Structural Information and Compression of Scale-Free Graphs*

July 12, 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

The increasing prevalence of structural information in various forms in the wild has spurred interest in the information theory and computer science communities in efficient compression and transmission of graph-structured data. Networks in diverse application areas have been observed to exhibit a power law degree distribution, and, hence, it is a natural first step to consider compression of such graphs (as well as their unlabeled counterparts), under the assumption that they are generated by the *preferential attachment* model (initially devised to capture precisely the power law phenomenon using a simple mechanism), with parameter $m \geq 1$ giving the number of attachment choices that each new vertex makes.

In this work, we give algorithmically efficient, asymptotically optimal algorithms for compression of both unlabeled and labeled preferential attachment graphs. Showing the optimality of our schemes entails new, precise estimates of the Shannon entropy of both models, which in turn require new results on quantities of independent interest: the typical size of the automorphism group, as well as some shape parameters of the directed version of the graph, which in turn allows us to estimate the typical number of *admissible labeled representatives* of a given graph structure. Our result on the automorphism group positively settles a conjecture to the effect that, provided that $m \geq 3$, preferential attachment graphs are asymmetric with high probability, and completes the characterization of the number of symmetries for a broad range of parameters of the model (i.e., for all fixed m).

Index Terms: graph compression, symmetry, preferential attachment, random graphs

*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

The increasing prevalence of network-structured data (e.g., protein interaction, social, and technological networks) in applications has stimulated interest among information theorists and computer scientists in the problem of efficient coding (in terms of expected code length) of labeled and unlabeled graphs (i.e., graph *structures*). Formally, the basic problem is as follows: fix a distribution $\mathbb{G}_n$ on (multi)graphs on n vertices. We would like to exhibit an efficiently computable *source code* [6] $(\mathcal{C}_n, \mathcal{D}_n)$ for $\mathbb{G}_n$, where $\mathcal{C}_n$ is a function mapping graphs in the support of $\mathbb{G}_n$ to bit strings, in such a way as to minimize the expected length of the output bit string when the input is a graph distributed according to $\mathbb{G}_n$, and $\mathcal{D}_n$ inverts $\mathcal{C}_n$ and is efficiently computable. A related problem, and the main focus of our paper, seeks to compress graph structures: here, the encoding function $\mathcal{C}_n$ is presented with a multigraph G isomorphic to a sample from $\mathbb{G}_n$, and $\mathcal{D}_n(\mathcal{C}_n(G))$ is only required to be a labeled multigraph isomorphic to G (that is, the labels are "discarded", leaving only the structural information referred to in the title). We again insist on a source code with the minimum possible expected code length (which is given by the Shannon entropy of the distribution on unlabeled graphs induced by $\mathbb{G}_n$, an often non-trivial quantity to estimate; we call this the *structural entropy* of the model).

Our contributions: Succinctly, our contribution in this work is threefold: (i) we give algorithmically efficient, asymptotically optimal algorithms for lossless compression (in the information theoretic sense described above) of both labeled and unlabeled *preferential attachment* graphs[1]; (ii) to establish optimality, we precisely determine the entropies for the distribution on labeled graphs, as well as on graph structures; (iii) and, to analyze our algorithms and give our entropy estimates, we analyze several structural parameters of independent interest (explained more precisely below).

Showing the optimality of our schemes (Theorems 8 and 9) entails new, precise estimates of the Shannon entropy of both the labeled and unlabeled models (Theorems 5 and 6), which in turn require new results on a few quantities of independent interest: the typical size of the automorphism group, as well as some structural characteristics of the *directed* version of the graph (e.g., the number of *admissible labeled representatives* of a given graph structure). Our result on the automorphism group positively settles a conjecture in [13] to the effect that preferential attachment graphs in which each node makes a sufficiently large number of choices are asymmetric with high probability. This completes the characterization of the number of symmetries for a broad range of parameters of the model: when the number of attachment choices m of each vertex is 1, with high probability, there are many symmetries; when $m = 2$, the probability of symmetry is asymptotically positive; and we show in this work that the only symmetry when $m \geq 3$ is the identity with high probability (see Theorem 2).

Regarding structural characteristics of the directed version of the graph (wherein edges are directed from younger nodes to those older nodes that they choose), we analyze a natural partitioning of the vertices into *layers*, which intervenes in the depth-first search process on the directed graph and in the estimation of the number of admissible labeled representatives of the graph (i.e., the number of isomorphic graphs which could have arisen by preferential attachment): in particular, we show that the order of growth of the number of layers is $\Theta(\log n)$ with high probability (see Theorem 4), and almost all vertices occur within the first few layers (Theorem 3). The result on the number of layers is important for our structural compression algorithm. The concentration result allows us to prove that the number of admissible representatives is typically $\exp(n \log n - O(n \log \log n))$ (which should be compared with $n! = e^{n \log n - n + o(n)}$), which intervenes in our derivation of the structural entropy.

[1] Preferential attachment models, though they have limitations, were initially devised to produce graphs with power law degree distributions (frequently observed in the aforementioned applications) via a natural mechanism [3] and continue to be well studied.

In addition to the role that the quantities above play in our analyses of the structural entropy and of our proposed algorithm, our results have implications for another problem of interest [11], wherein we want to recover, as precisely as possible, the original labeling of a randomly relabeled preferential attachment graph.

Finally, in order to obtain our main results, we prove a number of results on the degrees of nodes, as well as on the degree sequence, which may be of interest in other applications.

We provide details of proofs in the appendix. Full proofs can also be found in the journal version [10] of this work. In this conference version, we also present new results on compression algorithms and structural parameters relevant to their analyses.

