Asymmetry and Structural Information in Preferential Attachment Graphs*

May 23, 2017

Tomasz Łuczak
Faculty of Math. & Comp. Sci.
Adam Mickiewicz University
61-614 Poznań
Poland
tomasz@amu.edu.pl

Abram Magner
Coordinated Science Lab
U. Illinois at Urbana-Champaign
Urbana, IL 618015
U.S.A.
anmagner@illinois.edu

Wojciech Szpankowski
Dept.Computer Science
Purdue University
W. Lafayette, IN 47907
U.S.A.
spa@cs.purdue.edu

Abstract

Symmetries of graphs intervene in diverse applications, ranging from enumeration to graph structure compression, to the discovery of graph dynamics (e.g., inference of the arrival order of nodes in a growing network). It has been known for some time that Erdős-Rényi graphs are asymmetric with high probability, but it is also known that real world graphs (web, biological networks) have a significant amount of symmetry. So the first natural question to ask is whether preferential attachment graphs, in which in each step a new node with m edges is added, exhibit any symmetry. It turns out that the problem is harder than in the Erdős-Rényi case and reveals unexpected results: in recent work it was proved that preferential attachment graphs are symmetric for $m = 1$ (as expected), and there is some non-negligible probability of symmetry for $m = 2$. The question was left open for $m \geq 3$. It was conjectured, however, that the preferential attachment graphs are asymmetric with high probability when more than two edges are added each time a new node arrives. In this paper we settle this conjecture in the positive. We then use it to estimate the entropy of unlabelled preferential attachment graphs (called the structural entropy of the model). To do this, we prove a new, precise asymptotic result for the labeled graph entropy, and we give bounds on another combinatorial parameter of the graph model: the number of ways in which the given graph structure could have arisen by preferential attachment (a measure of the extent to which the graph structure encodes information about the order in which nodes arrived). These results have further implications for information theoretic and statistical problems of interest on preferential attachment graphs.

Index Terms: Preferential attachment graphs, entropy, graph automorphism, symmetry, degree distribution.

*This work was supported by NSF Center for Science of Information (CSoI) Grant CCF-0939370, and in addition by NSF Grants CCF-1524312, and NIH Grant 1U01CA198941-01, and NCN grants 2012/06/A/ST1/00261 and 2013/09/B/ST6/02258.

1 Introduction

Study of the asymptotic behavior of the symmetries of graphs, originally motivated by enumerative combinatorial problems, has recently found diverse applications in problems ranging from graph compression to discovering interesting motifs to understanding dynamics of growing graphs.

Let us explore some of these applications in more detail. The basic problem of structural (unlabeled graph) compression can be formulated as follows: given a probability distribution on labeled graphs, determine a binary encoding of unlabeled graphs so as to minimize expected description length. In [5] the authors studied this problem in the setting of Erdős-Rényi graphs. They showed that, under any distribution giving equal probability to isomorphic graphs, the entropy of the induced distribution on graph structures (i.e., isomorphism classes of graphs) is less than the entropy of the original distribution by an amount proportional to the expected logarithm of the number of automorphisms. Thus the solution to the above problem is intimately connected with the symmetries of the random graph model under consideration. We explore this topic in some detail in Lemma 1 of Section 2.

We mention also a few representative algorithmic motivations for the study of graph symmetries. The first involves the problem of *motif discovery*: given a graph G and a pattern graph H, the problem is to find all subgraphs of G that are isomorphic to H. It has been observed (see, e.g., [17]) that taking into account the symmetries of H can significantly decrease the time and space complexity of the task. The same is true for G if it has nontrivial symmetries.

In the area of Markov chains, the paper [4] studies the following problem: given a graph G, the task is to assign weights to edges of G so as to minimize the mixing time of the resulting Markov chain. The authors show that symmetries in G may be exploited to significantly reduce the size of a semidefinite program formulation which solves the problem. Moreover, they point out several references to literature in which symmetry plays a key role in reducing complexity for various problems.

Study of symmetries is further motivated by their connection to various measures of information contained in a graph structure. For instance, the *topological entropy* of a random graph, studied by [18] and [20], measures the uncertainty in the orbit class (i.e., the set of nodes having the same long-term neighborhood structure) of a node chosen uniformly at random from the node set of the graph. If the graph is asymmetric with high probability, then the topological entropy is maximized: if n is the size of the graph, then the topological entropy is, to leading order, $\log n$. In general, if the symmetries of the graph can be characterized precisely, then so can the topological entropy.

The present paper is a step in the direction of understanding symmetries of complex networks and toward extending graph structure compression results to random graph models other than Erdős-Rényi. In particular, many real-world graphs exhibit a power law degree distribution (see [8]). A commonly studied model for real-world networks is the *preferential attachment* mechanism introduced in [1], in which a graph is built one vertex at a time, and each new vertex t makes m choices of neighbors in the current graph, where it attaches to a given old vertex v with probability proportional to the current degree of v. We study here a simple variant of the preferential attachment model (see [8] and the conclusion section for other models), and in the conclusion of this paper we suggest that the symmetry behavior of other preferential models can be studied using the approach developed here. Our main result is the following: for the variant of the preferential attachment model under consideration, when each vertex added to the graph chooses a fixed number $m \geq 3$ neighbors, with high probability, there are no nontrivial symmetries. This is perhaps surprising in light of the many symmetries observed in real-world networks [14].

The problem of establishing asymmetry in preferential attachment graphs appears to be difficult. Literature on symmetries of preferential attachment graphs seems to be scarce. We are aware only of [16], where two of the authors of the present paper, with colleagues, proved that such graphs are *symmetric* for $m = 1$ and (with asymptotically positive probability) also for $m = 2$. The authors of [16] conjectured that preferential attachment graphs are indeed asymmetric for $m \geq 3$. In this paper we first settle this conjecture in the positive using different methods than the one applied in [11] and [16]. Namely, instead of relying on the graph defect (a measure of asymmetry defined in [11] which is necessarily bounded in preferential attachment graphs, and hence has poor concentration properties), we shall observe that symmetry would imply that certain vertices make the same choices with regard to an initial set of vertices uniquely identifiable by their degrees, which we prove is unlikely to happen for preferential attachment graphs whenever $m \geq 3$.

After settling the asymmetry question for preferential attachment graphs, we use it to address the issue of graph entropy. We first give in Theorem 2 a precise estimate of the labelled graph entropy (improving on the result of [19]), and then estimate the unlabelled graph entropy (also known as the structural entropy). In Lemma 1 we relate both entropies. Then, using our asymmetry result from Theorem 1, we estimate the structural entropy. To derive the structural entropy estimate, we study the characteristics of the directed, acyclic graph version of the preferential attachment process (culminating in Proposition 1, which may be of independent interest).

Now we review some of the literature on symmetries of random graphs. Study of the asymptotic behavior of the automorphism group of a random graph started with a paper of Erdős and Rényi [9], where they showed that $\mathbb{G}(n, M)$ (i.e., the uniformly random graph on n vertices with M edges) with constant density (i.e. when $M = \Theta(n^2)$) is asymmetric with high probability, a result motivated by the combinatorial question of determining the asymptotics of the number of unlabeled graphs on n vertices for $n \to \infty$. Then Wright [22] proved that $\mathbb{G}(n, M)$ whp becomes asymmetric as soon as the number of isolated vertices in it drops under 1. His result was later strengthened by Bollobás [2], who also proved asymmetry for r-regular graphs with $r \geq 3$. The asymptotic size of the automorphism group of $\mathbb{G}(n, M)$ for small M, where the graph is not connected, was given by Łuczak [12]. As a similar question motivated the investigation of symmetry properties of random regular graphs, Bollobás improved his result from [2] by showing in [3] that unlabelled regular graphs with degree $r \geq 3$ are whp asymmetric as well. Let us note that it is a substantially stronger theorem (see the discussion below after Theorem 1).

For general models, asymmetry results can be nontrivial to prove, due in part to the fact that asymmetry is a global property. Furthermore, the particular models considered here present difficulties not seen in the Erdős-Rényi case: there is significant dependence between edge events, and graph sparseness makes derivation of concentration results difficult. However, settling the symmetry/asymmetry question opens the door to several other lines of investigation, including, e.g., the design of optimal structural compression schemes and the precise characterization of the limits of inference problems (see, for example, [15]) for preferential attachment graphs. These applications crucially depend on our precise understanding of graph symmetry.

The paper is organized as follows. In the next Section 2 we present our main results regarding the graph asymmetry and structural entropy. In Section 3, we state and prove several results on the degree sequence which will be useful in subsequent proofs. We prove the graph asymmetry result in Secton 4 and the entropy results in Sections 5 and 6.

2 Main Result

In this section, we state the main problem, introduce the model that we consider, and formulate the main results. First, we review some standard graph-theoretic terminology and notation.

We start with the notion of structure-preserving transformations between labeled graphs: given two graphs (possibly with multiple edges between nodes) G_1 and G_2 with vertex sets $V(G_1)$ and $V(G_2)$, a mapping $\phi : V(G_1) \to V(G_2)$ is said to be an *isomorphism* if it is bijective and preserves edge relations; that is, for any $x, y \in V(G_1)$, the number of edges (possibly 0) between x and y is equal to the number of edges in G_2 between $\phi(x)$ and $\phi(y)$. When such a ϕ exists, G_1 and G_2 are said to be isomorphic.

An isomorphism from a graph G to itself is called an *automorphism* or *symmetry* of G. The set of automorphisms of G, together with the operation of function composition, forms a group, which is called the *automorphism group* of G, denoted by $\mathrm{Aut}(G)$. We then say that G is *symmetric* if it has at least one *nontrivial symmetry* and that G is *asymmetric* if the only symmetry of G is the identity permutation.

Our first main goal is to answer, for G distributed according to a *preferential attachment* model, the question of whether with high probability the automorphism group is trivial (i.e., $|\mathrm{Aut}(G)| = 1$) or not.

We say that a multigraph G on vertex set $[n] = \{1, 2, \ldots, n\}$ is *m-left regular* if the only loop of G is at the vertex 1, and each vertex v, $2 \leq v \leq n$, has precisely m neighbours in the set $[v-1]$. The *preferential attachment model* $\mathcal{PA}(m; n)$ is a dynamic model of network growth which gives a probability measure on the set of all m-left regular graphs on n vertices proposed in [1]. More precisely, for an integer parameter m we define graph $\mathcal{PA}(m; n)$ with vertex set $[n] = \{1, 2, \ldots, n\}$ using recursion on n in the following way: the graph $G_1 \sim \mathcal{PA}(m; 1)$ is a single node with label 1 with m self-edges (these will be the only self-edges in the graph, and we will only count each such edge once in the degree of vertex 1).

