means that most links are short, and hence can be compressed efficiently, while long links, though not impossible, are sufficiently rare. It turns out that a power law distribution of either lengths or gaps (with exponent > 1) is sufficient to show compressibility. For sake of simplicity of the model, we focus on the former in Section 6.

5 The rich get richer

In this section we characterize the indegree distribution of our graph model. We show that the expected indegree distribution follows a power law. We then show the distribution is tightly concentrated. Let

$$f(i) = \frac{k 2^{1+\frac{2}{k-1}} \Gamma\left(\frac{3}{2} + \frac{1}{k-1}\right)}{(k-1)\sqrt{\pi}} \cdot \frac{\Gamma\left(i+1 + \frac{1}{k-1}\right)}{\Gamma\left(i+3 + \frac{2}{k-1}\right)}.$$

It follows that $\lim_{i \rightarrow \infty} f(i) / \left(\frac{k 2^{1+\frac{2}{k-1}} \Gamma\left(\frac{3}{2} + \frac{1}{k-1}\right)}{(k-1)\sqrt{\pi}} \cdot i^{-2-\frac{1}{k-1}} \right) = 1$, i.e., $f(i) = \Theta(i^{-2-\frac{1}{k-1}})$. Let X_i^t denote the number of nodes of indegree i at time t . We first show that $\mathbb{E}[X_i^t]$ can be bounded by $f(i) \cdot t \pm c$, for some constant c .

Theorem 13. *There is a constant $c = c(G_{t_0})$ such that*

$$f(i) \cdot t - c \leq \mathbb{E}[X_i^t] \leq f(i) \cdot t + c, \quad (2)$$

for all $t \geq t_0$ and $i \in [t]$.

Proof. For now, assume $t > t_0$. Let x be the new node, and let y be the corresponding prototype. Recall that y is chosen u.a.r. First, focus on the case $i = 0$, i.e., consider the nodes of indegree 0. We have

$$\mathbb{E}[X_0^t | X_0^{t-1}] = X_0^{t-1} - \Pr[y \text{ had indegree } 0] + 1,$$

for at each time step a new node (i.e., x) of indegree 0 is added and the only node that can change its indegree to 1 is y . The probability of the latter event is exactly $X_0^{t-1}/(t-1)$. By the linearity of expectation,

$$\mathbb{E}[X_0^t] = \left(1 - \frac{1}{t-1}\right) \mathbb{E}[X_0^{t-1}] + 1. \quad (3)$$

Next, consider $i \geq 1$. According to our model, nodes z_1, \dots, z_{k-1} , will be chosen without replacement from $\Gamma(y)$, the successors of y . The successors of the new node x will then be $\Gamma(x) = \{y, z_1, \dots, z_{k-1}\}$. Since z_1, \dots, z_{k-1} are all distinct, the graph remains simple and $|\Gamma(x)| = k$.

For each $j = 1, \dots, k-1$, the node z_j is chosen with probability proportional to its indegree; this follows since node z_j was the endpoint of an edge chosen u.a.r. (all nodes have the same outdegree and to choose an edge, we first select the prototype uniformly at random, i.e., the tail of the edge, and then discard one of the prototype's outgoing edges uniformly at random). The probability that a particular node of indegree $i \geq 1$ is chosen as a successor is $\frac{1}{t-1} + \frac{i(k-1)}{k(t-1)}$ (recall that all the k successors of x will be distinct). Thus, for $i \geq 1$,

$$\mathbb{E}[X_i^t] = \left(1 - \frac{1}{t-1} - \frac{i}{t-1} \frac{k-1}{k}\right) \mathbb{E}[X_i^{t-1}] + \left(\frac{1}{t-1} + \frac{i-1}{t-1} \frac{k-1}{k}\right) \mathbb{E}[X_{i-1}^{t-1}]. \quad (4)$$

For the base cases, note that $X_i^t = 0$ for each $t \geq t_0$. Also, the variables $X_i^{t_0}$ are completely determined by G_{t_0} . For each fixed k , we have $f(t) = \Theta(t^{-2-\frac{1}{k-1}})$. Thus, there is a constant c_0 such that for any $c \geq c_0$, and for all $t \geq t_0$, $\mathbb{E}[X_i^t]$ satisfies (1). The base cases $\mathbb{E}[X_i^{t_0}]$, $i = 1, 2, \dots$, can also be covered with a sufficiently large c (that has to be greater than some function of the initial graph G_{t_0}).

For the inductive case, we have $f(0) = \frac{1}{2}$ (by applying $\Gamma(x)\Gamma(x + \frac{1}{2}) = \Gamma(2x) 2^{1-2x} \sqrt{\pi}$, and $\Gamma(2x + 1) = 2x \Gamma(2x)$, with $x = 1 + \frac{1}{k-1}$). Using this and Equation 3, it follows after some calculations that if X_0^{t-1} satisfies Equation 2, then X_0^t also satisfies it. For $i \geq 1$, we have $f(i-1) = f(i) \cdot (ik - i + 2k + 2)/(ik - i + 1)$. An induction on (4) completes the proof. \square

Thus, in expectation, the indegrees follow a power law with exponent $-2 - 1/(k-1)$. We now show a $O(1)$ -Lipschitz property for the r.v.'s X_i^t for $k = O(1)$. The concentration immediately follows using Theorem 1.

Lemma 14. *Each r.v. X_i^t satisfies the $(2k)$ -Lipschitz property.*

Proof. Our model can be interpreted as the following stochastic process: at step t , two independent dice, with $t-1$ and k faces respectively, are thrown. Let Q_t and R_t be the respective outcomes of these two trials. The new node x will position itself to the immediate left of the node y that was added at time Q_t . Suppose that the (ordered) list of successors of y is (z_1, \dots, z_k) . The ordered list of successors of x will be composed of y followed by the nodes z_1, \dots, z_k with the exception of node z_{R_t} . Thus, the number of nodes X_i^τ of indegree i at time τ can be interpreted as a function of the trials $(Q_1, R_1), \dots, (Q_\tau, R_\tau)$.

