Isomorphism and Embedding Problems for Infinite Limits of Scale-Free Graphs

Robert D. Kleinberg * Jon M. Kleinberg †

Abstract

The study of random graphs has traditionally been dominated by the closely-related models $\mathcal{G}(n, m)$, in which a graph is sampled from the uniform distribution on graphs with n vertices and m edges, and $\mathcal{G}(n, p)$, in which each of the $\binom{n}{2}$ edges is sampled independently with probability p. Recently, however, there has been considerable interest in alternate random graph models designed to more closely approximate the properties of complex real-world networks such as the Web graph, the Internet, and large social networks. Two of the most well-studied of these are the closely related "preferential attachment" and "copying" models, in which vertices arrive one-by-one in sequence and attach at random in "rich-get-richer" fashion to d earlier vertices.

Here we study the infinite limits of the preferential attachment process — namely, the asymptotic behavior of finite graphs produced by preferential attachment (briefly, *PA graphs*), as well as the infinite graphs obtained by continuing the process indefinitely. We are guided in part by a striking result of Erdős and Rényi on countable graphs produced by the infinite analogue of the $\mathcal{G}(n, p)$ model, showing that any two graphs produced by this model are isomorphic with probability 1; it is natural to ask whether a comparable result holds for the preferential attachment process.

We find, somewhat surprisingly, that the answer depends critically on the out-degree d of the model. For $d = 1$ and $d = 2$, there exist infinite graphs R_d^∞ such that a random graph generated according to the infinite preferential attachment process is isomorphic to R_d^∞ with probability 1. For $d \geq 3$, on the other hand, two different samples generated from the infinite preferential attachment process are non-isomorphic with positive probability. The main technical ingredients underlying this result have fundamental implications for the structure of finite PA graphs; in particular, we give a characterization of the graphs H for which the expected number of subgraph embeddings of H in an n-node PA graph remains bounded as n goes to infinity.

1 Introduction

For decades, the study of random graphs has been dominated by the closely-related models $\mathcal{G}(n, m)$, in which a graph is sampled from the uniform distribution on graphs with n vertices and m edges, and $\mathcal{G}(n, p)$, in which each of the $\binom{n}{2}$ edges is sampled independently with probability p. The first was introduced by Erdős and Rényi in [16], the second by Gilbert in [19]. While these random graphs have remained a central object of study and continue to have many important applications in combinatorics and theoretical computer science, recently there has also been a great deal of interest in alternative random graph models whose properties more closely resemble those of complex real-world networks such as the Web graph, the Internet, and large social networks. Two of the most well-studied of these are the closely related "preferential attachment" and "copying" models; the former was introduced by Barabási and Albert in [3] and subsequently formalized by Bollobás and Riordan in [8], while the latter was introduced by Kumar et al. in [22].

A random graph in the preferential attachment model (henceforth, the PA model) is built up one vertex at a time, with each new vertex v linking to the preceding ones by d new edges, where the out-degree d is a parameter of the model. Roughly, the head of each edge emanating from v is chosen by sampling from the preceding vertices with probabilities weighted according to their total degree (in-degree plus out-degree); this is the *preferential*, or "rich-get-richer," aspect of the model, since nodes of higher in-degree attract new incoming edges more readily. (We will sometimes use the term "PA graph" as an informal shorthand to refer to a random graph drawn from the distribution defined by the PA model.) As we discuss further below, there has been considerable work aimed at determining fundamental graph-theoretic properties in the PA model, exposing both similarities and contrasts

*Department of Mathematics, MIT, Cambridge MA 02139. Email: rdk@math.mit.edu. Supported by a Fannie and John Hertz Foundation Fellowship.

†Department of Computer Science, Cornell University, Ithaca NY 14853. Email: kleinber@cs.cornell.edu. Supported in part by a David and Lucile Packard Foundation Fellowship and NSF grants 0081334 and 0311333.

with the classical $\mathcal{G}(n,p)$ model.

In the present paper, we seek to understand the infinite limits of the PA model — namely, the asymptotic behavior of graphs produced by this model as the number of nodes goes to infinity, and the distribution $\mathcal{PA}_d^\infty$ on random graphs with countably many vertices obtained by continuing the PA process indefinitely. We were inspired by the following classical theorem about the "infinite version" of the $\mathcal{G}(n,p)$ model [17].

Theorem 1.1. *Let $\mathcal{G}(\infty,p)$ denote the probability distribution on graphs with vertex set $\mathbb{N}$, in which each edge (i,j) is included independently with probability p. (Here p is any constant in $(0,1)$.) There exists an infinite graph R, such that a random sample from $\mathcal{G}(\infty,p)$ is isomorphic to R with probability 1.*

When one first encounters this theorem, it can seem quite startling: infinite random graphs are not "random" at all; they are almost surely isomorphic to a single fixed graph R. A rich theory has developed around the infinite model $\mathcal{G}(\infty,p)$, with connections reaching into mathematical logic, algebra, and a number of other areas (see e.g. [13]).

On the other hand, essentially nothing is known about the the infinite version of the PA model. Does something analogous to Theorem 1.1 hold here as well, or is the situation fundamentally different? At a more fine-grained level, we are also interested in understanding what can be said about the local structure of finite graphs produced by the PA model as the number of nodes goes to infinity. As we discuss further below, the only prior work addressing the infinite graphs generated by such processes, as far as we are aware, are some interesting recent papers by Bonato and Janssen [11, 12], which proposed the notion of studying infinite limits of random graph evolution processes related to the copying model of [23]. These papers consider the relationship between such infinite random graphs and certain *deterministic* adjacency axioms. Some of these axioms have a unique infinite model up to isomorphism, while others are satisfied with probability 1 by the infinite limits of the random graph processes considered in these papers. However, none of their theorems resolve the question of whether an analogue of Theorem 1.1 holds for such infinite random graphs.

Our first result is the following, where again $\mathcal{PA}_d^\infty$ denotes the distribution associated with the infinite PA model.