Prior work: In previous decades, information theory (and the theory of source coding in particular) has dealt primarily with sequences (i.e., samples from a source taking values in a finite alphabet, with some relatively simple dependence structure imposed on distinct sequence elements). Most of the (relatively recent) prior work on compression of non-sequential data sources has either dealt with tree sources (e.g., [16]) or with graphs coming from models with strong edge independence assumptions: optimal compression for the unlabeled Erdős-Rényi model was studied in [5] (see also [1, 2] which handle compression of labeled graphs under the stochastic block model).

An exception is [7], which gives quite general information theoretic results but does not address issues of algorithmic efficiency or compression of unlabeled graphs. We mention also [4], which gives entropy computations for some models with dependent edges but also does not deal with algorithms.

2 Main results

We now introduce the model that we consider and formulate the main results.

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, $2 \le v \le n$, 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 [3]. More precisely, for an integer parameter $m \ge 1$ we define the 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 (with replacement) $v_1, ..., v_m$ of neighbors in G_n as follows: for each vertex $w \le n$ (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). Our proof techniques adapt to tweaks of the model in which multiple edges are not allowed.

For any graph G, we denote by $S(G)$ its unlabeled version (i.e., the equivalence class consisting of all labeled graphs isomorphic to G). Our structural compression/entropy results will be concerned with the unlabeled preferential attachment model, defined by first generating $G \sim \mathcal{PA}(m; n)$, then taking $S(G)$.

2.1 Entropy estimates and structural results

Our first concern will be to derive the fundamental lower bound on the expected code length for compression of unlabeled preferential attachment graphs, as described above. As usual, this is given by the *Shannon entropy* of the distribution on unlabeled graphs induced by $\mathcal{PA}(m;n)$. Recall that for a discrete random variable X with probability mass function $p(\cdot)$, its entropy $H(X)$ is given by $H(X) = -\mathbb{E}_X[\log p(X)]$. We are thus interested in $H(S(G))$, where $G \sim \mathcal{PA}(m;n)$.

By the chain rule for conditional entropy, $H(G) = H(S(G)) + H(G|S(G))$. 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 the automorphism group $|\mathrm{Aut}(G)|$ 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 $H(\sigma|\sigma^{-1}(G)) = H(\sigma|\sigma(G))$ (intuitively, the amount of uncertainty about the value of the random permutation σ upon seeing the result of its application to G) and $\mathbb{E}[\log|\mathrm{Aut}(G)|]$ using the chain rule for entropy, resulting in the following lemma (which, in fact, is not specific to preferential attachment models).

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*

$$H(G) - H(S(G)) = H(\sigma|\sigma(G)) - \mathbb{E}[\log|\mathrm{Aut}(G)|]. \tag{1}$$

To evaluate $H(S(G))$ and to analyze our compression algorithms, we are thus led to evaluate $\mathbb{E}[\log|\mathrm{Aut}(G)|]$, $H(\sigma|\sigma(G))$, and $H(G)$. The next few results give the structural properties that we need for this.

Structural results: The proof of Theorem 6 (our expansion of $H(S(G))$) below and the analyses of our algorithms depend on the following structural results.

The next theorem (whose proof we sketch in Section 3.1 and which we fully prove in the appendix) says that with high probability G has no symmetries, provided that $m \geq 3$. As mentioned in the introduction, this essentially completes the analysis of the precise behavior of the number of symmetries of $\mathcal{PA}(m;n)$ for constant m. For most of this paper, we will focus on the case $m \geq 3$, since the behaviors for $m = 1, 2$ are qualitatively different (for $m = 1, 2$, there are many symmetries with high probability and with asymptotically positive probability, respectively).

Theorem 2 (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 $m \geq 3$, and large n,*

$$P(|\mathrm{Aut}(G)| > 1) = O(n^{-0.004}). \tag{2}$$

We will also state some results on the *directed* version of G (denoted by $\mathrm{DAG}(G)$). This is the directed multigraph defined on $[n]$, with an edge from w to the older node $v < w$ for each edge between v and w in G. We can partition the vertices of $\mathrm{DAG}(G)$ into *levels* inductively 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. The *height* of $\mathrm{DAG}(G)$ is then defined to be the number of levels in this partition.

The next result (proven in Section 3.2) says that almost all of the vertices are concentrated within the first few levels. This will be instrumental in the proof of Theorem 6.

Theorem 3. *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 $\ell \geq \frac{15m}{2\delta^4}\log(3/(2\delta^2))$.*

Next, we find the order of growth of the typical height of DAG(G), which will be useful in the analysis of our structural compression algorithm. We give the proof in Section 3.3.

Theorem 4 (Height of DAG(G)). *Consider $G_n \sim \mathcal{PA}(m;n)$ for fixed $m \geq 1$. Then, with probability at least $1 - o(n^{-1})$, the height of* DAG(G_n) *is at most $Cm \log n$, for some absolute positive constant C.*

It is simple to show that with high probability the height is also lower bounded by $\Omega(\log n)$.

Using these results, we will be able to connect $H(\sigma|\sigma(G))$ in (1) to a combinatorial parameter of DAG(G) (the number of *linear extensions* of DAG(G), viewed as a partial order), which we will be able to show is estimated by $n \log n + R(n)$, where $C_1 n \leq |R(n)| \leq C_2 n \log\log n$.

Entropy results: The final quantity to evaluate in (1) is $H(G)$. Since, in many real applications, n is small enough that $n \log n$ is comparable to n, it is worthwhile (and theoretically interesting) to provide a few terms in the asymptotic expansion of $H(G)$. We give the proof of the following theorem in the appendix.