We want to show that changing the outcome of any single trial $(Q_{t'}, R_{t'})$, changes the r.v. X_i^τ (for fixed i) by an amount not greater than $2k$. Suppose we change $(q_{t'}, r_{t'})$ to $(q'_{t'}, r'_{t'})$, going from graph G to G' . Let x be the node added at time t' with the choice $(q_{t'}, r_{t'})$, and x' be the node added with the choice $(q'_{t'}, r'_{t'})$.

Let S, S' be the successors of x in G and x' in G' , respectively. The proof is complete by showing inductively that at any time step t , and for any nodes y, y' added at the same time respectively in G, G' , the (ordered) list of successors of y and y' are *close*, i.e., in each of their positions, they either have the same successor, or they have two different elements of $S \cup S'$.

If $t \leq t'$, then the proof is immediate. For $t > t'$, it follows that the only edges we need to consider are the copied edges. By induction, we know that at time $t-1$, the lists of successors of the node we are copying from, in G and G' , were close. Since the two lists are sorted, either the i th copied edges in G and G' will be the same, or they will both point to nodes in $S \cup S'$. Thus the lists of the time t node are close and the proof is complete. \square

6 The long get longer

In this section we analyze the distribution of the edge lengths for the graphs generated by our model, showing that it follows a power law with exponent more than 1. This is the crucial ingredient to establish high

compressibility, i.e., on average constant many bits per edge suffice. Let

$$g(\ell) = \frac{\Gamma(\ell + 1 - \frac{1}{k})}{\Gamma(2 - \frac{1}{k}) \Gamma(\ell + 2)}.$$

It holds that $\lim_{\ell \rightarrow \infty} g(\ell) / \left(\ell^{-1 - \frac{1}{k}} / \Gamma(2 - \frac{1}{k}) \right) = 1$, i.e., $g(\ell) = \Theta(\ell^{-1 - \frac{1}{k}})$. Recall that the *length* of an edge from a node in position i to a node in position j is equal to $|i - j|$; we define its *circular directed length*, denoted as *cd-length*, to be $j - i$ if $j > i$, and $t - (i - j)$ otherwise. Let Y_ℓ^t be the number of edges of length ℓ at time t . We aim to show that $Y_\ell^t \approx g(\ell) \cdot t$. It turns out to be useful to consider a related r.v. Z_ℓ^t that denotes the number of edges of cd-length ℓ at time t . We will first show that, w.h.p., $Z_\ell^t \approx g(\ell) \cdot t$. We will then argue that Y_ℓ^t is very close to Z_ℓ^t .

Theorem 15. *There exists some constant $c = c(G_{t_0})$ such that*

$$g(\ell) \cdot t - c \leq \mathbb{E}[Z_\ell^t] \leq g(\ell) \cdot t + c,$$

for all $t \geq t_0$ and $\ell \in [t]$.

Proof. As in the proof of Theorem 13, we begin by obtaining a recurrence for the r.v.'s Z_ℓ^t . Let x be the node added at time t , and let y, y' be the nodes to the immediate right and left of x respectively (where y' equals the last node in the ordering if x is placed before the first node y).

Consider Z_1^t . For $t > t_0$,

$$\mathbb{E}[Z_1^t | Z_1^{t-1}] = Z_1^{t-1} - \Pr[x \text{ lengthens an edge of cd-length 1}] + 1,$$

because an edge $x \rightarrow y$ of length 1 is necessarily added to the graph, and adding x can lengthen at most one edge of cd-length 1 (i.e., the edge $y' \rightarrow y$ if it exists). The probability of the latter event is equal to $Z_1^{t-1}/(t-1)$. By the linearity of expectation,

$$\mathbb{E}[Z_1^t] = \left(1 - \frac{1}{t-1}\right) \mathbb{E}[Z_1^{t-1}] + 1.$$

Now consider Z_ℓ^t , for $\ell \geq 2$ and $t > t_0$. We have,

$$\begin{aligned} \mathbb{E}[Z_\ell^t | Z_\ell^{t-1}, Z_{\ell-1}^{t-1}] &= Z_\ell^{t-1} - \mathbb{E}[\# \text{ of edges of cd-length } \ell \text{ that } x \text{ lengthened} | Z_\ell^{t-1}, Z_{\ell-1}^{t-1}] \\ &\quad + \mathbb{E}[\# \text{ of edges of cd-length } (\ell - 1) \text{ that } x \text{ lengthened} | Z_\ell^{t-1}, Z_{\ell-1}^{t-1}] \\ &\quad + \mathbb{E}[\# \text{ of edges of cd-length } (\ell - 1) \text{ that } x \text{ copied from } y | Z_\ell^{t-1}, Z_{\ell-1}^{t-1}]. \end{aligned}$$

Recall that x is placed to the left of a prototype node y chosen u.a.r. Thus, given a fixed edge of length ℓ , the probability that this edge is lengthened by x is $\ell/(t-1)$. Thus,

$$\mathbb{E}[\# \text{ of edges of length } \ell \text{ that } x \text{ lengthened} | Z_\ell^{t-1}, Z_{\ell-1}^{t-1}] = \frac{\ell}{t-1} Z_\ell^{t-1},$$

and

$$\mathbb{E}[\# \text{ of edges of length } (\ell - 1) \text{ that } x \text{ lengthened} | Z_\ell^{t-1}, Z_{\ell-1}^{t-1}] = \frac{\ell - 1}{t - 1} Z_{\ell-1}^{t-1},$$

so that

$$\begin{aligned} & \mathbb{E} [\# \text{ of edges of cd-length } (\ell - 1) \text{ that } x \text{ copied from } y \mid Z_\ell^{t-1}, Z_{\ell-1}^{t-1}] \\ &= \sum_{j=1}^{k-1} \Pr [\text{the } j\text{th copied edge had cd-length } (\ell - 1) \mid Z_\ell^{t-1}, Z_{\ell-1}^{t-1}]. \end{aligned}$$