Theorem 1.2. *For $d = 1,2$, there is a graph R_d^∞ such that a random sample from $\mathcal{PA}_d^\infty$ is isomorphic to R_d^∞ with probability 1.*

For $d = 1$ this is clear, since the outcome of the random process will almost surely be a tree with countably many nodes, in which each node has infinite degree. For the case of out-degree $d = 2$, the resulting graph R_2^∞ is much more complicated. Its structure can be characterized axiomatically, but it is also possible to give explicit constructions of graphs isomorphic to R_2^∞. For example, it is isomorphic to the graph whose vertices consist of all finite rooted binary trees with integer labels, where the vertex corresponding to a labeled tree T has edges to its left sub-tree and to its right sub-tree.

The global structure of the proof for the case $d = 2$ is a standard "back-and-forth" argument, which will be familiar to readers acquainted with Theorem 1.1. The key step, however — establishing that there is an adequate supply of vertices to sustain the back-and-forth construction of the isomorphism — is much more complicated than in the classical case, since the PA process introduces difficult conditioning problems.

One might imagine that for the cases of out-degrees $d = 3,4,5,\ldots$ one could establish isomorphisms with probability 1 to increasingly complex graphs R_3^∞, R_4^∞, and so on. But in fact, we have the following result.

Theorem 1.3. *For each out-degree $d \geq 3$, it is not the case that two independent random samples from $\mathcal{PA}_d^\infty$ are isomorphic with probability 1.*

This contrast between the cases of $d = 2$ and $d \geq 3$ comes to us as something of a surprise, since it does not have an obvious analogue in the prior work on graphs generated according to the PA process. There, typically, the out-degree d has a clear quantitative effect on the underlying graph parameters, but not a qualitative effect of this sort.

This contrasting pair of results is a particularly succinct consequence of one of the main technical components of the paper, which addresses a fundamental structural issue for both the finite and infinite versions of the PA model — a characterization of the graphs H for which the expected number of subgraph embeddings of H in an n-node PA graph remains bounded as n goes to infinity. Phrased equivalently as a statement about the infinite model $\mathcal{PA}_d^\infty$, we show that if a finite graph H is equal to its 3-core (i.e. the union of all subgraphs of H of minimum degree 3), then the number of subgraph embeddings of H in a random sample from $\mathcal{PA}_d^\infty$ has a positive finite expectation, while if H is not equal to its 3-core, then the number of embeddings is almost surely either zero or infinite.

The existence and relative abundance of small subgraphs is a topic of considerable interest for both empirical studies of real networks and for theoretical studies of their models (see e.g. [20, 23]). Our characterization theorem has a natural interpretation in this context, as a precise statement about the lack of dense local structure in PA graphs G. First, any graph H of minimum degree 3 appears a bounded number of times in expectation as a subgraph of G, independent of the size of G. Second, any graph $H = (V, E)$ for which $|E|/|V| > 2$ has a non-trivial 3-core, and so our result implies that in any PA graph G, there exists a set of nodes S in G of bounded expected size, such that any embedded copy of H in G includes at least one node from S. (In other words, S serves as a bounded set of "attachment points" for copies of H.)

This characterization theorem for subgraph embeddings yields the non-isomorphism theorem for $\mathcal{PA}_d^\infty$ with $d \geq 3$ fairly directly; it also has the following further consequence for finite PA graphs. (Here the distribution on n-vertex graphs produced by the PA process will be denoted by $\mathcal{PA}_d^{(n)}$.)

THEOREM 1.4. *For $d \geq 3$, there exist first-order graph properties which do not satisfy a zero-one law for $\mathcal{PA}_d^{(n)}$, i.e. there is a first-order formula $\phi(G)$ such that*

$$0 < \lim_{n\to\infty} \Pr_{G \leftarrow \mathcal{PA}_d^{(n)}}(\phi(G)) < 1.$$

This contrasts with the situation for $\mathcal{G}(n, p)$, where it is known that every first-order formula satisfies a zero-one law. (For a very interesting and deep analysis of first-order properties of $\mathcal{G}(n, p)$ when p is a function of n, we refer the reader to [28].)

Finally, it is worth briefly returning to the original motivation for these types of models — the complex structures of graphs such as the Web, the Internet, and large social networks. The (finite) PA and copying models are of course stylized abstractions designed to capture some of the observed properties of these networks; they were not intended as faithful representations of the complexities of the true structures. Our study of infinite analogues here follows a theme that is common in a number of areas, to try gaining insight into extremely large finite systems by modeling them as infinite — as, for example, when working with infinite lattice structures in physics, or with a continuum of agents in economics. Thus far, aside from the work of Bonato and Janssen [11, 12], this has not really been attempted for complex networks, but the results about finite structures that emerge from the study of infinite limits of the graph generation process here provide a suggestion for the kinds of results one can obtain from this style of investigation, and we feel there is clearly room for further study in this direction.

1.1 Relation to prior work

The preferential-attachment model of random graphs was introduced by Barabási and Albert in [3], motivated in part by the goal of explaining the power-law degree distribution observed in the Internet topology by Faloutsos et al [18] and in the Web topology by Kumar et al [21]. Barabási and Albert's original paper contained a heuristic argument establishing a power law for the degree distribution of random preferential-attachment graphs; rigorous mathematical proofs of this result subsequently appeared in [1, 10]. An alternative random graph model with power-law degree distribution, the "evolving copying model," was independently proposed and analyzed by Kumar et al [22, 23], with the aim of modeling the Web graph. Cooper and Frieze introduced a model which simultaneously generalizes these two random graph models, and again proved that the degree distribution obeys a power law [14]. A directed version of the preferential-attachment model was introduced and studied by Bollobas et al in [6], who again established a power-law distribution both for the in-degrees and the out-degrees.

In addition to their degree distribution, many other properties of preferential-attachment random graphs have been rigorously analyzed; these include their diameter [8], conductance [25], eigenvalues [24, 15], "clustering coefficient" [7], and "robustness" under random vertex deletions [9]. See [2, 7, 26] for various surveys of work in this area, focusing on different research communities.