Theorem 5 (Entropy of preferential attachment graphs). *Consider $G \sim \mathcal{PA}(m;n)$ for fixed $m \geq 1$. We have*

$$H(G) = mn \log n + m \left(\log 2m - 1 - \log m! - A\right) n + o(n), \tag{3}$$

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

This entropy should be compared with the naive method of encoding these graphs, which takes $mn \log(mn) = mn \log n + mn \log m$ space. As $m \to \infty$, compressing to the entropy saves $nm^2 \log m(1 + o_{n,m}(1))$ bits over the naive encoding. For even moderate m (say, $m = 5$ and $n = 10^8$), this is an appreciable difference. This is a more precise analysis than the one given in [14], which only recovers the first term and the order of the second.

Using the above results, we finally have the following expression for $H(S(G))$.

Theorem 6 (Structural entropy of preferential attachment graphs). *Let $m \geq 3$ be fixed. Consider $G \sim \mathcal{PA}(m;n)$. We have*

$$H(S(G)) = (m-1)n \log n + R(n), \tag{4}$$

where $R(n)$ satisfies $Cn \leq |R(n)| \leq O(n \log\log n)$ for some nonzero constant $C = C(m)$.

We sketch the proof of this in Section 3.4 and complete it in the appendix. Compared with the naive encoding method which simply stores a labeled representative of the structure using $mn \log(mn)$ bits, the structural entropy is smaller by $n \log n(1 + o(1))$ bits.

2.2 Optimal compression algorithms

We established (via a variant of Shannon's source coding theorem) in the previous section that *there exist* source codes for unlabeled and labeled graph compression for $\mathcal{PA}(m;n)$ with expected length within one bit of the entropies (4) and (3), respectively. In this section, we give our results on *efficient algorithms* for compression and decompression of unlabeled/labeled samples from $\mathcal{PA}(m;n)$ which asymptotically achieve these bounds.

First, we give an asymptotically optimal algorithm for compression of unlabeled graphs (see Theorem 8 below): that is, given an arbitrary labeled representative G isomorphic to $G' \sim \mathcal{PA}(m;n)$, we construct a code from which $S(G')$ can be efficiently recovered.

Structural compression algorithm (Proof of Theorem 8). Our algorithm starts with finding a certain orientation of the edges of the input graph G to produce a directed graph D which is m-left regular. In the case where G is isomorphic to a sample G' from $\mathcal{PA}(m; n)$ (say, $G = \pi(G')$), we have $D = \pi(\mathrm{DAG}(G'))$.

We accomplish this by a *peeling* procedure: at each step, consider the set D_m of degree-m nodes in the graph. We orient the edges incident on those nodes away from them, and then recurse on the subgraph excluding the nodes in D_m. This procedure terminates precisely when it arrives at a graph consisting of a single vertex with m self-edges.

That this yields the directed graph $D = \pi(\mathrm{DAG}(G'))$ is spelled out in detail in Lemma 2 of [12]. Hence, we are free to apply our structural results (such as Theorem 4) on $\mathrm{DAG}(G')$. We remark that it is not too hard to generalize our algorithm to tweaks of the model, since the only thing that is required is that the height of the resulting directed graph be at most $O(\log n)$; such an orientation of the edges of G exists with high probability, because of Theorem 4.

With this procedure in hand, the structural compression algorithm works as follows, on input G:

1. Construct the directed version $D = \mathrm{DAG}(G)$ by the procedure just described.

2. Starting from the "bottom" vertex (i.e., the vertex with no out-edges except for self-loops), we will do a depth-first search of D (following edges only from their destinations to their sources). To the jth vertex in this traversal, for $j = 1, ..., n$, we will associate a *backtracking number* B_j, which tells us how many steps to backtrack in the DFS process after visiting the jth node; e.g., when there is at least one in-edge leading to an unvisited node (so that we do not backtrack), $B_j = 0$.

   Upon visiting vertex w from vertex v in the DFS, we do the following:

   (a) Encode the names of the $m - 1$ vertex choices made by w, excluding one choice to connect to vertex v. Here, the *name* of a vertex is the binary expansion of its index in the DFS, which we can represent using exactly $\lceil \log n \rceil$ bits. These can be determined in a preprocessing step, by doing an initial DFS to label the nodes with their names.

   (b) We need to know what happens after we visit vertex w: do we go forward in the search, or is there nowhere left to go along the current route (i.e., do we need to backtrack)? Suppose w is the jth vertex to be visited. Then we output an encoding of B_j.

   Now, we need to more precisely examine how we encode these numbers, since it would be suboptimal to simply encode them in $\Theta(\log n)$ bits. Lemma 7 tells us how to more efficiently perform this encoding.

3. For the purposes of decoding, we store (once, for the entire graph) the sequence of code words for the code used for the backtracking numbers. This can be done in at most $O(n \log \log n)$ extra bits, at the beginning of the code.

Lemma 7. *The backtracking numbers* $B_1, ..., B_n$ *can be encoded using a total of* $O(n \log \log n)$ *bits on average.*

Proof. Consider a random variable X whose distribution is given by the empirical distribution of the collection $B = \{B_1, ..., B_n\}$. That is,

$$P_X(x) = \frac{|\{j \ : \ B_j = x\}|}{n} \tag{5}$$

for each x. Note that this empirical distribution is itself a random variable. We will show that $\mathbb{E}[n \cdot H(X)] = O(n \log \log n)$.