Inductively, to obtain a graph $G_{n+1} \sim \mathcal{PA}(m; n+1)$ from G_n, we add vertex $n+1$ and make m random choices $v_1, ..., v_m$ of neighbors in G_n as follows: for each vertex $w \leq n$ (i.e., vertices in G_n),

$$P(v_i = w | G_n, v_1, ..., v_{i-1}) = \frac{\deg_n(w)}{2mn},$$

where throughout the paper we denote by $\deg_n(w)$ the degree of vertex $w \in [n]$ in the graph G_n (in other words, the degree of w after vertex n has made all of its choices).

We note that our proof techniques adapt to tweaks of the model in which multiple edges are not allowed.

We next formulate our first main result regarding asymmetry of $\mathcal{PA}(m; n)$ for $m \geq 3$ that we prove in Section 4.

Theorem 1 (Asymmetry for preferential attachment model)**.** *Let $G \sim \mathcal{PA}(m; n)$ for fixed $m \geq 3$. Then, with high probability as $n \to \infty$,*

$$|\mathrm{Aut}(G)| = 1.$$

More precisely, for $m \geq 3$,

$$P(|\mathrm{Aut}(G)| > 1) = O(n^{-0.004}) \tag{1}$$

for large n.

One may wonder if one can strengthen the above statement and claim that for $m \geq 3$ we have $\mathbb{E}|\mathrm{Aut}(\mathcal{PA}(m;n))| = 1 + o(1)$; if this would be the case, then a natural unlabelled version of the model, which we denote by $\mathcal{PA}^u(m;n)$, defined below would be with high probability asymmetric too. However, somewhat surprisingly, it is not the case.

To make this precise, let us recall that in the case of the uniform random graph model $\mathbb{G}(n, M)$, where we choose a graph uniformly at random from the family of all graphs with n labeled vertices and M edges, the automorphism group becomes with high probability trivial just above the connectivity threshold; i.e., when $2M/n - \log n \to \infty$; in fact, at this moment the expected size of $\mathrm{Aut}(\mathbb{G}(n, M))$ is whp $1 + o(1)$. Moreover, almost precisely at this time the unlabeled uniform random graph $\mathbb{G}^u(n, M)$ which is chosen at random from the family of all unlabeled graphs with n vertices and M edges becomes asymmetric and, furthermore, the the structure of $\mathbb{G}^u(n, M)$ is almost identical to the structure of $\mathbb{G}(n, M)$; i.e., $\mathbb{G}^u(n, M)$ is basically $\mathbb{G}(n, M)$ with erased labels (for more information on this model, see [13]). As we have already mentioned above, the same is true for r-regular random graphs with $r \geq 3$, where the uniform labeled and unlabeled graph models have basically the same asymptotic properties [3].

Returning to the preferential attachment case, for any m-left regular graph G let $S(G)$ denote the class of all m-left regular graphs which are isomorphic to G, and let $\mathcal{S}$ denote the family containing all $S(G)$, i.e. the family of all 'unlabeled m-left regular graphs'. We define the unlabeled graph distribution $\mathcal{PA}^u(n;m)$ as the probability distribution on $\mathcal{S}$, where the probability of each class $S(G)$ is proportional to the average of the probabilities that a labeled version of $S(G)$ is $\mathcal{PA}(m;n)$, i.e. proportional to

$$\frac{1}{|S(G)|} \sum_{H \in S(G)} P(H = G(m;n)) \,.$$

Note that this is a different distribution from the one that samples a preferential attachment graph and takes its isomorphism class. Note, also, that $\mathbb{G}^u(n, M)$ is defined in the same way, but in that case all terms in the sum are the same, so each equivalence class is equally likely. Now we can ask if the typical structure of $\mathcal{PA}^u(m;n)$ is the same, or very close to that of $\mathcal{PA}(m;n)$ (i.e., a preferential attachment graph with labels removed); in particular if it is asymmetric. It seems that it is not this case. To see this, notice that the typical $\mathcal{PA}(m;n)$ is asymmetric and, furthermore, whp it contains $L = \Omega_m(n)$ vertices with label at least $3n/4$, such that they are of degree m and all their neighbors are among vertices of label smaller that $n/2$. For such a graph G we will show that $|S(G)| \geq L! = \exp(\Omega_m(n \log n))$ and so for every $H \in S(G)$ we have $\Pr(H = G(m;n)) \leq \exp(-\Omega_m(n \log n))$. On the other hand for the graph H' such that all vertices of labels $\ell \geq m+1$ has neighbors $\{1, 2, \ldots, m\}$ we have $\Pr(H' = G(m;n)) \geq \exp(-O_m(n))$. Thus, the very asymmetric H' is much more likely to appear as $G^u(m;n)$ than a 'typical' graph $G(m;n)$.

Here we will not investigate the properties of $\mathcal{PA}^u(m;n)$ but rather characterize the information content of the distribution on unlabeled graphs given by sampling from $\mathcal{PA}(m;n)$ and removing the labels (i.e., taking the isomorphism class of the sampled graph).

As a direct application of Theorem 1, we estimate the structural entropy $H(S(G))$. Recall that the entropy $H(G)$ of the labelled graph $G \sim \mathcal{PA}(m; n)$ is defined as

$$H(G) = -\sum_{G \in \mathcal{G}_n} P(G) \log P(G),$$

where $\mathcal{G}_n$ denotes the set of graphs on n vertices. The structural entropy $H(S(G))$ is then simply the entropy of the isomorphism type of G. We next show how to find a relation between these two entropies. By the chain rule for conditional entropy,

$$H(G) = H(S(G)) + H(G|S(G)). \tag{2}$$

The second term, $H(G|S(G))$, measures our uncertainty about the labeled graph if we are given its structure. We will give a formula for $H(G|S(G))$ in terms of $|\mathrm{Aut}(G)|$ and another quantity, defined as follows: suppose that, after generating G, we relabel G by drawing a permutation π uniformly at random from $\mathbb{S}_n$, the symmetric group on n letters, and computing $\pi(G)$. Then conditioning on $\pi(G)$ yields a probability distribution for possible values of $\pi^{-1} = \sigma$. We can write $H(G|S(G))$ in terms of $H(\sigma|\sigma^{-1}(G)) = H(\sigma|\sigma(G))$ and $\mathbb{E}[\log |\mathrm{Aut}(G)|]$ using the chain rule for entropy, resulting in the following lemma.

Lemma 1 (Structural entropy for preferential attachment graphs)**.** *Let $G \sim \mathcal{PA}(m; n)$ for fixed $m \geq 1$, and let σ be a uniformly random permutation from $\mathbb{S}_n$. Then we have*

$$H(G) - H(S(G)) = H(\sigma|\sigma(G)) - \mathbb{E}[\log |\mathrm{Aut}(G)|]. \tag{3}$$

Remark 1. *In the proof of Theorem 3 below, we prove an alternative, more combinatorial representation for $H(\sigma|\sigma(G))$; see (34).*

To estimate the structural entropy $H(S(G))$ using Lemma 1, we need to find an expression for the labeled graph entropy $H(G)$ and evaluate the two terms on the right-hand side of (3).

In Section 5 we prove the following asymptotic formula for the entropy $H(G)$ of the preferential attachment graphs.

Theorem 2 (Entropy of preferential attachment graphs)**.** *Consider $G \sim \mathcal{PA}(m; n)$ for fixed $m \geq 1$. We have*

$$H(G) = mn \log n + m(\log 2m - 1 - \log m! - A)\, n + o(n), \tag{4}$$

where

$$A = \sum_{d=m}^{\infty} \frac{\log d}{(d+1)(d+2)}.$$

This should be compared with Theorem 1 of [19], which gives the first term and upper and lower bounds on the second term. Our result goes further by pinning down the exact value of the second term. Our proof borrows some elements from [19] but requires a nontrivial extension.

Now, we are in the position to complete our computation of the structural entropy.

Theorem 3 (Structural entropy of preferential attachment graphs)**.** *Let $m \geq 3$ be fixed. Consider $G \sim \mathcal{PA}(m;n)$. We have*

$$H(S(G)) = (m-1)n\log n + R(n), \tag{5}$$

where $R(n)$ satisfies

$$Cn \leq |R(n)| \leq O(n \log\log n)$$

for some nonzero constant $C = C(m)$.

To do this, we evaluate (3) by relating $H(\sigma|\sigma(G))$ to a combinatorial parameter of the directed version of G. We show this derivation in Section 6.

3 Results on the Degree Sequence

In this section, we present results on the degree sequence of preferential attachment graphs which we will use in the proofs of our main results in subsequent sections.

First, recall that $\deg_t(s)$ is the degree of a vertex $s < t$ after time t (i.e., after vertex t has made its choices). We also define $\mathrm{dg}_t(s) = \deg_t(s) - m$.

Our first lemma gives a bound on the in-degree of each vertex at any given time. This will give a corollary (Corollary 1) that bounds the probability that two given vertices are adjacent at a given time.

Lemma 2. *For any v, w,*

$$\Pr(\mathrm{dg}_v(w) = d) \leq \binom{m+d-1}{m-1}\left(1 - \sqrt{\frac{w}{v}} + O\left(\frac{d}{\sqrt{vw}}\right)\right)^d$$

In particular,

$$\Pr(\deg_v(w) = d) \leq (2m+d)^m \exp\left(\sqrt{\frac{w}{v}}d + O\left(\frac{d^2}{\sqrt{vw}}\right)\right).$$

Proof. We estimate this probability as follows. Below we set $t_{d+1} = mv + 1$.

$$\begin{aligned}
\Pr(\mathrm{dg}_v(w) = d) &\leq \sum_{mw<t_1<t_2<\cdots<t_d\leq mv} \prod_{i=1}^{d} \frac{m+i-1}{2t_i} \prod_{j=t_i+1}^{t_{i+1}-1}\Big(1 - \frac{m+i}{2j}\Big) \\
&\leq \sum_{mw<t_1<t_2<\cdots<t_d\leq mv} \frac{(m+d-1)!}{(m-1)!} \prod_{i=1}^{d} \frac{1+O(d/t_i)}{2t_i} \exp\Big(-\sum_{j=t_i}^{t_{i+1}-1} \frac{i}{2j}\Big) \\
&= \sum_{mw<t_1<t_2<\cdots<t_d\leq mv} \frac{(m+d-1)!}{(m-1)!} \prod_{i=1}^{d} \frac{1+O(d/t_i)}{2t_i} \exp\Big(-\sum_{j=t_i}^{mv} \frac{1}{2j}\Big) \\
&\leq \binom{d+m-1}{m-1}\Big(\sum_{i=mw+1}^{mv} \frac{1+O(d/t)}{2t} \exp\Big(-\sum_{j=t}^{mv}\frac{1}{2j}\Big)\Big)^d.
\end{aligned}$$

Note that

$$
\begin{aligned}
&\sum_{i=mw+1}^{mv} \frac{1+O(d/t)}{2t} \exp\Big(-\sum_{j=t}^{mv} \frac{1}{2j}\Big) \\
&\quad \le \sum_{i=mw+1}^{mv} \frac{1+O(d/t)}{2t} \exp\left(-\frac{1}{2}\log\frac{mv}{t} + O\Big(\frac{1}{t}\Big)\right) \\
&\quad \le \sum_{i=mw+1}^{mv} \frac{1+O(d/t)}{2\sqrt{mvt}} \\
&\quad \le 1 - \sqrt{w/v} + O(d/\sqrt{vw})\,.
\end{aligned}
$$

Thus, the assertion follows. ■

Recall that for $t > s$, the expectation of $\deg_t(s)$ is $O(\sqrt{t/s})$. We first state a simple tail bound to the right of this expectation, which may be found in [10] (it also is a corollary of Lemma 2):

Lemma 3 (Right tail bound for a vertex degree at a specific time)**.** *Let $r < t$. Then*

$$P[\deg_t(r) \ge Ae^m (t/r)^{1/2} (\log t)^2] = O(t^{-A})$$

for any constant $A > 0$ and any t.