As discussed above, the only other work to our knowledge that addresses the infinite graphs which arise as the limit of such processes is [11, 12]. In [11], Bonato and Janssen formulate a copying model, similar to that proposed in [23], and they show that an infinite random graph generated according to this process satisfies a certain deterministic adjacency property which they label "Property (B)." They then study various model-theoretic and combinatorial properties of graphs satisfying property (B) and its generalizations. Of particular relevance, for our purposes, is their theorem that there are $2^{\aleph_0}$ many non-isomorphic graphs satisfying property (B). While this suggests the possibility that random samples from their copying model are not almost surely isomorphic, the authors explicitly refrain from addressing this question since their focus is on studying infinite graphs satisfying the deterministic property (B) and its generalizations, regardless of whether such graphs were generated by a random process or not.

The subsequent paper [12], written independently and concurrently with our work, generalizes the random graph process introduced in [11] and relates it to some new adjacency properties (ARO, near-ARO, local near-ARO, n-near-ARO). Only the ARO property has a unique infinite model up to isomorphism; in fact, the other properties are shown to be satisfied by $2^{\aleph_0}$ many non-isomorphic graphs. Moreover, the infinite random graphs considered in [12] have a positive probability of failing to satisfy the near-ARO property. Again, this suggests the possibility that random samples from this generalized copying model are not almost surely isomorphic, but again the authors refrain from answering this question, as they leave open the possibility that the infinite graphs generated by their random process are almost surely isomorphic to a single infinite graph which fails to satisfy the near-ARO property.

2 Definitions

We begin by defining, for each $d > 0$, a random graph process $\mathcal{PA}_d$ on graphs with vertex set $\{0, 1, \ldots\}$. $\mathcal{PA}_d$ is a probability distribution on sequences of connected undirected graphs, $G_0 \subset G_1 \subset \ldots$, where G_t has vertex set $\{0, 1, \ldots, t\}$. Our definition is closely modeled on the definition of the graph process $(G_m^t)_{t \geq 0}$ in [8]; however, it differs in some technical details because we want our graphs to be connected and theirs are potentially disconnected. Graphs in their model are allowed to have self-loops, and a new connected component is created every time a new vertex appears and connects to no vertices other than itself. Our graphs will have parallel edges but no self-loops, and they will be connected.

The graph process $\mathcal{PA}_d$ is defined recursively as follows. G_0 has one vertex (labeled 0) and no edges. G_{t+1} is obtained from G_t by adding a new vertex (labeled $t+1$) and joining it to vertices $0, 1, \ldots, t$ with d random edges, sampled independently at random from a probability distribution (the "preferential attachment" distribution) specified as follows:

$$\Pr(e = (t+1, s)) = d_t(s)/2dt,$$

where $d_t(s)$ denotes the degree of vertex s in G_t. In other words, each neighbor of $t+1$ is chosen according to a distribution which weights vertices by their current degree. The definition of the preferential attachment distribution makes no sense in the case $t = 0$, since G_0 has no edges. Accordingly, we stipulate that vertex 1 always links to vertex 0 with d parallel edges.

Given a sample $G_0 \subset G_1 \subset \ldots$ from $\mathcal{PA}_d$, let $G_\infty = \bigcup_{t=0}^{\infty} G_t$ and define $\mathcal{PA}_d^\infty$ to be the resulting probability distribution on graphs with vertex set $\{0, 1, 2, \ldots\}$. The edges of G_∞ may be numbered $1, 2, \ldots$, such that the edges of $G_{t+1} \setminus G_t$ are labeled $dt+1, dt+2, \ldots, dt+d$. An equivalent way of specifying the graph process $\mathcal{PA}_d$ would have been to say that edge $dt + j$ ($1 \leq j \leq d$) chooses an edge uniformly at random from the set $\{1, 2, \ldots, dt\}$, chooses an endpoint of this edge uniformly at random, and joins vertex $t+1$ to the chosen endpoint.

Although $\mathcal{PA}_d^\infty$ was defined as a probability distribution on undirected graphs, the edges of these graphs come equipped with a natural orientation, directed from the higher-numbered endpoint to the lower-numbered one. We will sometimes consider the graphs G_∞ as directed graphs, and it will be clear when we are doing so. The advantage of adopting this dual viewpoint on the graphs G_∞ is that it enables us to state stronger theorems: our isomorphism theorem holds for *directed graphs* sampled from $\mathcal{PA}_2^\infty$ and trivially implies the corresponding result for undirected graphs; while our non-isomorphism theorem holds for *undirected graphs* and trivially implies the corresponding result for directed graphs.

3 Growth rate of vertex degrees

The proofs in this paper hinge on a detailed understanding of the growth rate of vertex degrees, i.e. the asymptotics of $d_t(i)$ as a function of t, in a typical sequence $G_0 \subset G_1 \subset \ldots$ sampled from $\mathcal{PA}_d$. It has been known since the introduction of the Barabási-Albert model that $\mathbf{E}[d_t(i)] = \theta(\sqrt{t})$ for any fixed i. A non-rigorous argument using differential equations appears in [3], and a rigorous proof may be found in [8]. A key ingredient in our proof of Theorem 1.2 is the following stronger fact:

Proposition 3.1. *For any fixed vertex i, with probability 1, $\lim_{t\to\infty} d_t(i)/\sqrt{t}$ exists and is positive.*

Although the calculations arising in the proof are very similar to those used in establishing the asymptotics of $\mathbf{E}[d_t(i)]$ [8], we require two more techniques from martingale theory to establish a stronger result: the existence and positivity of the limit $\lim_{t\to\infty} d_t(i)/\sqrt{t}$. The existence of the limit is established using Doob's martingale convergence theorem, and its positivity comes from the Kolmogorov-Doob inequality combined with a second-moment computation. (See [5], Chapter 35, for an introduction to martingales including both of the aforementioned tools.) The calculations arising in these proofs are very similar to those used in Lemma 2 of [8], in which the authors prove (among other things) that $\mathbf{E}[d_t(i)] = \theta(\sqrt{t})$.