Note that, for each $j = 1, \dots, k-1$, the j th copied edge is chosen uniformly at random over all the edges (even though the $k-1$ edges are not copied independent). Thus,

$$\sum_{j=1}^{k-1} \Pr [\text{the } j\text{th copied edge had cd-length } (\ell - 1) \mid Z_\ell^{t-1}, Z_{\ell-1}^{t-1}] = \frac{(k-1)Z_{\ell-1}^{t-1}}{k(t-1)}.$$

By the linearity of expectation, we get for $\ell \geq 2$,

$$\mathbb{E} [Z_\ell^t] = \left(1 - \frac{\ell}{t-1}\right) \mathbb{E} [Z_\ell^{t-1}] + \left(\frac{\ell-1}{t-1} + \frac{1}{t-1} \frac{k-1}{k}\right) \mathbb{E} [Z_{\ell-1}^{t-1}].$$

The base cases can be handled as in Theorem 13. The inductive step for $\ell = 1$ can be directly verified. For $\ell \geq 2$, it suffices to note that $g(\ell-1) = k \cdot (\ell+1)/(\ell k - 1) \cdot g(\ell)$. \square

Thus, the expectation of the edge lengths follows a power law with exponent $-1 - 1/k$. To establish the concentration result, we need to analyze quite closely the combinatorial structure of the graphs generated by our model. Recall that the nodes in our graphs are placed contiguously on a discrete line (or list). At a generic time step, we use x_i to refer to the i th node in the left-to-right ordering. Given an ordering $\pi = (x_1, x_2, \dots, x_t)$ of the nodes, and an integer $0 \leq k < t$, a k -rotation, $\rho_k(x_i)$ maps the generic node x_i , $1 \leq i \leq t$, to position $1 + ((i+k) \bmod t)$.

We say that two nodes x, x' are *consecutive* if there exists a k such that $|\rho_k(x) - \rho_k(x')| = 1$, i.e., they are consecutive if in the ordering they are either adjacent or one is the first and the other is the last. Further, we say that an edge $x'' \rightarrow x'''$ *passes over* a node x if there exists k such that $\rho_k(x'') < \rho_k(x) < \rho_k(x''')$. Finally, two edges $x \rightarrow x'$ and $x'' \rightarrow x'''$ are said to *cross* if there exists a k such that after a k -rotation exactly one of x and x' lies between positions $\rho_k(x'')$ and $\rho_k(x''')$. We prove the following characterization that will be used later in the analysis.

Lemma 16. *At any time, given any two consecutive nodes x, x' , and any positive integer ℓ , the number of edges of cd-length ℓ that pass over x or x' (or both) is at most $C = (k+2)t_0 + 1$.*

Proof. Let us define G_t^- as the graph G_t minus the edges incident to the nodes that were originally in G_{t_0} . Note that for each cd-length ℓ , the number of edges of cd-length ℓ that we remove is upper-bounded by $2t_0$ as each node can be incident to at most two edges of cd-length ℓ , one going in, and one going out of the node. Unless otherwise noted, we will consider G_t^- for the rest of the proof.

Fix a time t , and consider any rotation ρ . Let x_1, \dots, x_t be the nodes in the left-to-right ordering given by the rotation (i.e., node x_i is in position i according to ρ). For a set of edges of the same cd-length to pass over at least one of two consecutive nodes x, x' it is necessary for every pair of them to cross. We will bound, for a generic edge e , the number of edges that cross e and have the same length as e . Let $t(x_a)$ be the time when x_a was added to the graph. First, by definition we have that if $x_a \rightarrow x_b$, then $t(x_a) > t(x_b)$.

Second, we claim that if there exists a rotation ρ' for which there are three nodes x_a, x_b, x_c such that $\rho'(x_a) < \rho'(x_b) < \rho'(x_c)$ and $t(x_c) > t(x_b)$, then the edge $x_a \rightarrow x_c$ cannot exist. To see this, note that for $x_a \rightarrow x_c$ to exist it must be that $t(x_a) > t(x_c)$. We want to show inductively that a node pointing to x_c in

the ordering implied by ρ' is either to the left of x_c or to the right of x_b . Note that x_c was not in G_{t_0} since its insertion time is greater than that of x_b . Thus, each node placed to the immediate left of x_c will point to it, and will satisfy the induction hypothesis. Furthermore, each node that copies an edge to x_c must be placed to the immediate left of a node pointing to x_c . Thus, the second claim is proved.

Third, we claim that if x_a, x_b, x_c, x_d are four nodes such that the edges $x_a \rightarrow x_c$ and $x_b \rightarrow x_d$ exist, and *cross* each other, then there exists an edge $x_c \rightarrow x_d$. To see this, first note that none of these four nodes could have been part of G_{t_0} , for otherwise at least one of the two edges could not have been part of G_t^- . Fix a rotation ρ'' s.t. $\rho''(x_a) < \rho''(x_b) < \rho''(x_c)$; by the second claim, it must be that $t(x_b) > t(x_c)$. Thus, the edge $x_b \rightarrow x_d$ has necessarily been copied from some node, say x_{b_1} . Note that $\rho''(x_{b_1}) \leq \rho''(x_c)$. Indeed by the assumption $\rho''(x_c) > \rho''(x_b)$ and it is impossible that $\rho''(x_c) < \rho''(x_{b_1})$, for otherwise x_b could not have copied from x_{b_1} as $t(x_b) > t(x_c)$. Now, we know that the edge $x_{b_1} \rightarrow x_d$ exists (as before, x_{b_1} is not part of G_{t_0}). If $x_{b_1} = x_c$, then we are done. Otherwise, there must exist an x_{b_2} pointing to x_d from which x_{b_1} has copied the edge. Note that $\rho''(x_{b_1}) < \rho''(x_{b_2}) \leq \rho''(x_c)$. By iterating this reasoning, the claim follows.