Denote by W the event that the number of levels in D is upper bounded by $O(\log n)$. Under conditioning on this event, X can take on at most $O(\log n)$ values, which implies that $H(X) = O(\log \log n)$. Then we have

$$\begin{aligned}\mathbb{E}[H(X)] &\leq \mathbb{E}[H(X)|W] + (1 - \Pr[W])\mathbb{E}[H(X)|\neg W] \\ &\leq \mathbb{E}[H(X)|W] + (1 - \Pr[W]) \log n = O(\log \log n),\end{aligned}$$

where we have used Theorem 4 to upper bound $1 - \Pr[W]$.

We can thus construct a prefix code (once, for the entire graph) for the observed values of B_i, whose empirical average length is given by

$$\sum_{x \,:\, \exists j, B_j = x} \ell_x P_X(x) \leq H(X) + 1,$$

where ℓ_x denotes the length of the code word for x. Now, recalling the definition of $P_X(x)$ in (5), this implies

$$\mathbb{E}\left[\sum_{x \,:\, \exists j, B_j = x} \ell_x |\{j \;:\; B_j = x\}|\right] \leq n\mathbb{E}[H(X)] + n = O(n \log \log n).$$

This completes the proof. ■

The code for $S(G)$ is uniquely decodable, as shown by the decompression algorithm sketched below. Furthermore, its expected length is at most $(m-1)n \log n + O(n \log \log n)$, which recovers the first term of the structural entropy and bounds the second. Let us analyze the running time. Construction of the Huffman code for the backtracking numbers takes time $O(n \log n)$, and each step of the DFS takes time at most $O(m \log n)$. Thus, the running time is $O(mn \log n)$.

We next sketch the decompression algorithm. Given a string $S = s_1 ... s_N$ from the compression algorithm, we produce a labeled graph as follows:

1. Read the prefix of S to recover m, n, and the prefix code for the backtracking numbers. Create a node called 1, with m self-edges. Initialize a stack $U \leftarrow 1$.

2. For $j = 2$ to n,

   (a) Set $x \leftarrow$ the top number on U. Push j onto the stack. Read the next $(m-1)\lceil \log n \rceil$ bits to recover a list of $m - 1$ choices made by vertex j, and append x to this list to produce a list ℓ of m vertices. Output $j \to \ell$.

   (b) Read the codeword for the next backtracking number B, and pop U B times.

The worst-case running time is $O(mn \log n)$: the first step takes time at most $O(n \log n)$ (as the backtracking numbers have length at most $\log n$), the total number of pops of the stack from the backtracking steps is n, and it takes time $\Theta(m \log n)$ to reconstruct the m choices made by each vertex in the loop. Thus, the total time taken in the loop is $O(mn \log n)$, as claimed. Since this algorithm produces precisely the adjacency list encoded by the compression algorithm, the output is a graph isomorphic to the original, with the isomorphism being given by the mapping from each vertex to its DFS number. ■

We have thus proven the following:

Theorem 8 (Structural compression)**.** *There exists an algorithm (given above) which, on input a graph G isomorphic to $G' \sim \mathcal{PA}(m; n)$, runs in time $O(mn \log n)$ and outputs a code of expected length $(m-1)n \log n + O(n \log \log n)$ from which we can recover $S(G)$ in time $O(mn \log n)$.*

Note, from Theorem 6, that our algorithm is optimal at least up to the first term of the lower bound, and we explicitly bound the second term.

For completeness, in Section 3.8 of the appendix, we give an algorithm for compression of labeled graphs, with the details in the appendix: given a labeled graph $G \sim \mathcal{PA}(m;n)$, we construct a code from which G can be recovered in polynomial time.

Theorem 9 (Labeled graph compression)**.** *There exists an algorithm (given in the appendix) which, on input $G \sim \mathcal{PA}(m;n)$, uses $O(n^{m+1})$ arithmetic operations on integers representable with at most $O(n \log n)$ bits and outputs a code of expected length $H(\mathcal{PA}(m;n)) + O(1)$ which is invertible in time $O(n^{m+1})$.*

Note that, while the code is optimal up to an additive constant term and runs in time polynomial in n for each fixed m, the running time may be quite expensive for practical applications. In essence, the algorithm works by arithmetic coding, and we decompose the probability of the input graph as a product of conditional probabilities, each one corresponding to the multiset of choices of connections made by the vertices of G.

3 Proofs

3.1 Proof of Theorem 2

We only sketch the proof of the asymmetry result here. The full proof is in the appendix. Let us define first two properties, $\mathfrak{A}$ and $\mathfrak{B}$ of $\mathcal{PA}(m;n)$ which are crucial for our argument. Here and below we set, for convenience, $k = k(n) = n^{0.01}$.

($\mathfrak{A}$) $\mathcal{PA}(m;n)$ 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 $[t_1 - 1]$.

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

It is easy to see that

$$P(|\mathrm{Aut}(\mathcal{PA}(m;n))| = 1) \ge P(\mathcal{PA}(m;n) \in \mathfrak{A} \cap \mathfrak{B}), \tag{6}$$

and so

$$P(|\mathrm{Aut}(\mathcal{PA}(m;n))| > 1) \le P(\mathcal{PA}(m;n) \notin \mathfrak{A}) + P(\mathcal{PA}(m;n) \notin \mathfrak{B}). \tag{7}$$

Indeed, let us suppose that $\mathcal{PA}(m;n)$ has both properties $\mathfrak{A}$ and $\mathfrak{B}$, and $\sigma \in \mathrm{Aut}(\mathcal{PA}(m;n))$. 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. $|\mathrm{Aut}(\mathcal{PA}(m;n))| = 1$ which proves (6).