Fix a vertex i, and consider how its degree changes at time $t+1$. Each of the d new edges attaches to i with probability $d_t(i)/2dt$, so

$$\mathbf{E}(d_{t+1}(i) \parallel d_t(i)) = d_t(i) + d\left(\frac{d_t(i)}{2dt}\right) = \left(1 + \frac{1}{2t}\right) d_t(i).$$

It follows that the sequence of random variables $(d_t(i))_{t \geq i}$ may be transformed into a martingale by rescaling, as follows. Define

$$c_t = \prod_{j=1}^{t-1}\left(1+\frac{1}{2j}\right) \qquad\qquad X_t = d_t(i)/c_t.$$

Now the sequence $(X_t)_{t\geq i}$ is a martingale (adapted to the σ-field $\mathcal{F}_t$ generated by the random variable G_t) since:

$$\begin{aligned}\mathbf{E}[X_{t+1} \parallel \mathcal{F}_t] &= \frac{1}{c_{t+1}}\mathbf{E}[d_{t+1}(i) \parallel \mathcal{F}_t] \\ &= \frac{1}{c_{t+1}}\left(1+\frac{1}{2t}\right)d_t(i) \\ &= \frac{1}{c_t}d_t(i) \\ &= X_t.\end{aligned}$$

The constant c_t is $\theta(\sqrt{t})$, as may be seen easily by taking the logarithm of both sides of the formula defining c_t, and using the identity

$$x - \frac{1}{2}x^2 < \log(1+x) < x.$$

The following two theorems are instrumental in the proof of Proposition 3.1. Proofs may be found in [5], or in most books on stochastic processes.

Theorem 3.1. (Doob's Martingale Convergence Theorem) *Let $X_1, X_2, \ldots$ be a submartingale. If $K = \sup_n \mathbf{E}(|X_n|) < \infty$, then $X_n \to X$ with probability 1, where X is a random variable satisfying $\mathbf{E}\left[|X|\right] \leq K$.*

Theorem 3.2. (Kolmogorov-Doob inequality) *If $X_1, \ldots, X_n$ is a submartingale, then for $\alpha > 0$,*

$$\Pr\left[\max_{i\leq n} X_i \geq \alpha\right] \leq \frac{1}{\alpha}\mathbf{E}\left[|X_n|\right].$$

Proof of Proposition 3.1. The random variables X_t are non-negative, so $\mathbf{E}(|X_t|) = \mathbf{E}(X_t) = \mathbf{E}(X_i) = d/c_i$ for all t. This establishes that the X_t satisfy the hypotheses of Doob's Martingale Convergence Theorem, so with probability 1 they approach a finite limit as $t \to \infty$. Given that $c_t = \theta(\sqrt{t})$, this implies that $\lim_{t\to\infty} d_t(i)/\sqrt{t}$ exists almost surely.

It remains to show that the limit is almost surely positive. The idea of the proof is simple, and conceptually similar to Zeno's Paradox of the Race Course [4]. We will show that after the degree of i exceeds some threshold, the value of X_t is very unlikely to drop by a factor of 2 from its current value. In order for $\lim_{t\to\infty} X_t$ to be zero, it must be the case that X_t decreases by a factor of two infinitely often, an event having probability 0. then you that you To make this notion precise, define a sequence of times $n_0 < n_1 < \ldots$ as follows. Let $n_0 = i$. Let n_1 be the smallest value of n such that $X_n < (1/2)X_{n_0}$, or ∞ if no such n exists. Continue defining $n_2, n_3, \ldots$ in the same manner, i.e. n_{j+1} is the smallest n such that $X_n < (1/2)X_{n_j}$, or ∞ if no such n exists or if $n_j = \infty$. We will prove that $\Pr(\text{all } n_j \text{ are finite}) = 0$, and to do so it is sufficient to prove that $\Pr(n_{j+1} < \infty \| n_j) < 1-\delta$ for some constant $\delta > 0$.

To do so, we use the Kolmogorov-Doob inequality applied to the submartingale $\tilde{X}_n = (X_{n_j} - X_n)^2$ ($n \geq n_j$). (Any convex function applied to a martingale yields a submartingale, by Jensen's inequality.) An estimate for $\mathbf{E}(\tilde{X}_n \parallel X_{n_j})$ is computed in Supplementary Section A. The result is:

$$\mathbf{E}(\tilde{X}_n \| X_{n_j}) < (C/\sqrt{n_j})X_{n_j}.$$

for some constant C. Now, by the Kolmogorov-Doob inequality,

$$\begin{aligned}&\Pr(\max_{n\geq n_j} \tilde{X}_n > (X_{n_j}/2)^2 \| X_{n_j}) \\ &= \lim_{N\to\infty} \Pr(\max_{n_j\leq n\leq N} \tilde{X}_n > (X_{n_j}/2)^2 \| X_{n_j}) \\ &\leq 4X_{n_j}^{-2} \lim_{N\to\infty} \mathbf{E}(\tilde{X}_N \| X_{n_j}) \\ &\leq 4X_{n_j}^{-2}(C/\sqrt{n_j})X_{n_j} \\ &= \frac{4C}{\sqrt{n_j}X_{n_j}} \\ &\leq C'/d_{n_j}\end{aligned}$$

for some constant C'. Recall that our goal is to show that $\Pr(n_{j+1} = \infty \| n_j) > \delta$ — or, equivalently, that $\Pr(\max_{n\geq n_j} \tilde{X}_n \leq (X_{n_j}/2)^2 \| n_j) > \delta$ — for some constant $\delta > 0$. We now see that this could be accomplished by establishing that $C'/d_{n_j} \leq 1-\delta$. So to finish, it suffices to prove that d_{n_j} grows unboundedly large as $j \to \infty$. (In fact, it would suffice to prove that d_{n_j} is eventually greater than $C'/(1-\delta)$.) But this is easy: the probability that $d_n = Y$ for all $n > N$ is bounded above by $\prod_{n=N}^{\infty}\left(1 - \frac{Y}{2dn}\right) = 0$, so with probability 1, $d_n \to \infty$ as $n \to \infty$.

4 An isomorphism theorem for $d = 1, 2$

We begin with the easy proof of Theorem 1.2 in the case $d = 1$. Let R_1^∞ denote a countable rooted arborescence in which each non-root vertex has infinite indegree and has a path to the root.