Take any set S of edges having the same length and such that any pair of them *cross*. Given an arbitrary ρ''' , let x be the node with the smallest $\rho'''(x)$ such that, for some x' , the edge $x \rightarrow x'$ is in S (the nodes x and x' are unique). For any other edge $y \rightarrow y'$ in S , by the third claim, there must exist the edge $x' \rightarrow y'$. As x' has outdegree k , it follows that $|S| \leq k + 1$.

Finally, since the seed graph G_{t_0} had $k \cdot t_0$ edges and we removed at most $2t_0$ edges of cd-length ℓ (for an arbitrary $\ell \geq 1$) in the cut $[G_{t_0}, G_t \setminus G_{t_0}]$, we have not counted at most $k \cdot t_0 + 2t_0$ edges of length ℓ passing over one of the nodes x, x' . The proof follows. \square

Now we prove the $O(1)$ -Lipschitz property of the r.v.'s Z_ℓ^t , provided that $t_0, k = O(1)$. The concentration of the Z_ℓ^t will follow from Theorem 1.

Lemma 17. *Each r.v. Z_ℓ^t satisfies the $((k + 2)t_0 + k + 1)$ -Lipschitz property.*

Proof. We use the stochastic interpretation of the model as in the proof of Lemma 14. For each τ , let Z_ℓ^τ be the r.v. representing the number of edges of cd-length ℓ at time τ . We consider Y_ℓ^τ as a function of the trials $(Q_1, R_1), \dots, (Q_\tau, R_\tau)$. We show that changing the outcome of any single trial $(Q_{t'}, R_{t'})$, changes the r.v. Z_ℓ^τ , for fixed ℓ , by an amount not greater than $C + k = (k + 2)t_0 + k + 1$.

Suppose we change $(q_{t'}, r_{t'})$ to $(q'_{t'}, r'_{t'})$, going from graph G to G' . Let x be the node added at time t' with the choice $(q_{t'}, r_{t'})$, and x' be its equivalent with the choice $(q'_{t'}, r'_{t'})$. We show that choosing two different positions for x and x' can change the number of edges of cd-length ℓ by at most $C + k$ at any time step. Note that before time step t' , the cd-lengths are all equal.

By Lemma 16, at time $t > t'$ and for all ℓ , the number of edges of cd-length ℓ that pass over x (resp., x') is at most C . For an edge e , let S_e be the set of edges that have been copied from e , directly or indirectly, including e itself, i.e., $e \in S_e$ and if an edge e' is copied from some edge in S_e , then $e' \in S_e$. Note that no two edges in S_e have the same cd-length, since they all start from different nodes, but end up at the same node.

For any node z , if e_1, \dots, e_k are the successors of z , we define $S_z = S_{e_1} \cup \dots \cup S_{e_k}$. The last observation implies that, for any fixed ℓ , no more than k edges of cd-length ℓ are in S_v (or $S_{v'}$) at any single time step. Now, consider the following edge bijection from G to G' : the i th edge of the j th inserted node in G is mapped to the i th edge of the j th inserted node in G' . It follows that if an edge e in G (resp., G') does not pass over x (resp., x') and is not in S_x (resp., $S_{x'}$), then e is mapped to an edge of the same cd-length in G' (resp., G). Thus, the difference in the number of edges of the cd-length ℓ in G and G' is at most $C + k$. \square

We now show that the number D_t of edges whose length and cd-length are different (at time t) is very small. Since the maximum absolute difference between Y_ℓ^t and Z_ℓ^t is bounded by D_t , this will show that these r.v.'s

are close to each other. First note that if an edge $x_i \rightarrow x_j$ has different length and cd-length, then $j < i$; call such an edge *left-directed* and let R_t be the set of left-directed edges. Since $D_t \leq R_t$, it suffices to bound the latter.

Lemma 18. *With probability $1 - O\left(\frac{1}{t^2}\right)$, $R_t = O\left(t^{1-\frac{1}{k}+\epsilon}\right)$, for each constant $\epsilon > 0$.*

Proof. Observe that each edge $x_i \rightarrow x_j$ counted by R_t is such that $j < i$. Thus, R_{t_0} is equal to the number of left-directed edges in G_{t_0} with its given embedding.

Further, R_t 's increase over R_{t-1} equals the number of left-directed edges copied at step t (the proximity edge is always not left-directed).

Thus, $E[R_t|R_{t-1}] = \left(1 + (k-1) \cdot \frac{1}{k(t-1)}\right) \cdot R_{t-1}$ and $E[R_t] = \left(1 + (k-1) \cdot \frac{1}{k(t-1)}\right) \cdot E[R_{t-1}]$, for each $t > t_0$. Therefore,

$$E[R_t] = R_{t_0} \cdot \prod_{i=t_0+1}^t \left(1 + \frac{k-1}{k} \cdot \frac{1}{i}\right) = R_{t_0} \cdot \prod_{i=t_0+1}^t \frac{i + \frac{k-1}{k}}{i} = R_{t_0} \cdot \frac{\Gamma\left(t + \frac{k-1}{k} + 1\right) \cdot \Gamma(t_0 + 1)}{\Gamma\left(t_0 + \frac{k-1}{k} + 1\right) \cdot \Gamma(t + 1)}.$$

Thus, $E[R_t] = \Theta\left(t^{1-\frac{1}{k}}\right)$. We note how an $O(1)$ -Lipschitz condition holds: at most $k-1$ new left-directed edges can be added at each step. Thus Theorem 1 can be applied with an error term of $O(\sqrt{t \log t}) = O\left(t^{\frac{1}{2}+\epsilon}\right) \subseteq O\left(t^{1-\frac{1}{k}+\epsilon}\right)$. The result follows. \square

By applying Theorem 1, Theorem 15, Lemma 17, and Lemma 18, we obtain the following.