Using the above lemma, we can show a stronger concentration result for the random variable $\deg_t(s)$ whenever $s \ll t$, as captured in the following lemma.

Lemma 4. *For $s < t$ we have*

$$P\left[|\deg_t(s) - \mathbb{E}[\deg_t(s)]| > y\right] \le \exp\left(-\frac{y^2}{O(t^{1/2+\epsilon_1}/s^{1/2})}\right) + \exp(-\mathrm{poly}(t)) \tag{6}$$

for any $y \le O(\frac{t^{1/2+\epsilon_1}}{s^{1/2}m})$ and any fixed $\epsilon_1 > 0$.

The proof uses the method of bounded variances. In particular, we will use the following result from [7].

Lemma 5 (Method of bounded variances)**.** *Let f be a function of n random variables $X_1, ..., X_n$, each X_i taking values in a set A_i, such that $\mathbb{E}[f] < \infty$.*

Assume that, for some numbers m and M,

$$m \le f(X_1, ..., X_n) \le M$$

almost surely. Let B be any event (which we think of as occurring only with low probability), and let V and c_i be defined as follows: first, we denote by F_τ, $\tau = 0, ..., n$, the σ-field generated by $X_1, ..., X_\tau$ and the event B^C (i.e., the complement of B). Then we denote by $\{Y_\tau\}$ the Doob martingale with respect to the filtration $\{\mathcal{F}_\tau\}$:

$$Y_\tau = \mathbb{E}[f(X_1, ..., X_n)|\mathcal{F}_\tau].$$

We define the difference bounds c_τ *to satisfy*

$$|Y_\tau - Y_{\tau-1}| \le c_\tau$$

with probability 1*, and the variance bounds* v_τ *to satisfy*

$$\sup_{x_1,\dots,x_{\tau-1}} \mathrm{Var}[Y_\tau - Y_{\tau-1} | X_1 = x_1, \dots, X_{\tau-1} = x_{\tau-1}] \le v_\tau.$$

We define $V = \sum_{\tau=1}^{n} v_\tau$*.*

Then, for any $t \le 2V/\max_i c_i$*,*

$$P(f < \mathbb{E}[f] - t - (M-m)P(B)) \le \exp\left(-\frac{t^2}{4V}\right) + P(B).$$

With this lemma, we can prove Lemma 4.

Proof of Lemma 4. Our choice of f in the theorem will be $\deg_t(s)$, which is a function of the random variables giving the number of times each vertex $s+\tau$, $\tau = 1, \dots, t-s$, chooses to connect to s. Each such random variable is denoted by $\deg(s+\tau \to s)$.

Easily enough, $m \le \deg_t(s) \le M = (t-s)m$, since vertex s chooses m neighbors, and each vertex after s may connect to s at most m times.

We choose B in Lemma 5 to be the unlikely event that the degree of s after any sufficiently large time is much larger than its expected value: in particular, for some constants $\epsilon, \epsilon_1 > 0$ which we will fix later,

$$B = \left[\bigcup_{\tau \ge t^\epsilon} \deg_{s+\tau}(s) > \frac{(s+\tau)^{1/2+\epsilon_1}}{s^{1/2}}\right].$$

Using Lemma 3, we can upper bound $P(B)$:

$$\begin{aligned} P(B) &\le \sum_{\tau \ge t^\epsilon} P\bigg(\deg_{s+\tau}(s) > (s+\tau)^{1/2+\epsilon_1}/s^{1/2}\bigg) \\ &\le t \cdot \exp(-\mathrm{poly}(t)), \end{aligned}$$

by plugging into the lemma $t := s+\tau$, $s := s$, and $A := e^{-m} t^{\epsilon_1}/\log^2(t+1)$ and union bounding.

Bounding the variances: We next estimate each v_τ. We define, for each $\tau \in \{0, \dots, t-s\}$, $\mathcal{F}_\tau$ to be the σ-field generated by the event B^C and the connection choices of the vertices $s+1, \dots, s+\tau$. We then define X_τ to be

$$X_\tau = \mathbb{E}[\deg_t(s) | \mathcal{F}_\tau].$$

Next, we express X_τ as a sum over vertices arriving later than s: defining $A = B^C$ for convenience,

$$X_\tau = m + \sum_{x=s+1}^{s+\tau} [\deg(x \to s) | A] + \sum_{x=s+\tau+1}^{t} \mathbb{E}[\deg(x \to s) | \mathcal{F}_\tau],$$

so that the difference $X_{\tau+1} - X_\tau$ is given by

$$\begin{aligned} X_{\tau+1} - X_\tau = & \sum_{x=s+1}^{s+\tau+1} [\deg(x \to s)|A] + \sum_{x=s+\tau+2}^{t} \mathbb{E}[\deg(x \to s)|\mathcal{F}_{\tau+1}] \\ & - (\sum_{x=s+1}^{s+\tau} [\deg(x \to s)|A] + \sum_{x=s+\tau+1}^{t} \mathbb{E}[\deg(x \to s)|\mathcal{F}_\tau]) \\ = & [\deg((s+\tau+1) \to s)|A] - \mathbb{E}[\deg(s+\tau+1) \to s|\mathcal{F}_\tau] \\ & + \sum_{x=s+\tau+2}^{t} (\mathbb{E}[\deg(x \to s)|\mathcal{F}_{\tau+1}] - \mathbb{E}[\deg(x \to s)|\mathcal{F}_\tau]). \end{aligned} \tag{7}$$

Now, recall that v_τ is, by definition,

$$v_\tau \geq \sup \mathrm{Var}[X_{\tau+1} - X_\tau|\mathcal{F}_\tau],$$

where the supremum is taken over all possible connection choices of the vertices $s+1, ..., s+\tau$. We first estimate the variances of the individual terms, and then we estimate the covariances.

Under the conditioning by $\mathcal{F}_\tau$ for $\tau > t^\epsilon$, the variance of the first term may be upper bounded as follows:

$$\mathrm{Var}[\deg(s+\tau+1 \to s)|\mathcal{F}_\tau] \leq O(\frac{(s+\tau)^{1/2+\epsilon_1}}{s^{1/2}(s+\tau)}) = O(\frac{(s+\tau)^{\epsilon_1 - 1/2}}{s^{1/2}})$$

where we have used the fact that the event A holds in the conditioning. For $\tau \leq t^\epsilon$, a cruder estimate suffices: since $0 \leq \deg(s+\tau+1 \to s) \leq m$, we have that $\mathrm{Var}[\deg(s+\tau+1 \to s)|\mathcal{F}_\tau] \leq O(m^2)$.

The variance of the second term, $\mathrm{Var}[-\mathbb{E}[\deg(s+\tau+1)|\mathcal{F}_\tau]\big|\mathcal{F}_\tau]$, is 0, because the random variable $\mathbb{E}[\deg(s+\tau+1)|\mathcal{F}_\tau]$ is a constant on the σ-field $\mathcal{F}_\tau$.

Finally, to compute the variance of the remaining sum, the plan is to upper bound it in absolute value, which will then yield a bound on the variance. In particular, we claim that the absolute value of the xth term of the sum is at most $C^{x-(s+\tau+1)}O((s+\tau)^{-1})$ for some constant $C < 1$, so that the entire sum is at most $O(1/(s+\tau))$. To prove this, we first note that we may safely ignore the conditioning on A (this incurs some error, but it is small enough to be ignored); in what follows, we denote by $\mathcal{F}'_x$ the σ-field $\mathcal{F}_x$ without the inclusion of A (i.e., the σ-field generated by the edge choices of the vertices $s+1, ..., s+x$).

Then to prove the claimed bound, we proceed by induction on x. For the base case of $x = s+\tau+2$, we have

$$\mathbb{E}[\deg(s+\tau+2 \to s)|\mathcal{F}'_{\tau+1}] = m\frac{\deg_{s+\tau+1}(s)}{2m(s+\tau+1)}\left(1 - \frac{\deg_{s+\tau+1}(s)}{2m(s+\tau+1)}\right).$$

On the other hand, $\mathbb{E}[\deg(s+\tau+2 \to s)|\mathcal{F}'_\tau]$ is given by the same expression with the degrees replaced by their expected values conditioned on $\mathcal{F}'_\tau$:

$$\mathbb{E}[\deg(s+\tau+2 \to s)|\mathcal{F}'_\tau] = m\frac{\mathbb{E}[\deg_{s+\tau+1}(s)|\mathcal{F}'_\tau]}{2m(s+\tau+1)}\left(1 - \frac{\mathbb{E}[\deg_{s+\tau+1}(s)|\mathcal{F}'_\tau]}{2m(s+\tau+1)}\right).$$

Since this degree can differ by at most m from this expected value, we have (after some calculation) that the first term of the sum is upper bounded by $\frac{m}{2x}(1+\frac{1}{2x}) \leq \frac{m}{x}$. This establishes the base case.