Theorem 4.1. *A random sample from $\mathcal{PA}_1^\infty$ is almost surely isomorphic to R_1^∞ as a directed graph.*

Proof. Let G_∞ be a random sample from $\mathcal{PA}_1^\infty$. By construction, every vertex except for 0 has outdegree 1, and vertex 0 has outdegree 0. With probability 1, the indegree of each vertex is infinite, by Proposition 3.1. By construction, each vertex except for 0 has a path to vertex 0. These properties uniquely determine the isomorphism type of G_∞ as a directed graph.

For the rest of this section, we focus on the case $d = 2$. Consider the following three axioms for an infinite directed graph K with countable vertex set.

1. There exists a vertex v_0 with outdegree 0. Every other vertex has outdegree 2.
2. For any pair of (not necessarily distinct) vertices v, w, there are infinitely many vertices whose two outgoing edges link to v and w.
3. K does not contain any infinite forward path.

PROPOSITION 4.1. *Any two countable directed graphs K_1, K_2 satisfying axioms (1)-(3) are isomorphic.*

Proof. Let $v_0, v_1, \ldots$ be the vertices of K_1, ordered so that all of the outgoing edges from v_j link to vertices in the set $\{v_0, \ldots, v_{j-1}\}$. Such an ordering may be constructed recursively as follows. Choose v_0 to be the vertex with outdegree 0. Given $v_0, \ldots, v_{j-1}$, start from an arbitrary vertex of $K_1 \setminus \{v_0, \ldots, v_{j-1}\}$ and follow outgoing edges until a vertex v_j with outdegree 0 is reached; this must happen after a finite number of steps, since otherwise K_1 would contain a cycle or an infinite forward path. Similarly, let $w_0, w_1, \ldots$ be the vertices of K_2, ordered so that all of the outgoing edges from w_j link to vertices in the set $w_0, \ldots, w_{j-1}$.

The proof now proceeds by a back-and-forth argument. We will construct an isomorphism $\phi : K_1 \to K_2$ by first selecting $\phi(v_0)$, then $\phi^{-1}(w_0)$, then $\phi(v_1)$, then $\phi^{-1}(w_1)$, and so on ad infinitum, until a one-to-one correspondence between $V(K_1)$ and $V(K_2)$ has been defined. The steps in which we select $\phi(v_i)$ will be called *forward steps*, those in which we select $\phi^{-1}(w_i)$ are *reverse steps.*

To start, set $\phi(v_0) = w_0$. The construction now proceeds in a series of steps, each of which starts with a one-to-one correspondence between finite subsets $S_1 \subset V(K_1), S_2 \subset V(K_2)$ inducing an isomorphism between the corresponding induced subgraphs, and extends this one-to-one correspondence to include a single additional element of each vertex set, while preserving the fact that it defines an isomorphism of induced subgraphs. The forward steps alternate with the reverse steps. For reasons which will soon become apparent, we add an additional claim into our induction hypothesis: every outgoing edge from a vertex of S_1 (resp. S_2) joins it to another vertex of S_1 (resp. S_2). This is satisfied vacuously in the base case where $S_1 = \{v_0\}, S_2 = \{0\}$.

To perform a forward step, take the lowest-numbered vertex v_j in $V(K_1) \setminus S_1$. This vertex has two outgoing edges pointing to vertices $v_{i_1}, v_{i_2} \in V(K_1)$. We have $i_1, i_2 < j$, so both v_{i_1} and v_{i_2} belong to S_1. Now choose $\phi(v_j)$ to be any vertex $w \in V(K_2) \setminus S_2$ such that w points to $\phi(v_{i_1})$ and $\phi(v_{i_2})$. (Such a vertex is guaranteed to exist, by Axiom 2.) It is now easy to check that the induction hypothesis is still satisfied. By construction, ϕ maps the outgoing edges from v_j to the outgoing edges from w. As for the incoming edges, ϕ is only defined at this stage as a mapping from $S_1 \cup \{v_j\}$ to $S_2 \cup \{w\}$, and neither v_j nor w have any incoming edges from vertices in these sets. (This is where we needed the additional fact that every outgoing edge from a vertex of S_1 (S_2) joins it to another vertex of S_1 (S_2). It is trivial to check that this fact remains true after extending S_1, S_2 to include v, i, respectively.)

This completes the proof of the induction hypothesis in the case of a forward step. By symmetry, the induction hypothesis is proved for reverse steps as well.

Any countable directed graph satisfying axioms (1)-(3) will be denoted by R_2^∞. Theorem 1.2 now follows from the following more precise result.

THEOREM 4.2. *If G_∞ is sampled at random from $\mathcal{PA}_2^\infty$, then G_∞ is almost surely isomorphic to R_2^∞.*

Proof Sketch. We must check that G_∞ satisfies axioms (1)-(3) almost surely. By construction, there is a single vertex with outdegree 0, all other vertices have outdegree 2, and no infinite forward paths exist in G_∞. Finally, we require the following:

PROPOSITION 4.2. *Given any two (not necessarily distinct) vertices j_1, j_2 in $V(G_\infty)$, there are infinitely many $i \in V(G_\infty)$ whose two outgoing edges point to j_1, j_2.*

The proof of this relies on Proposition 3.1. Informally, that proposition guarantees the existence of constants

$$x_1 = \lim_{t\to\infty} d_t(j_1)/\sqrt{t}$$
$$x_2 = \lim_{t\to\infty} d_t(j_2)/\sqrt{t}$$

so for sufficiently large t, the probability that vertex t links to j_1 and j_2 is approximately $(x_1\sqrt{t}/4t)(x_2\sqrt{t}/4t) = x_1x_2/16t$. The probability that no vertex after t_0 links to j_1 and j_2 is approximately $\prod_{t=t_0}^{\infty}\left(1 - \frac{x_1x_2}{16t}\right) = 0$. Thus, almost surely, there exists a vertex $i \in V(G_\infty) \setminus S_2$ whose two outgoing edges point to j_1, j_2.