Corollary 19. *With probability at least $1 - O\left(\frac{1}{t^2}\right)$, it holds that*

- i. $E[Z_\ell^t] - O(\sqrt{t \log t}) \leq Z_\ell^t \leq E[Z_\ell^t] + O(\sqrt{t \log t})$, and
- ii. $E[Z_\ell^t] - O\left(t^{1-1/k+\epsilon}\right) \leq Y_\ell^t \leq E[Z_\ell^t] + O\left(t^{1-1/k+\epsilon}\right)$.

Note that the concentration error term, $O(\sqrt{t \log t})$, is at most R_t , for each $k \geq 2$. Also, the corollary is vacuous if $\ell > t^{1/(k+2)}$.

7 Compressibility of our model

We now analyze the number of bits needed to compress the graphs generated by our model. Recall that the web graph has a natural embedding on the line via the URL ordering that experimentally gives very good compression [5, 6]. Our model generates web-like random graphs and an embedding ‘‘à-la-URL’’ on the line. We work with the following *BV-like compression scheme*: a node at position p on the line stores its list of successors at positions p_1, \dots, p_k as a list $(p_1 - p, \dots, p_k - p)$ of compressed integers. An integer $i \neq 0$ will be compressed using $O(\log(|i| + 1))$ bits, using the Elias γ -code (see, for instance, [36]). We show that our graphs can be compressed using $O(1)$ bits per link using the above scheme.

Theorem 20. *The above BV-like scheme compresses the graphs generated by our model using $O(n)$ bits, with probability at least $1 - O\left(\frac{1}{n}\right)$.*

Proof. Let $\epsilon > 0$ be a small constant. At time n , consider the number of edges of length at most $L = \lceil n^\epsilon \rceil$. Note that by Corollary 19, for each $1 \leq \ell \leq L$, we have that $|Y_\ell^n - E[Z_\ell^n]| = O\left(n^{1-1/k+\epsilon}\right)$, with probability $1 - O\left(n^{-1}\right)$. For the rest of the proof, we implicitly condition on these events.

By using the lower bound for $E[Z_\ell^n]$ given by Theorem 15, we obtain the following lower bound on the number of edges of length at most L . By Lemma 2,

$$\frac{1}{\Gamma(2 - \frac{1}{k})} \sum_{\ell=1}^L \frac{\Gamma(\ell + 1 - \frac{1}{k})}{\Gamma(\ell + 2)} = k \cdot \left(1 - \frac{\Gamma(L + 2 - \frac{1}{k})}{\Gamma(L + 2)\Gamma(2 - \frac{1}{k})} \right)$$

and thus,

$$\begin{aligned} S &\geq \sum_{\ell=1}^L \left(\frac{\Gamma(\ell + 1 - \frac{1}{k})}{\Gamma(2 - \frac{1}{k})\Gamma(\ell + 2)} \cdot n - c - O(n^{1-1/k+\epsilon}) \right) \\ &\geq nk \left(1 - \frac{\Gamma(L + 2 - \frac{1}{k})}{\Gamma(L + 2)\Gamma(2 - \frac{1}{k})} \right) - O(L \cdot n^{1-1/k+\epsilon}) \\ &\geq nk - O(n \cdot k \cdot L^{-1/k}) - O(L \cdot n^{1-1/k+\epsilon}) \geq nk - O(n^{1-\epsilon_1}), \end{aligned}$$

where ϵ_1 is a small constant.

At time n , the total number of edges of the graph is nk . Thus the number of edges of length more than L is at most $O(n^{1-\epsilon_1})$.⁶ The maximum edge length is $O(n)$ and so each edge can be compressed in $O(\log n)$ bits. The overall contribution, in terms of bits, of the edges longer than L will then be $o(n)$.

Now, we calculate the contribution B of the edges of length at most L .

$$\begin{aligned} B &\leq \sum_{\ell=1}^L \left(O(\log(\ell + 1)) \left(\frac{\Gamma(\ell + 1 - \frac{1}{k})}{\Gamma(2 + \frac{1}{k})\Gamma(\ell + 2)} n + c + O(n^{1-1/k+\epsilon}) \right) \right) \\ &\leq n \cdot O \left(\sum_{\ell=1}^L \log(\ell + 1) \cdot \ell^{-1-1/k} \right) + O(L \cdot n^{1-1/k+\epsilon} \cdot \log L) \\ &= O(n), \end{aligned}$$

where the penultimate inequality follows since the term $\frac{\Gamma(\dots)}{\Gamma(\dots)\Gamma(\dots)}$ is $O(\ell^{-1-1/k})$, and the last inequality from the fact that $O(\ell^{-1-2\epsilon} \cdot \log \ell) = O(\ell^{-1-\epsilon})$ and the convergence of the Riemann series. The claim follows. \square

Thus, given an ordering of nodes, we can compress the graph using $O(1)$ bits per edge on average using a linear-time algorithm. We then ask if it is still possible to compress this graph *without* knowing the ordering. We show that this is still possible.