Thus, in order to prove Theorem 2 it is enough to show that both probabilities $P(\mathcal{PA}(m;n) \notin \mathfrak{A})$ and $P(\mathcal{PA}(m;n) \notin \mathfrak{B})$ 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

$$\begin{aligned} P(\mathcal{PA}(m;n) \notin \mathfrak{A}) &\le \sum_{k<t_1<t_2} P(t_1, t_2 \text{ choose the same neighbours in } [t_1 - 1]) \\ &\le \sum_{k<t_1<t_2} \sum_{1 \le r_1 \le r_2 \ldots \le r_m < t_1} P(t_1, t_2 \text{ choose } r_1, \ldots, r_m). \end{aligned} \tag{8}$$

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

Let us define $\mathfrak{D}$ as an event that for some $\ell = 1, 2$, and $s = 1, 2, \ldots, m$, $\deg_{t_\ell}(r_s) \le \sqrt{t_\ell / r_s}(\log t_\ell)^3$. Then from Lemma 16 it follows that $P(\mathcal{PA}(m;n) \notin \mathfrak{D}) \le t_1^{-100m}$. Conditioning on $\mathfrak{D}$ and further manipulation shows that for $k < t_1 < t_2$ we get

$$P(t_1, t_2 \text{ choose } r_1, ..., r_m) \le (\log t_2)^{6m} \prod_{\ell=1}^{2} \prod_{s=1}^{m} \frac{1}{\sqrt{t_\ell r_s}} + n^{-m}$$

Thus, (8) becomes, again, after some work, $P(\mathcal{PA}(m;n) \notin \mathfrak{A}) \le k^{2-m}(\log k)^{9m} + n^{-1}$ Hence $P(\mathcal{PA}(m;n) \notin \mathfrak{A}) \le n^{-0.005}$.

Next we show that, with probability close to 1, the $k = n^{0.01}$ oldest vertices of $\mathcal{PA}(m;n)$ have unique degrees and so these are fixed points of every automorphism. The key ingredient of our argument is Lemma 21. To estimate the probability that $\mathcal{PA}(m;n) \notin \mathfrak{B}$, we reason as follows: from Lemma 21 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 17, 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. $\mathcal{PA}(m;n) \notin \mathfrak{B}$.

Finally, Theorem 2 follows directly from (7) and our estimates for $P(\mathcal{PA}(m;n) \notin \mathfrak{A})$ and $P(\mathcal{PA}(m;n) \notin \mathfrak{B})$.

3.2 Proof of Theorem 3

We define $X = X(\epsilon, k)$ to be the number of vertices $w > \epsilon n$ that are at level $\ge k$ in 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 DAG(G). We have the following lemma bounding $\mathbb{E}[X]$:

Lemma 10. *For any $\epsilon = \epsilon(n) > 0$, there exists $k = k(\epsilon)$ for which $\mathbb{E}[X(\epsilon, k)] \le \epsilon n$. In particular, we can take any k satisfying $k \ge 15 \frac{m}{\epsilon^2} \log(3/\epsilon)$.*

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

$$\binom{n}{k} \cdot \frac{\left((5m/\epsilon)\log(3/\epsilon)\right)^k}{n^k} \le \frac{e\left((5m/\epsilon)\log(3/\epsilon)\right)^k}{k^k}$$

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

To complete the proof, we have to extend the above to remove the assumption that vertices are $> \epsilon n$ (i.e., we need to study $Y = Y(k)$, the number of vertices $w \ge 1$ that are at level $\ge k$ in DAG(G)). This is a simple consequence of the above lemma, the fact that $X \le Y \le X + \epsilon n$ with probability 1, and Markov's inequality. This completes the proof.

3.3 Proof of Theorem 4

Let us start with the following, surprising at first sight, observation.

Fact 11. *Let $w < v$. Then the degree $\deg_v(w)$ as well as the probability that v is adjacent to w does not depend on the structure of the graph induced by the first w vertices.* ∎

Let $p_m(n,k)$ denote the probability that $\mathrm{DAG}(G_n)$ contains a path of length k. From Fact 11 and Corollary 1, it follows that

$$
\begin{aligned}
p_m(n,k) &\le \sum_{v_0<v_1<\dots<v_k} \prod_{i=1}^{k} \Pr(v_{i-1}\to v_i) \le \sum_{v_0<v_1<\dots<v_k} \prod_{i=1}^{k} \frac{5m\log(3v_i/v_{i-1})}{\sqrt{v_{i-1}v_i}} \\
&\le \sum_{v_0=1}^{n-k} \frac{1}{\sqrt{v_0}} \prod_{i=1}^{k} \sum_{v_i=v_{i-1}+1}^{n-k-i} \frac{5m\log(3v_i/v_{i-1})}{v_i}\,. \qquad (9)
\end{aligned}
$$

In order to estimate the above sum we split all the vertices $v_1,\dots,v_k$ of the path P into several classes. Namely we say that a vertex v_i is of type t in P if t is the smallest natural number such that $v_i/v_{i-1} \le (1+a)^t$, where a is a small constant to be chosen later, i.e. $t=\lceil \log(v_i/v_{i-1})/\log(1+a)\rceil$. Then, given v_{i-1}, the contribution of terms related to v_i can be estimated from above by

$$
\sum_{v_i=v_{i-1}(1+a)^{t-1}}^{v_{i-1}(1+a)^t} \frac{5m\log(3v_i/v_{i-1})}{v_i} \le 5m\log[(1+a)]\log[3(1+a)^t] \le \alpha t\,, \qquad (10)
$$

where, to simplify notation, we put $\alpha = 5m\log(1+a)\log(3(1+a))$. Let s_t denote the number of vertices of type t in P. Note that $\prod_{t\ge2}\left[(1+a)^{t-1}\right]^{s_t} \le n$ and so

$$
\sum_{t\ge2} t s_t \le 2\sum_{t\ge2}(t-1)s_t \le \frac{2\log n}{\log(1+a)}\,. \qquad (11)
$$