The biggest problem with making this informal argument rigorous is that, by conditioning on the values of x_1 and x_2, we change the distribution of the random outgoing edges from each vertex; it is no longer the preferential attachment distribution, so we have no justification for our estimate of the probability that t links to j_1 and j_2. While the informal argument gives the correct intuition, the rigorous version is surprisingly intricate; for details, see the full version of this paper. $\square$

Concrete constructions for R_2^∞. Theorem 4.2 supplies an axiomatic characterization of R_2^∞, but unlike Theorem 4.1 it does not concretely specify a graph which is isomorphic, almost surely, to random samples from $\mathcal{PA}_2^\infty$. In this section we present two such constructions.

The first construction produces R_2^∞ as the union of a countable chain of infinite graphs $R_0 \subset R_1 \subset \ldots$, defined recursively as follows. Let R_0 consist of a vertex v_0 of outdegree 0, and countably many other vertices, each with two parallel edges pointing to v_0. Given R_j, construct R_{j+1} as follows: for each pair of (not necessarily distinct) vertices $v, w \in V(R_j)$, adjoin a countable set of new vertices, each with two outgoing edges pointing to v, w. Finally, put

$$R_2^\infty = \bigcup_{j=0}^{\infty} R_j.$$

It is routine to verify that this graph R_2^∞ satisfies the axioms (1)-(3).

The second construction defines R_2^∞ as a graph whose vertex set is a set of labeled binary trees. Specifically, let Σ be a countable alphabet, and let the vertex set $V(R_2^\infty)$ be the set of finite rooted binary trees whose edges are labeled with elements of Σ. If $T \in V(R_2^\infty)$ is a tree with more than one node, then T has two outgoing edges in R_2^∞ pointing to its left and right subtrees. Again, it is straightforward to verify that this definition of R_2^∞ satisfies the axioms (1)-(3).

A model-theoretic characterization of R_2^∞. Our goal in this section is to specify a precise sense in which R_2^∞ is "axiomatically characterized" by the conditions given in Section 4. We will exhibit a first-order theory T, in the language of directed graphs, such that R_2^∞ is a prime model of T. (A model M of a first-order theory T is called *prime* if every other model of T contains a submodel isomorphic to M. If a countable theory has a prime model, this model is unique up to isomorphism.) Interestingly, T has many other countable models which are not isomorphic to R_2^∞. (Note the close thematic links between this section and [11].)

Let T denote the following set of first-order formulas in the language of directed graphs, consisting of one axiom **Outdegree** and two infinite families of axioms $(\mathbf{Acyclic}_n)_{n\geq 2}$ and $(\mathbf{Adjacency}_n)_{n\geq 1}$:

Outdegree: G has one vertex of outdegree 0, and all other vertices have outdegree 2.

$\mathbf{Acyclic}_n$**:** G does not contain an n-cycle.

$\mathbf{Adj}_n$**:** For any pair v, w of vertices of G, there are at least n distinct vertices whose two outgoing edges point to v, w.

PROPOSITION 4.3. *R_2^∞ is a prime model of T.*

Proof. Clearly R_2^∞ is a model of T. Given any other model G, an isomorphic embedding $\phi : R_2^\infty \to G$ is constructed in a manner similar to the back-and-forth argument used in proving Theorem 4.2, except that this time the construction is one-directional since we are not trying to make ϕ surjective. Let $\{v_0, v_1, \ldots\}$ be the vertex set of R_2^∞, numbered so that the edges from v_j point to elements of $\{v_0, \ldots, v_{j-1}\}$ as before. We construct ϕ inductively, by specifying that $\phi(v_0)$ is the unique vertex of G having outdegree zero; and that for $j > 0$, if the two edges from v_j in R_2^∞ point to w_1, w_2, then $\phi(v_j)$ is any vertex of G whose two outgoing edges point to $\phi(w_1), \phi(w_2)$. It is straightforward to verify that ϕ is an isomorphic embedding of R_2^∞ in G.

REMARK 4.1. *The theory T has many countable models which are not isomorphic to R_2^∞. To cite a specific example, let G be the graph whose vertex set $V(G)$ is the set of all countable or finite rooted binary trees whose edges are labeled by natural numbers, such that all but finitely many edges are labeled with the successor of their parent's label. For any such tree T with more than one node, the two outgoing edges from T in G point to its left and right subtrees.*

5 Subgraph embeddings and non-isomorphism theorem for $d \geq 3$

The aim of this section is to characterize, for each finite graph H, the number of embeddings of H in a random sample from $\mathcal{PA}_d^\infty$. The non-isomorphism theorem for $d > 2$, Theorem 1.3, will be derived as an easy corollary.

DEFINITION 5.1. *The* ordered arboricity *of an undirected graph G is the minimum k such that G admits a vertex ordering with the following property: for each vertex $v \in V(G)$, there are at most k edges connecting v to its predecessors. We will denote the ordered arboricity of G by $\eta(G)$.*

If G admits such a vertex ordering, and if we arbitrarily color the edges from each vertex v to its predecessors with distinct colors from a set of $\eta(G)$ colors, then the color classes constitute a partition of G's edge set into $\eta(G)$ acyclic subgraphs, so the ordered arboricity of G is bounded below by the arboricity. The ordered arboricity can be strictly greater than the arboricity, e.g. the edge set of a 4-clique K_4 may be partitioned into two disjoint paths, but $\eta(K_4) = 3$ since the last vertex in any ordering is joined to its precedessors by three edges.

THEOREM 5.1. *For a finite graph H, let $K \subseteq H$ denote the union of all subgraphs of H which have minimum degree 3. If G is a random sample from $\mathcal{PA}_d^\infty$, then:*

1. *If $\eta(H) > d$, there are no embeddings of H in G.*
2. *If $\eta(H) \le d$ and $K = H$ then, with probability 1, the number of embeddings of H in G is finite. In fact, the expected number of embeddings of H in G is finite and positive.*
3. *If $\eta(H) \le d$ and $K \subsetneq H$ then, with probability 1, the number of embeddings of H in G is either zero or infinite.*

Proof. If H is any finite subgraph of G and we order the vertices of H according to their arrival order, then each vertex has at most d edges to its predecessors, which proves that $\eta(H) \le d$ for any finite subgraph of G. Conversely, if H is a finite graph with $\eta(H) \le d$, let us label the vertices of H with the numbers $1, 2, \ldots, |V(H)|$ in such a way that each edge is joined to its predecessors by at most d edges, It is easy to see from the definition of $\mathcal{PA}_d^\infty$ that there is a positive probability the induced subgraph of G on vertex set $\{1, 2, \ldots, |V(H)|\}$ is precisely H (since any edges from $\{1, 2, \ldots, |V(H)|\}$ that don't contribute to the embedding of H can attach to vertex 0 of G).