Now, for the inductive step, we have, from the definition of the model,

$$\begin{aligned}\mathbb{E}[\deg(x \to s)|\mathcal{F}'_{x-1}] &= \mathbb{E}\left[\text{Binomial}\left(m, \frac{\deg_{x-1}(s)}{2m(x-1)}\right)\right] \\ &= m\frac{\deg_{x-1}(s)}{2m(x-1)}\left(1 - \frac{\deg_{x-1}(s)}{2m(x-1)}\right).\end{aligned}$$

Conditioning on the smaller σ-field $\mathcal{F}'_\tau$, we get

$$\mathbb{E}[\deg(x \to s)|\mathcal{F}'_{\tau+1}] = m\frac{\mathbb{E}[\deg_{x-1}(s)|\mathcal{F}'_{\tau+1}]}{2m(x-1)}\left(1 - \frac{\mathbb{E}[\deg_{x-1}(s)|\mathcal{F}'_{\tau+1}]}{2m(x-1)}\right)$$

We can now apply the inductive hypothesis to the right-hand side to approximate the conditional expectations:

$$\begin{aligned}&\mathbb{E}[\deg(x \to s)|\mathcal{F}'_{\tau+1}] \\ &= \frac{\mathbb{E}[\deg_{x-1}(s)|\mathcal{F}'_\tau] + \frac{C^{x-s-\tau-2}D}{x-1}}{2(x-1)}\left(1 - \frac{\mathbb{E}[\deg_{x-1}(s)|\mathcal{F}'_{\tau+1}]}{2m(x-1)}\right) \\ &\leq \mathbb{E}[\deg(x \to s)|\mathcal{F}'_\tau] + \frac{C^{x-s-\tau-2}D}{2(x-1)^2}.\end{aligned}$$

Since $x \geq s + \tau + 3 \geq 5$, we have that $\frac{1}{(x-1)^2} < \frac{1}{x}$, so that

$$\mathbb{E}[\deg(x \to s)|\mathcal{F}'_{\tau+1}] - \mathbb{E}[\deg(x \to s)|\mathcal{F}'_\tau] \leq C^{x-s-\tau-1}D/x,$$

as desired.

Now, to convert this bound on the absolute value of the sum to a bound on its variance, we use the following inequality: for any random variable X with $\mathbb{E}[X] = 0$ such that $|X| \leq r$, we have

$$\text{Var}[X] \leq r^2/4.$$

Since the expected value of the sum is at most $O((s+\tau)^{-1})$, this implies that the variance of the sum is at most $O((s+\tau)^{-1})$.

We next bound the covariances of (7). The only nontrivial covariance is between the first term and the sum. We may again use the upper bound on the absolute value of the sum to upper bound the covariance: letting Y denote the sum,

$$\begin{aligned}&\text{Cov}[\deg(s+\tau+1 \to s), Y|\mathcal{F}_\tau] \\ &= \mathbb{E}[\deg(s+\tau+1 \to s)Y|\mathcal{F}_\tau] - \mathbb{E}[\deg(s+\tau+1 \to s)|\mathcal{F}_\tau]\mathbb{E}[Y|\mathcal{F}_\tau] \\ &\leq \mathbb{E}[\deg(s+\tau+1 \to s)|\mathcal{F}_\tau](O(s+\tau)^{-1} - \mathbb{E}[Y|\mathcal{F}_\tau]) \\ &= \mathbb{E}[\deg(s+\tau+1 \to s)|\mathcal{F}_\tau]O(s+\tau)^{-1}.\end{aligned}$$

To conclude, we have that

$$\text{Var}[X_{\tau+1} - X_\tau|\mathcal{F}_\tau] \leq v_\tau = \begin{cases} O(m^2) & \tau \leq t^\epsilon \\ O\left(\frac{(s+\tau)^{\epsilon_1 - 1/2}}{s^{1/2}}\right) & \tau > t^\epsilon \end{cases}$$

This implies that

$$V = \sum_{\tau=1}^{t-s} v_\tau = O(m^2 t^\epsilon) + O(t^{1/2+\epsilon_1}/s^{1/2}) = O(t^{1/2+\epsilon_1}/s^{1/2}),$$

where the last equality holds provided that we choose ϵ small enough with respect to ϵ_1. This concludes the derivation of V.

Bounding the differences: We next bound the differences c_τ. We start with the expression (7). The first two terms may be easily upper bounded by $2m$:

$$\begin{aligned} &|[\deg((s+\tau+1) \to s)|A] - \mathbb{E}[\deg(s+\tau+1) \to s|\mathcal{F}_\tau]| \\ &\leq [\deg((s+\tau+1) \to s)|A] + \mathbb{E}[\deg(s+\tau+1) \to s|\mathcal{F}_\tau] \\ &\leq 2m, \end{aligned}$$

using the triangle inequality and the fact that the maximum value for the number of times any vertex chooses another is m.

Exactly as before, the remaining sum is $O((s+\tau)^{-1})$ in absolute value, so that

$$|X_{\tau+1} - X_\tau| = c_\tau = O(m).$$

Putting everything together: Combining the estimates on c_τ and V and $P(B)$, and invoking Lemma 5, we find that

$$\begin{aligned} &P(\deg_t(s) < \mathbb{E}[\deg_t(s)] - y - (t-s)\exp(-\text{poly}(t))) \\ &\leq \exp\left(-\frac{y^2}{O(V)}\right) + \exp(-\text{poly}(t)) \\ &= \exp\left(-\frac{y^2}{O(t^{1/2+\epsilon_1}/s^{1/2})}\right) + \exp(-\text{poly}(t)) \end{aligned}$$

for any $y \leq 2V/O(m)$, as desired. ∎

Next, we give a lemma on the expected number of vertices of degree d at time t. We denote this quantity by $\bar{N}_{t,d}$ and the random variable itself by $N_{t,d}$. We start by recalling an approximation result on this quantity [21].

Lemma 6 (Expected value of $N_{t,d}$)**.** *We have, for $t \geq 1$ and $1 \leq d \leq t$ and for any fixed $m \geq 1$,*

$$\left| \bar{N}_{t,d} - \frac{2m(m+1)t}{d(d+1)(d+2)} \right| \leq C,$$

for some fixed $C = C(m) > 0$.

This approximation is useful whenever $d = o(t^{1/3})$. For larger d, the error term C dominates. For our proofs, we need to extend this result for larger d as $t \to \infty$. We have the following result along these lines.

Lemma 7 (Upper bound on $\bar{N}_{t,d}$)**.** *We have, for $t \to \infty$, $d \geq t^{1/15}$, and fixed $m \geq 1$,*

$$\bar{N}_{t,d} = O\left(\frac{t}{d(d+1)(d+2)}\right) = O\left(\frac{t}{d^3}\right). \tag{8}$$

Proof. We start with $m = 1$. First, we note that $\bar{N}_{t,d}$ satisfies the following recurrence relation:

$$\bar{N}_{t,d} = \bar{N}_{t-1,d} \cdot \left(1 - \frac{d}{2(t-1)}\right) + \bar{N}_{t-1,d-1} \cdot \frac{d-1}{2(t-1)}. \tag{9}$$

This is a consequence of the observation that a vertex at time t has degree precisely d if and only if either it had degree d at time $t-1$ and was not chosen by vertex t, or it had degree $d-1$ at time $t-1$ and was chosen by vertex t.

We prove (8) by induction on t and d. A base case for this recurrence is provided by Lemma 6. Now, we prove the claim by induction. Having established the base case, we assume that the claim holds for $t', d' < t, d$, and we prove it for t, d. Applying the recurrence and the inductive hypothesis, we have

$$\bar{N}_{t,d} \leq \frac{C(t-1)}{d(d+1)(d+2)}\left(1 - \frac{d}{2(t-1)}\right) + \frac{C(t-1)}{(d-1)d(d+1)}\frac{d-1}{2(t-1)} := \bar{N}^*_{t,d}.$$

We must check that the given upper bound is further upper bounded by

$$\frac{Ct}{d(d+1)(d+2)}.$$

In particular, we require that

$$\Delta_{t,d} := \frac{Ct}{d(d+1)(d+2)} - \bar{N}^*_{t,d} \geq 0.$$

Some algebra yields

$$\Delta_{t,d} = \frac{C}{2(t-1)} \cdot \left(\frac{1}{2d} + \frac{1}{2(d+2)} - \frac{1}{d+1}\right).$$

In particular, it is clear that we must verify that

$$\frac{1}{2d} + \frac{1}{2(d+2)} - \frac{1}{d+1} \geq 0.$$

This follows in a straightforward manner from convexity of the function $f(x) = 1/x$. From this, it is clear that $\Delta_{t,d} \geq 0$, as desired. This completes the proof of the claimed inequality on $\bar{N}_{t,d}$ for $m = 1$.

The inequality for $m > 1$ follows from the one just proven using the correspondence between $\mathcal{PA}(m;t)$ and $\mathcal{PA}(1;mt)$: a vertex of degree d in $\mathcal{PA}(m;t)$ corresponds to a particular sequence of m consecutive vertices in $\mathcal{PA}(1;mt)$ whose degrees sum to d. Thus, we mimic the proof for the $m = 1$ case, this time proving by induction that the expected number of m-tuples $(mi+1, ..., m(i+1))$ in $\mathcal{PA}(1;t)$ corresponding to single vertices in $\mathcal{PA}(m;t/m)$ with degree d is at most $O\left(\frac{t}{d(d+1)(d+2)}\right)$. The base case is again provided by Lemma 6. ∎

Lemma 2 has the following useful consequence.

Corollary 1. *Let $w < v$. Then the probability that v is adjacent to w is bounded above by $5m\sqrt{1/(vw)}\log(3v/w)$. In particular, each two vertices $v, w \geq \epsilon n$ are adjacent with probability smaller than $(5m/\epsilon)\log(3/\epsilon)/n$.*

Proof. The probability that v and w are adjacent is bounded from above by

$$\sum_{d \geq 0} \frac{md}{2mv} \Pr(\mathrm{dg}_v(w) = d - m)\,.$$

When $d \leq d_0 = 8m\sqrt{v/w}\log(3v/w)$ the above sum is clearly smaller $d_0/2 = 4m\sqrt{1/vw}\log(3v/w)$. If $d \geq d_0$ one can use Lemma 2 to estimate this sum by $m\sqrt{1/vw}\log(3v/w)$. ∎

The next result gives a bound on the probability that two early vertices have the same degree.