Let us set $J = 2\log n/\log(1+a)$. Thus, we arrive at the following estimate for $p_m(n,k)$

$$
\begin{aligned}
p_m(n,k) &\le \sum_{v_0=1}^{n-k} \frac{1}{\sqrt{v_0}} \binom{k}{s_1} \alpha^{s_1} \sum_{\sum_t s_t t\le J} \binom{k-s_1}{s_2, s_3, \dots, s_k} \prod_{t\ge2} (\alpha t)^{s_t} \\
&\le 3\sqrt{n}\binom{k}{s_1}\alpha^{s_1} \sum_{\sum_t s_t t\le J} \binom{k-s_1}{s_2, s_3, \dots, s_k} \exp\Big(\sum_{t\ge2} s_t\log(\alpha t)\Big) \\
&\le 3\sqrt{n}\binom{k}{s_1}\alpha^{s_1} 2^{2J} \max_{\sum_t s_t t\le J} \exp\Big(\sum_{t\ge2} s_t \log\Big(\frac{e\alpha t(k-s_1)}{s_t}\Big)\Big)\,.
\end{aligned}
$$

In order to estimate the expression

$$
\sigma(J,S) = \max_{\sum_t s_t t\le J} \exp\Big(\sum_{t\ge2} s_t\log\Big(\frac{e\alpha t S}{s_t}\Big)\Big)
$$

where $S=\sum_{t\ge2} s_t$, we split the set of all t's into two parts. Thus, let

$$
T_1 = \{t : \log(e\alpha t S/s_t) \le t\} \quad \text{and} \quad T_2 = \{2,3,\dots,k\}\setminus T_1\,.
$$

Then, clearly,

$$
\max_{\sum_t s_t t\le J} \exp\Big(\sum_{t\in T_1} s_t\log\Big(\frac{e\alpha t S}{s_t}\Big)\Big) \le \max_{\sum_t s_t t\le J} \exp\Big(\sum_{t\in T_1} s_t t\Big) \le \exp(J)\,.
$$

Observe that for every $t\in T_2$ we have $\log(eS\alpha t/s_t)\ge t$ and so $s_t \le e\alpha t e^{-t} S$. It is easy to check that then

$$
s_t \log\Big(\frac{e\alpha t S}{s_t}\Big) \le 6\cdot 2^{-t} S\,,
$$

so

$$\max_{\sum_t s_t t \le J} \exp\left(\sum_{t\in T_2} s_t \log\left(\frac{e\alpha t S}{s_t}\right)\right) \le \max_{\sum_t s_t t \le J} \exp\left(6S\sum_{t\in T_2} 2^{-t}\right) \le \exp(3S) \le \exp(3J)\,.$$

Thus, $\sigma(J,S) \le \exp(4J)$, and, since $s_1 = k - S \ge k - J$,

$$\begin{aligned} p_m(n,k) &\le 3\sqrt{n}\binom{k}{s_1}\alpha^{s_1}2^{2J}\sigma(J,k-s_1) \le 3\sqrt{n}2^k\alpha^{k-J}\exp(6J) \\ &\le 3\exp(\log n + k + (k-J)\log\alpha + 6J)\,. \end{aligned}$$

Since for $0 < a < 1$ we have $a/2 < \log(1+a) < a$, if we set $a = 1/(310m)$, then $\alpha < 1/61$ and $\log\alpha < -4$. Now let us recall that $J = 2\log n/\log(1+a)$ and $k = 5000m\log n > 4J$. Thus,

$$\begin{aligned} p_m(n,k) &\le 3\exp(\log n + k + (k-J)\log\alpha + 6J) \\ &\le 3\exp(\log n + k - 3k + 3k/2) = \exp(\log n - k/2) = o(n^{-1})\,. \end{aligned}$$

3.4 Proof of Theorem 6

We only sketch the derivation of the structural entropy here. The full proof is in the appendix.

We start from Lemma 1. Since Theorem 5 precisely gives $H(G)$, and Theorem 2 implies that $\mathbb{E}[\log|\mathrm{Aut}(G)|] = o(n)$, it remains to estimate $H(\sigma|\sigma(G))$ for a uniformly random $\sigma \in \mathbb{S}_n$. To do this, we show that it can be written in terms of a combinatorial parameter of the directed version of G. To describe it, we make a few (somewhat nontrivial) observations: (i) the probability assigned to any graph g by $\mathcal{PA}(m;n)$ only depends on its unlabeled directed graph structure; (ii) for any unlabeled graph generated by $\mathcal{PA}(m;n)$, there is precisely one positive-probability orientation of the edges (i.e., one unlabeled directed graph structure). The latter is a consequence of the fact that our model starts with a vertex having m self-edges; the full proof does not rely on such small details of the model, and in fact replaces this observation with the more general one that there are at most $2^{O(n)}$ unlabeled directed graphs associated with a given unlabeled, undirected graph.

From observation (ii), we find that $H(G|S(G)) = H(\mathrm{DAG}(G)|S(G)) + H(G|\mathrm{DAG}(G)) = H(G|\mathrm{DAG}(G))$, since $\mathrm{DAG}(G)$ is fully determined by $S(G)$. Then observation (i) says that $H(G|\mathrm{DAG}(G)) = \mathbb{E}[\log|\mathrm{Adm}(G)|]$, where we define $\mathrm{Adm}(G)$ to be the set of labeled graphs isomorphic to G which could have arisen by preferential attachment (we call these the *admissible representatives* of $S(G)$). More formally, a labeled graph could have arisen by preferential attachment if, for any $t \le n$, the subgraph induced by the vertices $\{1,...,t\}$ is such that the degree of vertex t is m.