Theorem 21. *The graphs generated by our model can be compressed using $O(n)$ bits in linear time, even if the ordering of the nodes is not available.*

Proof. Given a node v in G , if we look at its two-neighborhood, we can either find an out-neighbor w of v with exactly $k - 1$ out-neighbors in common with v or conclude that v was part of the “seed” graph G_{t_0} (having constant order). This follows because, if v were not part of G_{t_0} , during its insertion, v added a proximity edge to its “real prototype” w , and copied $k - 1$ of w ’s outlinks. If more than one out-neighbor of v has $k - 1$ out-neighbors in common with v , we choose one arbitrarily and call it the “possible prototype” of v . Note that this step can be implemented in $O(k^2) = O(1)$ time.

⁶Notice that, for this argument to work, it is crucial to have a very strong bound on the behavior of the Y_ℓ^n random variables.

To compress the graph, we create an unlabeled rooted forest out of the nodes in G_{t_0} . A node v will look for a possible prototype w . If such a w is found, then v will choose w as its parent. Otherwise v will be a root in the forest. Then, to represent G it suffices to (i) represent the unlabeled rooted forest, (ii) represent the subgraph induced by the roots of the trees in the forest, and (iii) for each non-root node v in the forest, use $\lceil \log k \rceil$ bits to represent which of its parent's out-neighbors was not copied by v in G . The forest can be described with $O(n)$ bits, for instance, by writing down the *down / up* steps made when visiting each tree in the forest, disregarding edge orientations (as each edge is directed from the child to the parent). The graph induced by the roots of the trees (i.e., a subgraph of G_{t_0}) can be stored in a non-compressed way using $O(t_0^2) = O(1)$ bits. The third part of the encoding will require at most $O(n \log k) = O(n)$ bits. Note that it is possible to compute each of the three encodings in linear time. \square

8 Other properties of our model

In this section we prove some additional properties of our model: that it has a large number of bipartite cliques, high clustering coefficient, and small undirected diameter. We also extend our model to accommodate power law outdegree distributions.

8.1 Bipartite cliques

Recall that a *bipartite clique* $K(a, b)$ is a set A of a nodes and a set B of b nodes such that each node in A has an edge to every node in B . We can show that the graphs generated by our model contain a large number of bipartite cliques. The proof is similar to the one of [23] for the linear growth model.

Theorem 22. *There exists a $\beta > 0$, such that the number of bipartite cliques $K(\Omega(\log n), k)$ in our model is $\Omega(n^\beta)$, w.h.p. .*

Proof. Take any fixed node x of the seed graph G_{t_0} and a subset S of $k - 1$ of its successors. Divide the time steps $t - t_0$ into disjoint epochs of exponentially increasing size, i.e., of sizes $c\tau, c^2\tau, c^3\tau, \dots$, for a large enough τ . Let j be the number of epochs; then, $j = \Omega(\log n)$. Note that for $i \leq j$, the probability that at least one node added in epoch i will attach itself to x and copy exactly the edges in S is at least a constant; also, for each $i \neq i'$, these events are independent. Thus, w.h.p., at least $\Omega(\log n)$ nodes will be *good*, i.e., will have $S \cup \{v\}$ as successors.

Now, any subset of the good nodes will form a bipartite clique with $S \cup \{v\}$. The number of subsets of size $\Omega(\log n)$ is easily shown to grow as $\Omega(n^\beta)$ for some $\beta > 0$. \square

8.2 Clustering coefficient

Watts and Strogatz [35] introduced the concept of clustering coefficient. The *clustering coefficient* $C(x)$ of a node x is the ratio of the number of edges between neighbors of x and the maximum possible number of such edges, i.e., $\frac{1}{2} \deg(x)(\deg(x) - 1)$ in the undirected case and $\deg(x)(\deg(x) - 1)$ in the directed case. The clustering coefficient $C(G)$ of a (simple) graph G is the average clustering coefficient of its nodes. Snapshots of the real web graph have been observed to possess a pretty high clustering coefficient. Thus, having a high clustering coefficient is a desirable property of a web graph model.

Theorem 23. *Let G be a (directed) graph generated by our model. The clustering coefficient of G is $\Theta(1)$ w.h.p. .*

Proof. By Theorem 13 and Lemma 14, there are $q = \Theta(n)$ nodes of indegree 0, w.h.p; these nodes have been inserted after time t_0 . Take any such node x and let y be its prototype node. Then, x and y share $k - 1$ out-neighbors (the “copied” ones). The total degree of x is k , thus the clustering coefficient of x is at least $\frac{k-1}{k(k-1)} = \frac{1}{k} \in \Omega(1)$. The clustering coefficient of the graph is the average of the clustering coefficients of its nodes; thus, in our case, it is at least $\frac{1}{n} \cdot q \cdot \frac{1}{k} \geq \Omega(1)$. The proof follows since 1 is the maximum clustering coefficient. \square

The previous proof also shows that, if we remove orientations from the edges of the graphs generated by our model, the clustering coefficient of the resulting undirected graphs is $\Theta(1)$ w.h.p. .

8.3 Undirected diameter

We now argue that, w.h.p., the undirected diameter of our random graphs is $O(\log n)$ (provided that the seed graph G_{t_0} was weakly connected). By undirected diameter, we mean the diameter of the undirected graph obtained by removing edge orientations from our graphs. Note that our graphs are almost DAGs, i.e., they are DAGs perhaps except for the nodes in the seed graph G_{t_0} and therefore the directed diameter is not an interesting notion in our case.

Consider the so-called *random recursive trees*: the process starts with a single node and at each step, a node is chosen uniformly at random and a new leaf is added as a child of that node; the process ends at the generic time n . Consider the “proximity” edges added in step (ii) in our model, i.e., those added from the new node, to a node chosen uniformly at random. Now, these edges induce a random recursive forest with t_0 different roots corresponding to the nodes of the seed graph G_{t_0} . A result of [33] states that the height of a random recursive tree on n nodes is $O(\log n)$ w.h.p. . Thus, assuming that G_{t_0} is weakly connected implies that the (undirected) diameter of our graphs is $O(\log n)$ w.h.p. .