Lemma 8. *The probability that for some $s < s' < k^2 = n^{0.02}$ we have $\deg_n(s) = \deg_n(s')$ is $O(n^{-0.004})$.*

Proof. Let $s < s' < k^2 = n^{0.02}$. We first estimate the probability that $\deg_n(s) = \deg_n(s')$. In order to do so we set $n' = n^{0.6}$ and define

$$\underline{\deg}(s) = \deg_{n-n'}(s) \quad \text{and} \quad \underline{\underline{\deg}}(s) = \deg_n(s) - \underline{\deg}(s)\,.$$

Note that

$$\begin{aligned} P(\deg_n(s) = \deg_n(s')) &= \sum_{\underline{d},\underline{d}',\underline{\underline{d}}'} P(\deg_n(s) = \deg_n(s') | \underline{\deg}(s) = \underline{d}, \underline{\deg}(s') = \underline{d}', \underline{\underline{\deg}}(s') = \underline{\underline{d}}') \\ &\qquad \times P(\underline{\deg}(s) = \underline{d}, \underline{\deg}(s') = \underline{d}', \underline{\underline{\deg}}(s') = \underline{\underline{d}}') \\ &= \sum_{\underline{d},\underline{d}',\underline{\underline{d}}'} P(\underline{\underline{\deg}}(s) = \underline{d}' + \underline{\underline{d}}' - \underline{d} | \underline{\deg}(s) = \underline{d}, \underline{\deg}(s') = \underline{d}', \underline{\underline{\deg}}(s') = \underline{\underline{d}}') \\ &\qquad \times P(\underline{\deg}(s) = \underline{d}, \underline{\deg}(s') = \underline{d}', \underline{\underline{\deg}}(s') = \underline{\underline{d}}')\,. \end{aligned} \tag{10}$$

Observe that due to Lemma 4, with probability $1 - O(n^{-1})$ a vertex $s \in [k^2]$ has degree between $n^{0.488}$ and $n^{0.51}$ at any time in the interval $[n - n', n]$. Furthermore, one can estimate the random variable $\underline{\underline{\deg}}(s)$ conditioned on $\underline{\deg}(s) = \underline{d}$ from above and below by binomial distributed random variables and use Chernoff bound to show that with probability at least $1 - O(n^{-1})$ we have

$$\left|\frac{dn'}{2mn} - \underline{\underline{\deg}}(s)\right| = \left|0.5m\underline{d}n^{-0.4} - \underline{\underline{\deg}}(s)\right| \leq \left(\frac{dn'}{2mn}\right)^{0.6} \leq n^{0.08}\,. \tag{11}$$

Thus, in order to estimate $P(\deg_n(s) = \deg_n(s'))$, it is enough to bound

$$\rho(\underline{d}', \underline{\underline{d}}', \underline{d}) = P(\underline{\underline{\deg}}(s) = \underline{d}' + \underline{\underline{d}}' - \underline{d} | \underline{\deg}(s) = \underline{d}, \underline{\deg}(s') = \underline{d}', \underline{\underline{\deg}}(s') = \underline{\underline{d}}')$$

for $n^{0.488} \leq \underline{d}, \underline{d}' \leq n^{0.51}$ and

$$|0.5\underline{d}n^{-0.4}/m - (\underline{d}' + \underline{\underline{d}}' - \underline{d})| \leq n^{0.08}\,.$$

In order to simplify the notation set $\ell = \underline{d}' + \underline{\underline{d}}' - \underline{d}$. Let us estimate the probability that $\underline{\underline{\deg}}(s) = \ell$ conditioned on $\underline{\deg}(s) = \underline{d}$ and $\underline{\deg}(s') = \underline{d}'$. The probability that some vertex $v > n - n'$ is connected to s by more than one edge is bounded from above by

$$Cn'\Big(\frac{m\deg_n(s)}{n-n'}\Big)^2 \le n^{0.6}\,O(n^{-0.98}) = O(n^{-0.38})$$

so we can omit this case in further analysis. The probability that we connect a given vertex $v > n - n'$ with s is given by

$$\frac{m\deg_{v-1}(s)}{2m(v-1)} = \frac{\underline{d} + O(\underline{d}n^{-0.4})}{2(n - O(n'))} = \frac{\underline{d}}{2n}\Big(1 + O(n^{-0.4})\Big). \tag{12}$$

Consequently, the probability that $\underline{\underline{\deg}}(s) = \ell$ conditioned on $\underline{\deg}(s) = \underline{d}$ and $\underline{\deg}(s') = \underline{d}'$ is given by

$$\binom{n'}{\ell}\rho^\ell(1-\rho)^{n'-\ell}\Big(1+O(n^{-0.4})\Big)^\ell\Big(1+O(n^{-0.4}\underline{d}/n)\Big)^{n'-\ell},$$

where $\rho = \underline{d}/2n$.

If we additionally condition on the fact that $\underline{\underline{\deg}}(s') = \underline{\underline{d}}'$ (so that we now have conditioned on $\underline{\deg}(s) = \underline{d}, \underline{\deg}(s') = \underline{d}'$, and $\underline{\underline{\deg}}(s') = \underline{\underline{d}}'$), it will result in an extra factor of the order $\Big(1 + O(\underline{d}/2n)\Big)^{\underline{\underline{d}}'}$ since it means that some $\underline{\underline{d}}'$ vertices already made their choice (and selected s' as their neighbour). Note however that, since $\ell, \underline{\underline{d}}' = O(\underline{d}n'/n) = O(n^{0.11})$ we have

$$\begin{aligned}\Big(1+O(n^{-0.4})\Big)^\ell &= 1 + O(n^{-0.29})\\ \Big(1+O(n^{-0.4}\underline{d}/n)\Big)^{n'-\ell} &= 1 + O(n^{-0.29})\\ \Big(1+O(\underline{d}/2n)\Big)^{\underline{\underline{d}}'} &= 1 + O(n^{-0.48})\,.\end{aligned}$$

Hence, the probability that $\underline{\underline{\deg}}(s) = \ell$ conditioned on $\underline{\deg}(s) = \underline{d}$, $\underline{\deg}(s') = \underline{d}'$, and $\underline{\underline{\deg}}(s') = \underline{\underline{d}}'$ is given by

$$\binom{n'}{\ell}\rho^\ell(1-\rho)^{n'-\ell}\Big(1 + O(n^{-0.29})\big)\,,$$

and so it is well approximated by the binomial distribution. On the other hand, the probability that the random variable with binomial distribution with parameters n' and ρ takes a particular value is bounded from above by $O(1/\sqrt{n'\rho})$. Thus, for a given pair of vertices $s < s' < k^2 = n^{.02}$ we have

$$P(\deg_n(s) = \deg_n(s')) = O(\sqrt{n/n'\underline{d}}) + O(n^{-1}) = O(n^{-.044})\,.$$

Hence, the probability that such a pair of vertices, $s < s' < k^2 = n^{.02}$ exists is bounded from above by $O(k^4 n^{-.044}) = O(n^{-.004})$. ■

4 Proof of Theorem 1

In this section we shall give a complete proof of Theorem 1. Let us define first two properties, $\mathfrak{A}$ and $\mathfrak{B}$ of $G_n(m)$ which are crucial for our argument. Here and below we set, for convenience, $k = k(n) = n^{0.01}$.

($\mathfrak{A}$) $G_n(m)$ has property $\mathfrak{A}$ if no two vertices t_1, t_2, where $k < t_1 < t_2$, are adjacent to the same m neighbors from the set $[t_1 - 1]$.

($\mathfrak{B}$) $G_n(m)$ has property $\mathfrak{B}$ if the degree of every vertex $s \leq k$ is unique in $G_n(m)$, i.e. for no other vertex s' of $G_n(m)$ we have $\deg_n(s) = \deg_n(s')$.

It is easy to see that

$$P(|\mathrm{Aut}(G_n(m))| = 1) \geq P(G_n(m) \in \mathfrak{A} \cap \mathfrak{B})\,, \tag{13}$$

and so

$$P(|\mathrm{Aut}(G_n(m))| > 1) \leq P(G_n(m) \notin \mathfrak{A}) + P(G_n(m) \notin \mathfrak{B})\,. \tag{14}$$

Indeed, let us suppose that $G_n(m)$ has both properties $\mathfrak{A}$ and $\mathfrak{B}$, and $\sigma \in \mathrm{Aut}(G_n(m))$. Let us assume also that σ is not the identity, and let t_1 be the smallest vertex such that $t_2 = \sigma(t_1) \neq t_1$. Note that $\mathfrak{B}$ implies that for all $s \in [k]$ we have $\sigma(s) = s$, so that we must have $k < t_1 < t_2$. On the other hand from $\mathfrak{A}$ it follows that t_1 and $t_2 = \sigma(t_1)$ have different neigbourhoods in the set $[k]$ which consists of fixed point of σ. This contradiction shows that σ is the identity, i.e. $|\mathrm{Aut}(G_n(m))| = 1$ which proves (13).

Thus, in order to prove Theorem 1 it is enough to show that both probabilities $P(G_n(m) \notin \mathfrak{A})$ and $P(G_n(m) \notin \mathfrak{B})$ tend to 0 polynomially fast as $n \to \infty$.

Let us study first the property $\mathfrak{A}$. Our task is to estimate from above the probability that there exist vertices t_1 and t_2 such that $k < t_1 < t_2$, which select the same m neighbours (which, of course, belong to $[t_1 - 1]$). Thus we conclude

$$\begin{aligned} P(G_n(m) \notin \mathfrak{A}) &\leq \sum_{k<t_1<t_2} P(t_1, t_2 \text{ choose the same neighbours in } [t_1 - 1]) \\ &\leq \sum_{k<t_1<t_2} \sum_{1\leq r_1 \leq r_2 \ldots \leq r_m < t_1} P(t_1, t_2 \text{ choose } r_1, ..., r_m)\,. \end{aligned} \tag{15}$$

The event in the last expression is an intersection of dependent events but, if we condition on the degrees $\deg_{t_\ell}(r_s)$ of the chosen vertices r_s at times t_1, t_2, then the choice events become independent.

Let us define $\mathfrak{D}$ as an event that for some $\ell = 1, 2$, and $s = 1, 2, \ldots, m$,

$$\deg_{t_\ell}(r_s) \leq \sqrt{t_\ell / r_s} (\log t_\ell)^3\,.$$

Then from Lemma 3 it follows that

$$P(G_n(m) \notin \mathfrak{D}) \leq t_1^{-100m}\,.$$

Consequently, for $k < t_1 < t_2$ we get

$$\begin{aligned} P(t_1, t_2 \text{ choose } r_1, ..., r_m) &\leq P(t_1, t_2 \text{ choose } r_1, ..., r_m | \mathfrak{D}) + P(\neg \mathfrak{D}) \\ &\leq \prod_{\ell=1}^{2} \prod_{s=1}^{m} \frac{\sqrt{t_\ell / r_s} \log^3 t_\ell}{2 t_\ell} + k^{-100m} \\ &\leq (\log t_2)^{6m} \prod_{\ell=1}^{2} \prod_{s=1}^{m} \frac{1}{\sqrt{t_\ell r_s}} + n^{-m} \end{aligned}$$

Thus, (15) becomes

$$
\begin{aligned}
P(G_n(m) \notin \mathfrak{A}) &\le \sum_{k<t_1<t_2} (\log t_2)^{6m} \sum_{1\le r_1 \le r_2 \ldots \le r_m < t_1} \prod_{\ell=1}^{2} \prod_{s=1}^{m} \frac{1}{\sqrt{t_\ell r_s}} + n^{-1} \\
&\le \sum_{k<t_1<t_2} (t_1 t_2)^{-m/2} (\log t_2)^{6m} \sum_{1\le r_1 \le r_2 \ldots \le r_m < t_1} \prod_{s=1}^{m} \frac{1}{r_s} + n^{-1} \\
&\le \sum_{k<t_1} t_1^{-m+1} (\log t_1)^{9m} + n^{-1} \\
&\le k^{2-m} (\log k)^{9m} + n^{-1}
\end{aligned}
$$

Hence

$$
P(G_n(m) \notin \mathfrak{A}) \le n^{-0.005}\,. \tag{16}
$$

In this section we show that, with probability close to 1, the $k = n^{0.01}$ oldest vertices of $G_n(m)$ have unique degrees and so these are fixed points of every automorphism. The key ingredient of our argument is Lemma 8.