Now, $|\mathrm{Adm}(G)|$ can be written in terms of $|\mathrm{Aut}(G)|$ and another quantity: $|\Gamma(G)|$, which is the set of permutations $\pi \in \mathbb{S}_n$ such that $\pi(G) \in \mathrm{Adm}(G)$; alternatively, viewing $\mathrm{DAG}(G)$ as a partial order, this is the number of *linear extensions* of $\mathrm{DAG}(G)$. Precisely, we have

$$\mathbb{E}[\log|\mathrm{Adm}(G)|] = \mathbb{E}[\log|\Gamma(G)|] - \mathbb{E}[\log|\mathrm{Aut}(G)|],$$

which implies that $H(\sigma|\sigma(G)) = \mathbb{E}[\log|\Gamma(G)|]$. Thus, to estimate $H(S(G))$, it suffices to estimate $\mathbb{E}[\log|\Gamma(G)|]$. A trivial upper bound is $\mathbb{E}[\log|\Gamma(G)|] \le \log n! = n\log n - n + o(n)$. The lower bound follows by noting that any product of permutations that only permute vertices within levels is a member of $\Gamma(G)$. That is, recalling that L_j denotes the jth level of $\mathrm{DAG}(G)$,

$$|\Gamma(G)| \ge \prod_{j\ge 1}|L_j|!.$$

It follows from Theorem 3 and some work that this lower bound, in turn, is at least $\exp(n\log n - O(n\log\log n))$. Putting all of this together completes the proof.

References

[1] Emmanuel Abbe. Graph compression: the effect of clusters. In *Proc. of the fifty-fourth annual Allerton conference*, 2016.

[2] Emmanuel Abbe and Sergio Verdú. Compressing data on graphs with clusters. In *Proc. of the International Symposium on Information Theory (to appear)*, 2017.

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

[4] David Aldous and Nathan Ross. Entropy of some models of sparse random graphs with vertex-names. *Probability in the Engineering and Informational Sciences*, 28:145–168, 2014.

[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] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006.

[7] Payam Delgosha and Venkat Anantharam. Universal lossless compression of graphical data. In *Proc. of the International Symposium on Information Theory (to appear)*, 2017.

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

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

[10] T. Łuczak, A. Magner, and W. Szpankowski. Asymmetry and structural information in preferential attachment graphs. `http://arxiv.org/abs/1607.04102`.

[11] 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.

[12] Abram Magner, Ananth Grama, Jithin K. Sreedharan, and Wojciech Szpankowski. Times: Temporal information maximally extracted from structure. `http://www.cs.purdue.edu/homes/spa/papers/timesnips2017-subm.pdf`.

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

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

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

[16] Jie Zhang, En-Hui Yang, and John C. Kieffer. A universal grammar-based code for lossless compression of binary trees. *IEEE Transactions on Information Theory*, 60(3):1373–1386, 2014.

Appendix

3.5 Proof of Theorem 2

In this section we shall give a complete proof of Theorem 2. Let us define first two properties, $\mathfrak{A}$ and $\mathfrak{B}$ of $\mathcal{PA}(m;n)$ which are crucial for our argument. Here and below we set, for convenience, $k = k(n) = n^{0.01}$.

($\mathfrak{A}$) $\mathcal{PA}(m;n)$ 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 $[t_1 - 1]$.

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

It is easy to see that

$$P(|\mathrm{Aut}(\mathcal{PA}(m;n))| = 1) \ge P(\mathcal{PA}(m;n) \in \mathfrak{A} \cap \mathfrak{B}), \tag{12}$$

and so

$$P(|\mathrm{Aut}(\mathcal{PA}(m;n))| > 1) \le P(\mathcal{PA}(m;n) \notin \mathfrak{A}) + P(\mathcal{PA}(m;n) \notin \mathfrak{B}). \tag{13}$$

Indeed, let us suppose that $\mathcal{PA}(m;n)$ has both properties $\mathfrak{A}$ and $\mathfrak{B}$, and $\sigma \in \mathrm{Aut}(\mathcal{PA}(m;n))$. Let us assume also that σ is not the identity, and let t_1 be the smallest vertex such that $t_2 = \sigma(t_1) \ne t_1$. 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. $|\mathrm{Aut}(\mathcal{PA}(m;n))| = 1$ which proves (12).

Thus, in order to prove Theorem 2 it is enough to show that both probabilities $P(\mathcal{PA}(m;n) \notin \mathfrak{A})$ and $P(\mathcal{PA}(m;n) \notin \mathfrak{B})$ 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

$$\begin{aligned} P(\mathcal{PA}(m;n) \notin \mathfrak{A}) &\le \sum_{k<t_1<t_2} P(t_1, t_2 \text{ choose the same neighbours in } [t_1 - 1]) \\ &\le \sum_{k<t_1<t_2} \sum_{1\le r_1\le r_2\ldots\le r_m<t_1} P(t_1, t_2 \text{ choose } r_1, \ldots, r_m). \end{aligned} \tag{14}$$