If $\eta(H) \le d$ and $K = H$, we have already shown that the expected number of embeddings of H in G is positive. The fact that it is finite is contained in the following lemma.

LEMMA 5.1. *If K is a graph of minimum degree 3, then the expected number of embeddings of K in G is finite.*

This lemma forms the crux of the theorem; the complete proof is given in the full version of this paper. The basis of the proof is the observation that, while any *two* vertices in G almost surely have *infinitely* many common neighbors, any *three* vertices in G almost surely have *finitely* many common neighbors. Informally, this is because any three vertices have degrees $\theta(\sqrt{t})$ when vertex t is added, and so it links to all three with probability $\theta(t^{-3/2})$; summing over all t then gives a finite expected value. Making this precise, however, requires dealing with the conditioning on the degrees, which also posed difficulties in the proof of Theorem 4.2. To extend this argument to embeddings of a graph K of minimum degree 3, we "dismantle" K by removing one node at a time, controlling the number of embeddings in this dismantling through a bound composed of monomials over the random variables X_t defined in Section 3. Bounding the expectations of such monomials requires a delicate argument by induction over the set of all monomials, ordered by a "dominance ordering". (For the details of the proof, including the definition of the monomial ordering, we refer the reader to the full version of the paper.)

We now complete the final case in the proof of Theorem 5.1, when $\eta(H) \le d$ and $K \ne H$. We claim that if G contains an embedded copy of K, then the number of embeddings of H in G is infinite with probability 1. The proof is by induction on the number of vertices in $H \setminus K$. By assumption, $K \ne H$ so there exists a vertex $v \in H$ whose degree is less than 3. By the induction hypothesis or by the assumption that K embeds in G, we may assume that $H \setminus \{v\}$ embeds in G. Now by Proposition 4.2, the number of embeddings of H in G is infinite with probability 1 (since this proposition asserts that the event "there are only a finite number of ways to extend the embedding of $H \setminus \{v\}$" has unconditional probability 0, and here we're conditioning on a positive-probability event). Note that this establishes case 3, and concludes the proof of the Theorem.

A non-isomorphism theorem for $d > 2$. The non-isomorphism theorem for $d \ge 3$ (Theorem 1.3) follows easily from Theorem 5.1. Let $N_0 > 0$ be the expected number of distinct embeddings of K_4 in G, choose $N > N_0$. and let K denote the graph consisting of N disjoint copies of K_4. We now consider the probability that G contains a copy of K as a subgraph. Theorem 5.1 asserts that the expected number of copies of K is positive, and hence there is a positive probability that G contains a copy of K. On the other hand, since $N > N_0$, Markov's inequality ensures that the probability of finding N distinct embeddings of K_4 is less than 1, and this implies that the probability that G contains a copy of K is less than 1.

Thus the property "G contains K as a subgraph" is an isomorphism-invariant property, whose truth value has a positive probability of distinguishing two independent random samples from $\mathcal{PA}_d^\infty$.

Proof of Theorem 1.4. The graph property specified in the previous paragraph is expressible by a first-order formula $\phi(G)$. We claim now that

$$0 < \lim_{n\to\infty} \Pr_{G \leftarrow \mathcal{PA}_d^{(n)}} (\phi(G)) < 1.$$

The limit is greater than zero for the same reason as before: Theorem 5.1 implies that G contains K as a subgraph with positive probability. It is also easy to see that $\Pr_{G\leftarrow\mathcal{PA}_d^\infty}(\phi(G)) \geq \Pr_{G\leftarrow\mathcal{PA}_d^{(n)}}(\phi(G))$, since a random sample from $\mathcal{PA}_d^\infty$ is the union of a chain of graphs whose n-th member is a random sample from $\mathcal{PA}_d^{(n)}$, and ϕ is a monotone property. Now the fact that $\phi(G)$ is bounded away from 1, for graphs of finite size, follows from the fact that $\Pr_{G\leftarrow\mathcal{PA}_d^\infty}(\phi(G)) < 1$.

References

[1] W. Aiello, F. Chung, L. Lu. Random evolution of massive graphs. *Handbook of Massive Data Sets*, (Eds. James Abello et al.), Kluwer, 2002, pages 97-122.

[2] Reka Albert and Albert-Laszlo Barabasi. Statistical mechanics of complex networks, Reviews of Modern Physics 74, 47 (2002).

[3] A.-L. Barabasi and R. Albert, Emergence of scaling in random networks. Science, 286:509, October 1999.

[4] Aristotle, *Physics* 239b11-13

[5] P. Billingsley. *Probability and Measure.* John Wiley and Sons, New York, 3rd edition, (1995).

[6] B. Bollobas, C. Borgs, J. Chayes, and O. Riordan. Directed scale-free graphs. Proceedings of the 14th ACM-SIAM Symposium on Discrete Algorithms (2003), 132-139.

[7] B. Bollobas. Mathematical results on scale-free random graphs. pre-print at http://stat-www.berkeley.edu/users/aldous/Networks/

[8] B. Bollobas and O. Riordan. The diameter of scale-free graphs. Combinatorica, To appear.

[9] B. Bollobas and O. Riordan. Robustness and Vulnerability of Scale-Free Random Graphs. Internet Mathematics 1:1(2003) 1-35.

[10] B. Bollobas, Oliver Riordan, Joel Spencer, G. E. Tusnady. The degree sequence of a scale-free random graph process. Random Structure and Algorithms 18(2001) 279-290.

[11] A. Bonato and J. Janssen. Infinite limits of copying models of the web graph. Internet Mathematics, 1:2(2003) 193-213.

[12] A. Bonato and J. Janssen. Limits and power laws of models for the web graph and other networked information spaces, *Proceedings of Combinatorial and Algorithmic Aspects of Networking*, 2004.