To estimate the probability that $G_n(m) \notin \mathfrak{B}$, we reason as follows: from Lemma 8 we know that with probability at least $1 - O(n^{-.004})$ the degrees of all vertices smaller than $k^2 = n^{0.02}$ are pairwise different. Furthermore, using Lemma 4, one can deduce that with probability at least $1 - O(n^{-1})$ all vertices $s < k$ have degrees larger than those of all vertices $t > k^2$ (in particular using the left tail bound to show that vertices $< k$ all have high degree and the right tail bound to show that vertices $> k^2$ have low degree whp). Consequently, with probability $1 - O(n^{-0.004})$ degrees of vertices from $[k]$ are unique, i.e. $G_n(m) \notin \mathfrak{B}$.

Finally, Theorem 1 follows directly from (14) and our estimates for $P(G_n(m) \notin \mathfrak{A})$ and $P(G_n(m) \notin \mathfrak{B})$.

5 Proof of Theorem 2

In this section we prove Theorem 2 on the entropy of labeled preferential attachment graphs.

We start by noting that, using the chain rule for entropy, we can write

$$
H(G_n) = \sum_{t=1}^{n} H(v_{t+1}|G_t), \tag{17}
$$

where we denote by v_{t+1} the multiset of connection choices of vertex $t+1$ (i.e., a value for v_{t+1} takes the form of a multiset of m vertices $< t+1$). This follows because G_n corresponds precisely to exactly one n-tuple $(v_1, v_2, ..., v_n)$ of vertex choice multisets.

To calculate the remaining conditional entropy for each t, we first note that it would be simpler if v_{t+1} were a sequence of vertex choices, rather than a multiset (i.e., an equivalence class of sequences). First, let us denote by $\tilde{v}_{t+1}$ the sequence of m choices made by vertex $t+1$. I.e., $\tilde{v}_{t+1,1}$ is the first choice that it makes, and so on. Then we have the following observation:

$$
H(\tilde{v}_{t+1}|G_t) = H(\tilde{v}_{t+1}, v_{t+1}|G_t) = H(v_{t+1}|G_t) + H(\tilde{v}_{t+1}|v_{t+1}, G_t), \tag{18}
$$

where the first equality is because v_{t+1} is a deterministic function of $\tilde{v}_{t+1}$, and the second is by the chain rule for conditional entropy. We thus have

$$
H(v_{t+1}|G_t) = H(\tilde{v}_{t+1}|G_t) - H(\tilde{v}_{t+1}|v_{t+1}, G_t). \tag{19}
$$

The second term on the right-hand side is at most a constant with respect to n, so its total contribution to $H(G_n)$ is at most $O(n)$. We will estimate it precisely later, but will first compute $H(\tilde{v}_{t+1}|G_t)$.

By definition of conditional entropy,

$$H(\tilde{v}_{t+1}|G_t) = \sum_{G \text{ on } t \text{ vertices}} P(G_t = G)H(\tilde{v}_{t+1}|G_t = G).$$

Next, note that, conditioned on $G_t = G$, the m choices that vertex $t+1$ makes are independent and identically distributed. So the remaining conditional entropy is just m times the conditional entropy of a single vertex choice made by $t+1$. Using the definition of entropy (as a sum over all possible vertex choices, from 1 to t) and grouping together terms corresponding to vertices of the same degree (which all have the same conditional probability), we get

$$H(\tilde{v}_{t+1}|G_t) = m \sum_G P(G_t = G) \sum_{d=m}^{t} N_d(G) p_{t,d} \log(1/p_{t,d}), \tag{20}$$

where $N_d(G)$ denotes the number of vertices of degree d in the fixed graph G, and we define (using the notation of [19])

$$p_{t,d} = \frac{d}{2mt}.$$

Note that the d sum starts from $d = m$, since m is the minimum possible degree in the graph.

Next, we bring the G sum inside the d sum, and we note that

$$\sum_G P(G_t = G)N_d(G) = \mathbb{E}[N_d(G)],$$

which we denote by $\bar{N}_{t,d}$.

Thus, we can express $H(\tilde{v}_{t+1}|G_t)$ as

$$H(\tilde{v}_{t+1}|G_t) = m \sum_{d=m}^{t} \bar{N}_{t,d} p_{t,d} \log(1/p_{t,d}), \tag{21}$$

Plugging this into (17), we get

$$H(G_n) + \sum_{t=1}^{n} H(\tilde{v}_{t+1}|v_{t+1}, G_t) = m \sum_{t=1}^{n} \sum_{d=m}^{t} \bar{N}_{t,d} p_{t,d} \log(1/p_{t,d}). \tag{22}$$

Now, we split the inner sum into two parts:

$$\begin{aligned} H(G_n) + \sum_{t=1}^{n} H(\tilde{v}_{t+1}|v_{t+1}, G_t) = m \sum_{t=1}^{n} \sum_{d=m}^{\lfloor t^{1/15} \rfloor} \bar{N}_{t,d} p_{t,d} \log(1/p_{t,d}) \\ + m \sum_{t=1}^{n} \sum_{d=\lfloor t^{1/15} \rfloor + 1}^{t} \bar{N}_{t,d} p_{t,d} \log(1/p_{t,d}). \end{aligned} \tag{23}$$

The first part provides the dominant contribution, of order $\Theta(n \log n)$, and we will show that the second part is $o(n)$, due to the smallness of $\bar{N}_{t,d}$.

Estimating the small d terms: To estimate the contribution of the first sum, we apply Lemma 6 to estimate $\bar{N}_{t,d}$ and we use the definition of $p_{t,d}$:

$$\begin{aligned}
&\sum_{t=1}^{n} \sum_{d=m}^{\lfloor t^{1/15} \rfloor} \bar{N}_{t,d} p_{t,d} \log(1/p_{t,d}) + \sum_{t=1}^{n} \sum_{d=m}^{\lfloor t^{1/15} \rfloor} \frac{Cd}{2mt} \log(2mt/d) \\
&= 2m(m+1) \sum_{t=1}^{n} t \sum_{d=m}^{\lfloor t^{1/15} \rfloor} \frac{1}{d(d+1)(d+2)} \frac{d}{2mt} \log\left(\frac{2mt}{d}\right) + o(n) \\
&= (m+1) \sum_{t=1}^{n} \sum_{d=m}^{\lfloor t^{1/15} \rfloor} \frac{d(\log t + \log 2m - \log d)}{d(d+1)(d+2)} + o(n).
\end{aligned}$$

Here, the second sum on the left-hand side is the error in approximation incurred by invoking Lemma 6. It is easily seen to be $o(n)$.

Now, the tail sum is

$$\sum_{t=1}^{n} \sum_{d=\lfloor t^{1/15} \rfloor + 1}^{\infty} \frac{d(\log t + \log 2m - \log d)}{d(d+1)(d+2)} \le \sum_{t=1}^{n} \sum_{d=\lfloor t^{1/15} \rfloor + 1}^{\infty} \frac{O(\log d)}{(d+1)(d+2)} = o(n),$$

so we have

$$\sum_{t=1}^{n} \sum_{d=m}^{\lfloor t^{1/15} \rfloor} \bar{N}_{t,d} p_{t,d} \log(1/p_{t,d}) = \log n! + (\log 2m - A)\, n + o(n),$$

where we define A as in the statement of Theorem 2.

Upper bounding the large d terms: Our goal is now to show that the second sum of (23), which we denote by E, is $o(n)$.

We apply Lemma 7 to upper bound $\bar{N}_{t,d}$, which yields

$$E \le C \sum_{t=1}^{n} \sum_{d=\lfloor t^{1/15} \rfloor + 1}^{t} \frac{t}{d^3} \cdot \frac{d}{2tm} \log(2tm/d) \le C' \sum_{t=1}^{n} \log t \sum_{d=\lfloor t^{1/15} \rfloor + 1}^{t} d^{-2},$$

where we canceled factors in the numerator and denominator of each term, and we upper bounded the expression inside the logarithm using the fact that $d > \lfloor t^{1/15} \rfloor$.

The inner sum is easily seen to be $O(t^{-1/15})$, so that, finally,

$$E \le C' \sum_{t=1}^{n} t^{-1/15} \log t = o(n),$$

as desired.

We thus end up with

$$\sum_{t=1}^{n} H(\tilde{v}_{t+1} | G_t) = m \log n! + m(\log 2m - A)n + o(n). \tag{24}$$

Estimating $H(\tilde{v}_{t+1}|v_{t+1}, G_t)$**:** The final step is to estimate the contribution of $H(\tilde{v}_{t+1}|v_{t+1}, G_t)$. Let $\mathcal{C}_t$ denote the set of multisets of m elements coming from $[t]$ having no repeated elements. Then we can write

$$\begin{aligned} H(\tilde{v}_{t+1}|v_{t+1}, G_t) &= \sum_{G, v \in \mathcal{C}_t} P(G_t = G, v_{t+1} = v) H(\tilde{v}_{t+1}|v_{t+1} = v, G_t = G) \\ &+ \sum_{G, v \notin \mathcal{C}_t} P(G_t = G, v_{t+1} = v) H(\tilde{v}_{t+1}|v_{t+1} = v, G_t = G). \end{aligned} \tag{25}$$

The first sum can be estimated as follows: we trivially upper bound $H(\tilde{v}_{t+1}|v_{t+1} = v, G_t = G) \leq \log m!$ and take it outside the sum. This gives

$$\begin{aligned} \sum_{G, v \in \mathcal{C}_t} P(G_t = G, v_{t+1} = v) H(\tilde{v}_{t+1}|v_{t+1} = v, G_t = G) &\leq \log m! \sum_{G, v \in \mathcal{C}_t} P(G_t = G, v_{t+1} = v) \\ &= \log m! P(v_{t+1} \in \mathcal{C}_t). \end{aligned}$$

Now we can upper bound the remaining probability in this expression by noting that with high probability, the maximum degree in G_t is $\tilde{O}(\sqrt{t})$ [10]. Using this fact, we have, for arbitrarily small fixed $\epsilon > 0$,

$$\begin{aligned} P(v_{t+1} \in \mathcal{C}_t) &= P(v_{t+1} \in \mathcal{C}_t, \text{max. degree of } G_t \leq Ct^{1/2+\epsilon}) \\ &+ P(v_{t+1} \in \mathcal{C}_t, \text{max. degree of } G_t > Ct^{1/2+\epsilon}) \end{aligned} \tag{26}$$

The first term is at most

$$\begin{aligned} P(v_{t+1} \in \mathcal{C}_t, \text{max. degree of } G_t \leq Ct^{1/2+\epsilon}) &\leq 1 - \left(1 - \frac{Ct^{1/2+\epsilon}}{2mt}\right)^{m-1} \\ &= 1 - \left(1 - \Theta(t^{-1/2+\epsilon}/m)\right)^{m-1} \\ &= \Theta(t^{-1/2+\epsilon}). \end{aligned}$$

Now, the second term of (26) is at most

$$\begin{aligned} P(v_{t+1} \in \mathcal{C}_t, \text{max. degree of } G_t > Ct^{1/2+\epsilon}) &\leq P(\text{max. degree of } G_t > Ct^{1/2+\epsilon}) \\ &= O(e^{-t^{\epsilon}}) \end{aligned}$$

and is thus negligible compared to the first term.