8.4 Power law outdegree

We now observe how it is possible to obtain a power law outdegree by modifying our model slightly. Let $G = (V, E)$ be a graph generated according to our original model. Fix $0 \leq p \leq 1$ and initialize $E' \leftarrow E$. Then, for each node $v \in V$, independently with probability p , update $E' \leftarrow E' \cup \{(v, u) \mid \exists u \text{ s.t. } (u, v) \in E\}$, i.e., independently with probability p , each node u adds an edge to all v 's such that (u, v) is an edge in the original model. (This action of reciprocal linking is a common practice in web page creation.) Let $G' = (V, E')$. Observe that, if $p = 0$, then $G' = G$. And, if $p = 1$, then every edge of G gets reversed.

We will show that for each constant $0 < p \leq 1$, the outdegree distribution follows a power law. First, we bound the maximum indegree in G .

Lemma 24. *For each $\epsilon > 0$, with probability at least $1 - O\left(\frac{1}{n}\right)$, the maximum indegree in G is $O\left(n^{1 - \frac{1}{k} + \epsilon}\right)$.*

Proof. Let D_j^t be the indegree at time t of a node that had degree d at time $j \leq t$. We have $D_j^j = d$. Moreover, for each $t > j$, we have

$$E[D_j^{t+1}] = E[D_j^t] \left(1 + \frac{k-1}{kt}\right) + \frac{1}{t}.$$

It is easy to check that the following is the only solution to the recurrence relation for $t \geq j$:

$$E[D_j^t] = \frac{k}{k-1} \left(\left(d + 1 - \frac{d}{k}\right) \cdot \frac{\Gamma(j)}{\Gamma\left(j + 1 - \frac{1}{k}\right)} \cdot \frac{\Gamma\left(t + 1 - \frac{1}{k}\right)}{\Gamma(t)} - 1 \right).$$

It follows from the monotonicity of $\Gamma(x)$ that

$$E[D_j^t] \leq \frac{k}{k-1} \left(\left(d+1 - \frac{d}{k} \right) \cdot \frac{\Gamma\left(t+1 - \frac{1}{k}\right)}{\Gamma(t)} - 1 \right) = O\left(t^{1-\frac{1}{k}}\right).$$

Since all the nodes in G_{t_0} have constant degree and every new nodes start from degree 0, an easy application of Theorem 1 completes the proof. \square

We now prove that both the indegree and the outdegree distributions of G' follow a power law.

Lemma 25. *For each constant $p > 0$ there exists a constant $c > 0$ such that with probability $1 - O\left(\frac{1}{n}\right)$, for each $i < O(n^c)$, the number of nodes with indegree (resp., outdegree) i in G' is $\Theta\left(n \cdot i^{-2-\frac{1}{k-1}}\right)$.*

Proof. Let v be any node in G that was not in G_{t_0} and let $d(v)$ be the indegree of v in G . Note that the outdegree of v in G is k . Let $d_i(v)$ and $d_o(v)$ be, respectively, the indegree and the outdegree of v in G' .

First, we have that $d_o(v) = k + d(v)$ with probability p and $d_o(v) = k$ with probability $1 - p$. Also, the $d_o(v)$'s are mutually independent. Therefore, if X_i^n is the number of nodes of indegree i in G , then by Theorem 1, we have that the number of nodes of outdegree $i + k$ in G' is tightly concentrated around $p \cdot X_i^n$. The outdegree of G' then follows a power law since the indegree of G follows a power law (Lemmas 13 and 14).

Next, consider the indegree of a node v in G' : $d_i(v)$ will be equal to one of $d(v), d(v) + 1, \dots, d(v) + k$. Therefore, the number X'_i of nodes having in degree $i \geq k$ can be upper bounded by $X_i^n + X_{i-1}^n + \dots + X_{i-k}^n$. By Lemmas 13 and 14, the latter upper bound on X'_i is concentrated around $n \cdot O(f(i)) = n \cdot \Theta\left(i^{-2-\frac{1}{k-1}}\right)$.

Next, observe that we can lower bound X'_i by the number of nodes that (i) had in degree i in G , and such that (ii) none of its outlinks got reversed in the final step of the process. The probability that (ii) happens for a node that satisfies (i) is at least $(1 - p)^k = \Theta(1)$. Therefore the expected value of the lower bound is $n \cdot \Omega(f(i))$. The concentration of the value of the lower bound can be proved via Theorem 1. Indeed, let the original graph G be given. By Lemma 24, with high probability, the maximum indegree in G will be $O\left(n^{1-\frac{1}{k}+\epsilon}\right)$. For $v \in G$, let R_v be the random variable with value 1 if v adds its reversed links and 0 otherwise. The number of nodes with indegree i in G and that satisfy (ii) is a function of the R_v 's. This function satisfies the $O\left(n^{1-\frac{1}{k}+\epsilon}\right)$ -Lipschitz condition. Concentration is thus guaranteed by Theorem 1.

In summary, for some $c > 0$, with high probability and for each $i \in O(n^c)$, the number of nodes having indegree (resp., outdegree) i can be upper bounded by $n \cdot O(f(i))$ and lower bounded by $n \cdot \Omega(f(i))$. Since $f(i) = \Theta\left(i^{-2-\frac{1}{k-1}}\right)$, the proof is complete. \square

It is easy to show that all the properties that we proved for our original model (including compressibility) hold even in this modified model.

References

- [1] M. Adler and M. Mitzenmacher. Towards compressing web graphs. In *Data Compression Conference*, pages 203–212, 2001.
- [2] W. Aiello, F. Chung, and L. Lu. Random evolution in massive graphs. In *Proc. 42nd IEEE Symposium on Foundations of Computer Science*, pages 510–519, 2001.