[13] P. Cameron. The random graph. In Algorithms and Combinatorics 14 (R.L. Graham and J. Nesetril, eds.), Springer Verlag, New York (1997) 333-351.

[14] C. Cooper, A. M. Frieze. A general model of web graphs. Proceedings of 9th European Symposium on Algorithms, 2001, 500-511.

[15] F. Chung, L. Lu and V. Vu. Eigenvalues of Random Power law Graphs Annals of Combinatorics 7(2003) 21-33

[16] P. Erdos and A. Renyi. On the Evolution of Random Graphs. Mat. Kutato Int. Kozl 5 (1960), 17-60.

[17] Paul Erdös and Alréd Rényi. Asymetric graphs. Acta Math. Acad. Sci. Hung., 14:295-315, 1963.

[18] Michalis Faloutsos, Petros Faloutsos, Christos Faloutsos. On Power-law Relationships of the Internet Topology. Proc. SIGCOMM 1999, 251-262.

[19] Gilbert, E.N., Random graphs, Annals of Mathematical Statistics 30(1959), 1141-1144.

[20] J. Kleinberg, S.R. Kumar, P. Raghavan, S. Rajagopalan, A. Tomkins. The Web as a graph: Measurements, models and methods. Proc. International Conference on Combinatorics and Computing, 1999.

[21] Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, and Andrew Tomkins. Trawling the web for emerging cyber-communities. In *Proceedings of the Eighth International World Wide Web Conference*, 1999.

[22] Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, Andrew Tomkins: Extracting Large-Scale Knowledge Bases from the Web. VLDB 1999: 639-650

[23] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal. Stochastic models for the Web graph. Proc. 41st IEEE Symp. on Foundations of Computer Science, 2000, pp. 57-65.

[24] M. Mihail and C. H. Papadimitriou. On the eigenvalue power law. In Proceedings of RANDOM 2002.

[25] M. Mihail, C. Papadimitriou, and A. Saberi. On Certain Connectivity Properties of the Internet Topology. Proc. 44th IEEE FOCS 2003.

[26] M. Newman. The structure and function of networks. Comput. Phys. Commun., 147:40–45, 2002.

[27] P. Raghavan and R. Motwani, *Randomized Algorithms*, Cambridge University Press, 1995.

[28] J. Spencer and S. Shelah. *The strange logic of random graphs.* Spinger 2001.

A Bounding $\mathbf{E}[\tilde{X}_n \parallel X_{n_j}]$

To estimate $\mathbf{E}(\tilde{X}_n \| X_{n_j})$, we transform it into a telescoping sum:

$$\begin{aligned}
\mathbf{E}(\tilde{X}_n \| X_{n_j}) &= \mathbf{E}(X_{n_j}^2 - 2X_{n_j}X_n + X_n^2 \| X_{n_j}) \\
&= X_{n_j}^2 - 2X_{n_j}\mathbf{E}(X_n \| X_{n_j}) + \mathbf{E}(X_n^2 \| X_{n_j}) \\
&= X_{n_j}^2 - 2X_{n_j}^2 + \mathbf{E}(X_n^2 \| X_{n_j}) \\
&= \mathbf{E}(X_n^2 \| X_{n_j}) - X_{n_j}^2 \\
&= \sum_{k=n_j}^{n-1} \mathbf{E}(X_{k+1}^2 \| X_{n_j}) - \mathbf{E}(X_k^2 \| X_{n_j}).
\end{aligned}$$

We bound the sum on the right side term-by-term, using the following computation. Let $Z_k = d_{k+1}(i) - d_k(i)$; this is a sum of d independent Bernoulli random variables, each with mean $d_k(i)/2dk$. Writing d_k for $d_k(i)$, a simple computation yields

$$\begin{aligned}
\mathbf{E}(Z_k) &= d_k/2k \\
\mathbf{E}(Z_k^2) &= d_k/2k + d(d-1)d_k^2/4d^2k^2
\end{aligned}$$

hence

$$\mathbf{E}(d_{k+1}^2 \| d_k) = \mathbf{E}(Z_k^2 \| d_k) + 2d_k E(Z_k \| d_k) + d_k^2$$

$$
\begin{aligned}
&= d_k/2k + \left(\frac{d-1}{d}\right)\left(\frac{d_k^2}{4k^2}\right) + 2d_k^2/2k + d_k^2 \\
&= d_k/2k + \left(1 + \frac{1}{k} + \frac{1}{4k^2} - \frac{1}{4dk^2}\right) d_k^2 \\
&< d_k/2k + \left(1 + \frac{1}{2k}\right)^2 d_k^2 \\
\mathbf{E}(X_{k+1}^2 \| X_k) &< \left(\frac{1}{c_{k+1}}\right)^2 \mathbf{E}(d_{k+1}^2 \| X_k) \\
&< \left(\frac{c_k}{2kc_{k+1}^2}\right) X_k + \left(1 + \frac{1}{2k}\right)^2 \left(\frac{c_k}{c_{k+1}}\right)^2 X_k^2 \\
&= \left(\frac{c_k}{2kc_{k+1}^2}\right) X_k + X_k^2 \\
\mathbf{E}(X_{k+1}^2 \| X_{n_j}) &= \left(\frac{c_k}{2kc_{k+1}^2}\right) \mathbf{E}(X_k \| X_{n_j}) + \mathbf{E}(X_k^2 \| X_{n_j}) \\
&= \left(\frac{c_k}{2kc_{k+1}}\right) X_{n_j} + \mathbf{E}(X_k^2 \| X_{n_j})
\end{aligned}
$$

This means that

$$
\begin{aligned}
\mathbf{E}(\tilde{X}_n \| X_{n_j}) &< \left(\sum_{k=n_j}^{n-1} \frac{c_k}{2kc_{k+1}}\right) X_{n_j} \\
&< \left(\sum_{k=n_j}^{\infty} \frac{c_k}{2kc_{k+1}}\right) X_{n_j} \\
&< (C/\sqrt{n_j}) X_{n_j}.
\end{aligned}
$$

for some constant C, using the fact that each term of the infinite sum is $O(k^{-3/2})$.