Thus, the first sum in (25) is at most

$$\sum_{G, v \in \mathcal{C}_t} P(G_t = G, v_{t+1} = v) H(\tilde{v}_{t+1}|v_{t+1} = v, G_t = G) = O(t^{-1/2+\epsilon}). \tag{27}$$

We will now show that the second sum in (25), over all multisets v of size m with no repeated elements, is $(1 + o(1)) \log m!$. This is trivial, since vertex $t+1$ is equally likely to have chosen the elements of v in any order. Thus,

$$H(\tilde{v}_{t+1}|v_{t+1} = v, G_t = G) = \log m!. \tag{28}$$

This implies that

$$\sum_{G,v\notin\mathcal{C}_t} P(G_t = G, v_{t+1} = v)H(\tilde{v}_{t+1}|v_{t+1} = v, G_t = G) = \log m! \cdot P(v_{t+1} \notin \mathcal{C}_t)$$

$$= \log m!(1 - O(t^{-1/2+\epsilon})).$$

Thus,

$$H(\tilde{v}_{t+1}|v_{t+1}, G_t) = \log m!(1 + O(t^{-1/2+\epsilon})).$$

Summing over all t yields a total contribution of

$$-\sum_{t=1}^{n} H(\tilde{v}_{t+1}|v_{t+1}, G_t) = -n\log m! + o(n). \tag{29}$$

Putting everything together: From (22), (24), and (29), we get

$$H(G_n) = mn\log n + m(\log 2m - 1 - A - \log m!)n + o(n), \tag{30}$$

where A is as in the statement of Theorem 2. ∎

6 Proof of Theorem 3

We now prove the claimed estimate of the structural entropy.

We first show that the contribution of $\mathbb{E}[\log|\mathrm{Aut}(G)|]$ is negligible (in particular, $o(n)$). From Theorem 1, we immediately have

$$\mathbb{E}[\log|\mathrm{Aut}(G)|] \le n\log n \cdot n^{-0.004} = o(n).$$

We now move on to estimate $H(\sigma|\sigma(G))$, which we will show to satisfy

$$n\log n - O(n\log\log n) \le H(\sigma|\sigma(G)) \le n\log n - n + O(\log n). \tag{31}$$

To go further, we need to define a few sets which will play a role in our derivation. We define the *admissible set* $\mathrm{Adm}(S)$ of a given unlabeled graph S to be the set of all labeled graphs g with $S(g) = S$ such that g could have been generated according to the preferential attachment model with given parameters. That is, denoting by g_t the subgraph of g induced by the vertices $1, ..., t$ for each $t \in [n]$, we have that the degree of vertex t in g_t is exactly m. We can similarly define $\mathrm{Adm}(g) = \mathrm{Adm}(S(g))$. Then, for a graph g, we define $\Gamma(g)$ to be the set of permutations π such that $\pi(g) \in \mathrm{Adm}(g)$. We will also define, for an arbitrary set of graphs B,

$$\mathrm{Adm}_B(g) = \mathrm{Adm}(g) \cap B, \qquad \Gamma_B(g) = \{\pi \ : \ \pi(g) \in \mathrm{Adm}_B(g)\}.$$

For a given graph g, these sets are related by the following formula (the simple proof of this fact is a tweak of that given in [15]):

$$|\mathrm{Adm}_B(g)| = \frac{|\Gamma_B(g)|}{|\mathrm{Aut}(g)|}. \tag{32}$$

We next need to consider some directed graphs associated with G: we start with $\mathrm{DAG}(G)$, which is defined on the same vertex set as G; there is an edge from u to $v < u$ in $\mathrm{DAG}(G)$ if and only if there is an edge between u and v in G (in other words, $\mathrm{DAG}(G)$ is simply the graph G before we remove edge directions). Note that, if we ignore self-loops, $\mathrm{DAG}(G)$ is a directed, acyclic graph.

We denote the *unlabeled* version of $\mathrm{DAG}(G)$ (i.e., the set of all labeled directed graphs with the same structure as $\mathrm{DAG}(G)$) by $\mathrm{UDAG}(G)$. We will also, at times, abuse notation and write $\mathrm{UDAG}(G)$ as the set of all labeled, undirected graphs with the same structure as $\mathrm{UDAG}(G)$ and with labeling consistent with $\mathrm{UDAG}(G)$ as a partial order.

We have the following observations regarding these directed graphs.

Lemma 9. *For any two graphs g_1, g_2 satisfying $\mathrm{UDAG}(g_1) = \mathrm{UDAG}(g_2)$, we have*

$$P(G = g_1) = P(G = g_2).$$

Proof. This can be seen by deriving a formula for the probability assigned to a given graph g by the model and noting that it only depends on the structure and admissibility (a graph is said to be admissible if it is in $\mathrm{Adm}(S)$ for some unlabeled graph S). If g is not admissible, then there exists some $t \in [n]$ such that the degree of vertex t at time t is not equal to m. This has probability 0, so $P(G = g) = 0$.

Now, if g is an admissible graph, then we can write $P(G = g)$ as a product over possible degrees of vertices at time n: let $\deg_g(v)$ denote the degree of vertex v in g. We consider the immediate ancestors (i.e., the parents, the vertices that chose to connect to v) of v in $\mathrm{DAG}(g)$, denoting the number of edges that they supply to v by $d_1(v), ..., d_{k(v)}(v)$, where $k(v)$ is the number of parents of v. We also denote by $K_g(v)$ the number of orders in which the parents of v could have arrived in the graph (which is only a function of $\mathrm{UDAG}(g)$. Then we can write $P(G = g)$ as follows:

$$P(G = g) = \frac{\prod_{d \geq m} \prod_{v \,:\, \deg_g(v) = d} K_g(v) \prod_{j=1}^{k_g(v)} \binom{m}{j} (m + d_1(v) + \cdots d_{j-1}(v))^{d_j(v)}}{\prod_{i=1}^{n-1} (2mi)^m}. \tag{33}$$

Here, each factor of the v product corresponds to the sequence of $d - m$ choices to connect to vertex v, which can be ordered in a number of ways determined by the structure of $\mathrm{DAG}(g)$. The innermost product gives the contribution of each such choice. Since this formula is only in terms of the degree sequence of the graph and $\mathrm{UDAG}(g)$, two graphs that are admissible and have the same unlabeled DAG must have the same probability, which completes the proof. ∎

Lemma 10. *Fix an unlabeled graph S on n nodes with $P(S(G) = S) > 0$ with some fixed $m \geq 1$. Then the number of distinct unlabeled directed graphs with undirected structure S is at most $e^{\Theta(n)}$.*

Proof. Observe that the number of edges in S is $\Theta(n)$, as it arises with positive probability from $\mathcal{PA}(m; n)$ and m is fixed.

Then note that each of the $\Theta(n)$ edges may be given one of two orientations, resulting in at most $2^{\Theta(n)}$ distinct directed graphs, which completes the proof. ∎

The next lemma shows that $H(\sigma|\sigma(G))$ may be expressed in terms of the quantities just defined.

Lemma 11. *Fix $m \geq 1$ and consider $G \sim \mathcal{PA}(m; n)$. Let $\sigma \in \mathbb{S}_n$ be a uniformly random permutation. Then*

$$H(\sigma|\sigma(G)) = \mathbb{E}[\log|\Gamma_{\mathrm{UDAG}(G)}(G)|] + O(n). \tag{34}$$

Proof. First, we give an alternative representation of $H(\sigma|\sigma(G))$. Recall that $H(G|S(G)) = H(\sigma|\sigma(G)) - \mathbb{E}[\log|\mathrm{Aut}(G)|]$. The plan is to derive an alternative expression for $H(G|S(G))$ as follows: by the chain rule for entropy, we have

$$\begin{aligned} H(G|S(G)) &= H(G, \mathrm{UDAG}(G)|S(G)) \\ &= H(\mathrm{UDAG}(G)|S(G)) + H(G|\mathrm{UDAG}(G)) \\ &= O(n) + H(G|\mathrm{UDAG}(G)). \end{aligned}$$

Here, the last equality is a result of Lemma 10. Now, by Lemma 9, we have

$$H(G|\mathrm{UDAG}(G)) = \mathbb{E}[\log|\mathrm{Adm}_{\mathrm{UDAG}(G)}(G)|] = \mathbb{E}[\log|\Gamma_{\mathrm{UDAG}(G)}|] - \mathbb{E}[\log|\mathrm{Aut}(G)|] + O(n),$$

where the second equality is an application of (32). This completes the proof. ∎

Remark 2. *Note that Lemma 11 is robust to small variations in the model.*

Now, to calculate $H(\sigma|\sigma(G))$, it thus remains to estimate $\mathbb{E}[\log|\Gamma_{\mathrm{UDAG}(G)}(G)|]$.

We will lower bound $|\Gamma_{\mathrm{UDAG}(G)}(G)|$ in terms of the sizes of the *levels* of $\mathrm{DAG}(G)$, defined as follows: L_1 consists of the vertices with in-degree 0 (i.e., with total degree m). Inductively, L_j is the set of vertices incident on edges coming from vertices in L_{j-1}. Equivalently, a vertex w is an element of some level $\geq j$ if and only if there exist vertices $v_1 < \cdots < v_j$ such with $v_1 > w$ and the path $v_j v_{j-1} \cdots v_1 w$ exists in G.