- [3] R. Albert and A.-L. Barabasi. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.
- [4] P. Boldi, M. Santini, and S. Vigna. Permuting web graphs. In *Proc. of the 6th International Workshop on Algorithms and Models for the Web-Graph*, pages 116–126, 2009.
- [5] P. Boldi and S. Vigna. The webgraph framework I: Compression techniques. In *Proc. 13th International World Wide Web Conference*, pages 595–601, 2004.
- [6] P. Boldi and S. Vigna. Codes for the world-wide web. *Internet Mathematics*, 2(4):405–427, 2005.
- [7] B. Bollobás, O. Riordan, J. Spencer, and G. E. Tusnády. The degree sequence of a scale-free random graph process. *Random Structures and Algorithms*, 18(3):279–290, 2001.
- [8] C. Borgs, J. T. Chayes, C. Daskalakis, and S. Roch. First to market is not everything: An analysis of preferential attachment with fitness. In *Proc. 39th Annual ACM Symposium on Theory of Computing*, pages 135–144, 2007.
- [9] A. Z. Broder, S. R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener. Graph structure in the web. *Computer Networks*, 33:309–320, 2000.
- [10] G. Buehrer and K. Chellapilla. A scalable pattern mining approach to web graph compression with communities. In *Proc. 1st International Conference on Web Search and Data Mining*, pages 95–106, 2008.
- [11] J. Carlson and J. Doyle. Highly optimized tolerance: A mechanism for power laws in designed systems. *Phys. Rev. E*, 60:1412, 1999.
- [12] F. Chierichetti, R. Kumar, S. Lattanzi, A. Panconesi, and P. Raghavan. Models for the compressible web. In *Proc. 50th Symposium on Foundations of Computer Science*, pages 331–340, 2009.
- [13] F. Chierichetti, R. Kumar, M. Mitzenmacher, A. Panconesi, P. Raghavan, and S. Lattanzi. On compressing social networks. In *Proc. 15th Conference on Knowledge Discovery and Data Mining*, pages 219–228, 2009.
- [14] C. Cooper and A. M. Frieze. A general model of web graphs. *Random Structures and Algorithms*, 22(3):311–335, 2003.
- [15] C. Cooper and A. M. Frieze. The cover time of the preferential attachment graph. *Journal of Combinatorial Theory, Ser. B*, 97(2):269–290, 2007.
- [16] M. Dodds and Watts. An experimental study of search in global social networks. *Science*, 301(5634):827–829, 2003.
- [17] D. Dubhashi and A. Panconesi. *Concentration of Measure for the Analysis of Randomized Algorithms*. Cambridge University Press, 2009.
- [18] A. Fabrikant, E. Koutsoupias, and C. H. Papadimitriou. Heuristically optimized trade-offs: A new paradigm for power laws in the internet. In *Proc. 29th International Colloquium on Automata, Languages and Programming*, pages 110–122, 2002.
- [19] C. Karande, K. Chellapilla, and R. Andersen. Speeding up algorithms on compressed web graphs. In *Proc. 2nd International Conference on Web Search and Data Mining*, pages 272–281, 2009.

- [20] J. Kleinberg. Navigation in a small world. *Nature*, 406:845, 2000.
- [21] J. Kleinberg. The small-world phenomenon: An algorithmic perspective. In *Proc. 37th Annual ACM Symposium on Theory of Computing*, pages 163–170, 2000.
- [22] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal. Stochastic models for the web graph. In *Proc. 41st IEEE Symposium on Foundations of Computer Science*, pages 57–65, 2000.
- [23] R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins. Trawling the web for emerging cybercommunities. In *Proc. 8th International World Wide Web Conference*, pages 403–416, 1999.
- [24] S. Lattanzi and D. Sivakumar. Affiliation networks. In *Proc. 41st ACM Symposium on Theory of Computing*, pages 427–434, 2009.
- [25] J. Leskovec, D. Chakrabarti, J. Kleinberg, and C. Faloutsos. Realistic, mathematically tractable graph generation and evolution, using kronecker multiplication. In *European Conference on Principles and Practice of Knowledge Discovery in Databases*, pages 133–145, 2005.
- [26] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graphs over time: Densification laws, shrinking diameters. *ACM Transactions on Knowledge Discovery from Data*, 1:1–41, 2007.
- [27] D. Liben-Nowell, J. Novak, R. Kumar, P. Raghavan, and A. Tomkins. Geographic routing in social networks. *Proc. National Academy of Sciences*, 102(33):11623–11628, 2005.
- [28] B. Mandelbrot. An informational theory of the statistical structure of languages. In W. Jackson, editor, *Communication Theory*, pages 486–502. Butterworth, 1953.
- [29] M. Mitzenmacher. A brief history of generative models for power law and lognormal distributions. *Internet Mathematics*, 1(2), 2003.
- [30] L. Mutafchiev. The largest tree in certain models of random forests. *Random Structures and Algorithms*, 13(3-4):211–228, 1998.
- [31] H. Simon. On a class of skew distribution functions. *Biometrika*, 42:425–440, 1955.
- [32] T. Suel and J. Yuan. Compressing the graph structure of the web. In *Data Compression Conference*, pages 213–222, 2001.
- [33] J. Szymanski. On the complexity of algorithms on recursive trees. *Theoretical Computer Science*, 74(3):355–361, 1990.
- [34] J. Travers and S. Milgram. An experimental study of the small world problem. *Sociometry*, 32(4):425–443, 1969.
- [35] D. Watts and S. Strogatz. Collective dynamics of ‘small-world’ networks. *Nature*, 393:409–410, 1998.
- [36] I. H. Witten, A. Moffat, and T. C. Bell. *Managing Gigabytes: Compressing and Indexing Documents and Images*. Morgan Kauffman Publishers, 1999.
- [37] G. K. Zipf. *Human Behavior and the Principle of Least Effort*. Addison-Wesley, 1949.