Then it is not too hard to see that any product of permutations that only permute vertices within levels is a member of $\Gamma_{\mathrm{UDAG}(G)}(G)$. Thus, we have, with probability 1,

$$|\Gamma_{\mathrm{UDAG}(G)}(G)| \geq \prod_{j \geq 1} |L_j|!.$$

To continue, we will prove a proposition (Proposition 1 below), to the effect that almost all vertices lie in low levels of $\mathrm{DAG}(G)$. We define $X = X(\epsilon, k)$ to be the number of vertices $w > \epsilon n$ that are at level $\geq k$ in $\mathrm{DAG}(G)$. In other words, w is counted in X if there exist vertices $v_1 < v_2 < \cdots < v_k$ for which $w < v_1$ and the path $v_k \cdots v_1 w$ exists in $\mathrm{DAG}(G)$.

We have the following lemma bounding $\mathbb{E}[X]$:

Lemma 12. *For any $\epsilon = \epsilon(n) > 0$, there exists $k = k(\epsilon)$ for which*

$$\mathbb{E}[X(\epsilon, k)] \leq \epsilon n.$$

In particular, we can take any k satisfying

$$k \geq 15 \frac{m}{\epsilon^2} \log(3/\epsilon). \tag{35}$$

Proof. Suppose that $w > \epsilon n$. We want to upper bound the probability that there exist vertices $v_1 < \cdots < v_k$, with $w < v_1$, such that there is a path $v_k \cdots v_1 w$ in G. Applying Corollary 1, this probability is upper bounded by

$$\binom{n}{k} \cdot \frac{((5m/\epsilon)\log(3/\epsilon))^k}{n^k} \leq \frac{e\,((5m/\epsilon)\log(3/\epsilon))^k}{k^k}$$

Now, it is sufficient to show that we can choose k so that this is $\leq \epsilon$. In fact, we can choose $k \geq 3 \cdot \frac{5m}{\epsilon^2} \log(3/\epsilon)$. This completes the proof. ■

Now, we define $Y = Y(k)$ to be the number of vertices $w \geq 1$ that are at level $\geq k$ in $\mathrm{DAG}(G)$. The variables X and Y are related by the following inequalities, which hold with probability 1:

$$X \leq Y \leq X + \epsilon n.$$

Now, to get a bound on Y, we apply Markov's inequality:

$$\Pr[Y \geq \delta n] \leq \frac{\mathbb{E}[Y]}{\delta n} \leq \frac{\mathbb{E}[X] + \epsilon n}{\delta n},$$

and provided that (35) holds, we can further bound by

$$\Pr[Y \geq \delta] \leq 2\epsilon/\delta$$

using Lemma 12. Then, provided that we choose $\delta = \sqrt{2\epsilon}$, we have shown that

$$\Pr[Y \geq \delta] \leq \delta.$$

This is summarized in the following proposition.

Proposition 1. *For any $\delta = \delta(n) > 0$, there exists $\ell = \ell(\delta)$ for which the number of vertices that are not in the first ℓ layers of $\mathrm{DAG}(G)$ is at most δn, with high probability.*

In particular, we can take $\ell \geq \frac{15m}{2\delta^4} \log(3/(2\delta^2))$.

We now use Proposition 1 to finish our lower bound on $\mathbb{E}[\log |\Gamma_{\mathrm{UDAG}(G)}(G)|]$. Fix $\epsilon = \frac{1}{\log^2 n}$, so that $\delta = \sqrt{2\epsilon} = \Theta(1/\log n)$, and choose $\ell = \frac{15m}{2\delta^4} \log(3/(2\delta^2))$. Then, defining A to be the event that the number of vertices in layers $> \ell$ is at most $\delta n = \Theta(n/\log n)$, we have

$$\mathbb{E}[\log |\Gamma_{\mathrm{UDAG}(G)}(G)|] \geq \mathbb{E}[\log |\Gamma_{\mathrm{UDAG}(G)}(G)| \;\big|\; A](1 - \delta).$$

Among the ℓ layers, there are at most $\ell - 1$ that satisfy, say, $|L_i| < \log\log n$, since $\sum_{i=1}^{\ell} |L_i| \geq (1-\delta)n$. So we have the following:

$$\sum_{i=1}^{\ell} \log(|L_i|!) = O(\ell \log\log n \log\log\log n) + \sum_{i \in B} (|L_i| \log |L_i| + O(|L_i|)),$$

where $B = \{i \leq \ell \;:\; |L_i| \geq \log\log n\}$, and we used Stirling's formula to estimate the terms $i \in B$.

The sum $\sum_{i \in B} O(|L_i|) = O((1-\delta)n) = O(n)$, so it remains to estimate

$$\sum_{i \in B} |L_i| \log |L_i|.$$

Let $N = \sum_{i \in B} |L_i|$. Then, multiplying and dividing each instance of $|L_i|$ by N in the above expression, it becomes

$$\sum_{i \in B} |L_i| \log |L_i| = N \sum_{i \in B} \frac{|L_i|}{N} \log \frac{|L_i|}{N} + N \sum_{i \in B} \frac{|L_i|}{N} \log N.$$

The first sum is simply $-NH(X)$, where X is a random variable distributed according to the empirical distribution of the vertices on the levels $i \in B$. Since $|B| \leq \ell$, we have that $|-NH(X)| \leq N \log \ell$. Thus, the first term in the above expression is $O(N \log \ell) = O(n \log \log n)$. Meanwhile, the second term is $N \log N \sum_{i \in B} \frac{|L_i|}{N} = N \log N = n \log n - O(n \log \log n)$. Thus, in total, we have shown

$$\mathbb{E}[\log |\Gamma_{\mathrm{UDAG}(G)}(G)|] \geq n \log n - O(n \log \log n).$$

Compare this with the trivial upper bound on $\mathbb{E}[\log |\Gamma_{\mathrm{UDAG}(G)}(G)|]$:

$$\mathbb{E}[\log |\Gamma_{\mathrm{UDAG}(G)}(G)|] \leq \log n! = n \log n - n + O(\log n).$$

This implies that we have recovered the first term, but there is a gap in our lower and upper bounds on the second term.

7 Conclusion and Further Work

In this paper, we just proved that a version of the standard preferential attachment graph is asymmetric if every node adds *more* than two edges. It is easy to extend this statement to the case when the attachment is uniform and a mixture of uniform and preferential: e.g., for a fixed $\beta \in [0,1]$, the probability that a connection choice goes to node w at time $n+1$ is

$$P(v_i = w | G_n, v_1, ..., v_{i-1}) = \beta \frac{\deg_n(w)}{2mn} + (1-\beta)\frac{1}{n}.$$

Another, possibly more practical, model was introduced by Cooper and Frieze [6] in which essentially the number of edges added follows a given distribution. We believe our methodology can handle this case, too.

However, consider a model in which the weight of a vertex when m new edges are generated is proportional to the degree raised to some power α. In this paper we considered $\alpha = 1$. We are confident our approach could be adopted to work for all $\alpha > 0$ to find the threshold m_α for the asymmetry which, clearly, will grow with α. However, in the case $\alpha \neq 1$ the problem becomes much harder since, for instance, the probability that t chooses vertex s as its neighbor depends not only on the degree $\deg_t(s)$ but on the whole degree sequence at the time t. Nonetheless, these difficulties could be overcome by modern combinatorial methods and we plan to deal with this model in the nearest future.

References

[1] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys*, 74:47–97, 2002.

[2] B. Bollobás. Distinguishing vertices of random graphs. *Ann. Discrete Mathematics*, 13:33–50, 1982.

[3] Béla Bollobás. The asymptotic number of unlabelled regular graphs. *Journal of the London Mathematical Society*, 26:201–206, 1982.

[4] Stephen Boyd, Perci Diaconis, Pablo Parrilo, and Lin Xiao. Fastest mixing markov chain on graphs with symmetries. *SIAM Journal on Optimization*, 20(2):792–819, 2009.

[5] Yongwook Choi and W. Szpankowski. Compression of graphical structures: Fundamental limits, algorithms, and experiments. *Information Theory, IEEE Transactions on*, 58(2):620–638, 2012.

[6] Colin Cooper and Alan Frieze. A general model of web graphs. *Random Structures and Algorithms*, 22:311–335, 2003.

[7] Devdatt P. Dubhashi and Allesandro Panconesi. *Concentration of Measure for the Analysis of Randomized Algorithms*. Cambridge University Press, New York, NY, USA, 2009.

[8] Rick Durrett. *Random Graph Dynamics (Cambridge Series in Statistical and Probabilistic Mathematics)*. Cambridge University Press, New York, NY, USA, 2006.

[9] Paul Erdős and Alfred Rényi. Asymmetric graphs. *Acta Math. Acad. Sci. Hungar.*, 14:295–315, 1963.

[10] Alan Frieze and Michał Karoński. *Introduction to Random Graphs*. Cambridge University Press, 2016.

[11] Jeong Han Kim, Benny Sudakov, and Van H. Vu. On the asymmetry of random regular graphs and random graphs. *Random Structures & Algorithms*, 21(3-4):216–224, 2002.

[12] T. Łuczak. The automorphism group of random graphs with a given number of edges. *Math. Proc. Camb. Phil. Soc.*, 104:441–449.

[13] Tomasz Łuczak. How to deal with unlabelled random graphs. *J. Graph Theory*, 15:303–316, 1991.

[14] B. D. MacArthur and J. W. Anderson. Symmetry and Self-Organization in Complex Systems. *eprint arXiv:cond-mat/0609274*, September 2006.

[15] Abram Magner, Ananth Grama, Jithin Sreedharan, and Wojciech Szpankowski. Recovery of vertex orderings in dynamic graphs. *Proceedings of the IEEE International Symposium on Information Theory (to appear)*, 2017.

[16] Abram Magner, Svante Janson, Giorgos Kollias, and Wojciech Szpankowski. On symmetries of uniform and preferential attachment graphs. *Electronic Journal of Combinatorics*, 21, 2014.

[17] F. Picard, J. J. Daudin, M. Koskas, S. Schbath, and S. Robin. Assessing the exceptionality of network motifs. *Journal of Computational Biology*, 15(1):1–20, 2008.

[18] N. Rashevsky. Life, information theory, and topology. *The bulletin of mathematical biophysics*, 17(3):229–235, 1955.

[19] Martin Sauerhoff. On the entropy of models for the web graph. *In submission*, 2016.

[20] Ernesto Trucco. A note on the information content of graphs. *The bulletin of mathematical biophysics*, 18(2):129–135, 1956.

[21] Remco van der Hofstad. *Random Graphs and Complex Networks: Volume 1*. Cambridge University Press, 2016.

[22] E. M. Wright. Asymmetric and symmetric graphs. *Glasgow Math. J.*, pages 69–73, 1974.