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 $s = 1, 2, \ldots, m$,

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

Then from Lemma 16 it follows that

$$P(\mathcal{PA}(m;n) \notin \mathfrak{D}) \le t_1^{-100m}.$$

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

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

Thus, (14) becomes

$$\begin{aligned}
P(\mathcal{PA}(m;n) \notin \mathfrak{A}) &\leq \sum_{k<t_1<t_2} (\log t_2)^{6m} \sum_{1\leq r_1\leq r_2 \ldots \leq r_m < t_1} \prod_{\ell=1}^{2}\prod_{s=1}^{m} \frac{1}{\sqrt{t_\ell r_s}} + n^{-1} \\
&\leq \sum_{k<t_1<t_2} (t_1t_2)^{-m/2}(\log t_2)^{6m} \sum_{1\leq r_1\leq r_2 \ldots \leq r_m < t_1} \prod_{s=1}^{m} \frac{1}{r_s} + n^{-1} \\
&\leq \sum_{k<t_1} t_1^{-m+1}(\log t_1)^{9m} + n^{-1} \\
&\leq k^{2-m}(\log k)^{9m} + n^{-1}
\end{aligned}$$

Hence

$$P(\mathcal{PA}(m;n) \notin \mathfrak{A}) \leq n^{-0.005}\,. \tag{15}$$

Next we show that, with probability close to 1, the $k = n^{0.01}$ oldest vertices of $\mathcal{PA}(m;n)$ have unique degrees and so these are fixed points of every automorphism. The key ingredient of our argument is Lemma 21.

To estimate the probability that $\mathcal{PA}(m;n) \notin \mathfrak{B}$, we reason as follows: from Lemma 21 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 17, 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. $\mathcal{PA}(m;n) \notin \mathfrak{B}$.

Finally, Theorem 2 follows directly from (13) and our estimates for $P(\mathcal{PA}(m;n) \notin \mathfrak{A})$ and $P(\mathcal{PA}(m;n) \notin \mathfrak{B})$.

3.6 Proof of Theorem 5

In this section we prove Theorem 5 on the entropy of labeled preferential attachment graphs.

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

$$H(G_n) = \sum_{t=1}^{n} H(v_{t+1}|G_t), \tag{16}$$

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:

$$H(\tilde{v}_{t+1}|G_t) = H(\tilde{v}_{t+1}, v_{t+1}|G_t) = H(v_{t+1}|G_t) + H(\tilde{v}_{t+1}|v_{t+1}, G_t), \tag{17}$$

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

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

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

$$H(\tilde{v}_{t+1}|G_t) = m \sum_G P(G_t = G) \sum_{d=m}^{t} N_d(G) p_{t,d} \log(1/p_{t,d}), \tag{19}$$

where $N_d(G)$ denotes the number of vertices of degree d in the fixed graph G, and we define (using the notation of [14])

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

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

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

Plugging this into (16), we get

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

Now, we split the inner sum into two parts:

$$\begin{aligned} H(G_n) + \sum_{t=1}^{n} H(\tilde{v}_{t+1}|v_{t+1}, G_t) = m &\sum_{t=1}^{n} \sum_{d=m}^{\lfloor t^{1/15} \rfloor} \bar{N}_{t,d} p_{t,d} \log(1/p_{t,d}) \\ + m &\sum_{t=1}^{n} \sum_{d=\lfloor t^{1/15} \rfloor + 1}^{t} \bar{N}_{t,d} p_{t,d} \log(1/p_{t,d}). \end{aligned} \tag{22}$$

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 19 to estimate $\bar{N}_{t,d}$ and we use the definition of $p_{t,d}$:

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

Here, the second sum on the left-hand side is the error in approximation incurred by invoking Lemma 19. 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 5.

Upper bounding the large d terms: Our goal is now to show that the second sum of (22), which we denote by E, is $o(n)$.

We apply Lemma 20 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

$$\sum_{t=1}^{n} H(\tilde{v}_{t+1}|G_t) = m\log n! + m(\log 2m - A)n + o(n). \tag{23}$$

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

$$H(\tilde{v}_{t+1}|v_{t+1}, G_t) = \sum_{G, v\in\mathcal{C}_t} P(G_t = G, v_{t+1} = v)H(\tilde{v}_{t+1}|v_{t+1} = v, G_t = G)$$
$$+ \sum_{G, v\notin\mathcal{C}_t} P(G_t = G, v_{t+1} = v)H(\tilde{v}_{t+1}|v_{t+1} = v, G_t = G). \tag{24}$$

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

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

Now we can upper bound the remaining probability in this expression by noting that with high probability, the maximum degree in G_t is $\tilde{O}(\sqrt{t})$ [9]. Using this fact, we have, for arbitrarily small fixed $\epsilon > 0$,

$$\begin{aligned} P(v_{t+1} \in \mathcal{C}_t) = P(v_{t+1} \in \mathcal{C}_t, \text{max. degree of } G_t \leq Ct^{1/2+\epsilon}) \\ + P(v_{t+1} \in \mathcal{C}_t, \text{max. degree of } G_t > Ct^{1/2+\epsilon}) \end{aligned} \quad (25)$$

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 (25) 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 (24) is at most

$$\sum_{G, v \in \mathcal{C}_t} P(G_t = G, v_{t+1} = v) H(\tilde{v}_{t+1}|v_{t+1} = v, G_t = G) = O(t^{-1/2+\epsilon}). \quad (26)$$

We will now show that the second sum in (24), 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,

$$H(\tilde{v}_{t+1}|v_{t+1} = v, G_t = G) = \log m!. \quad (27)$$

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

$$-\sum_{t=1}^{n} H(\tilde{v}_{t+1}|v_{t+1}, G_t) = -n \log m! + o(n). \quad (28)$$

Putting everything together: From (21), (23), and (28), we get

$$H(G_n) = mn \log n + m(\log 2m - 1 - A - \log m!)n + o(n), \tag{29}$$

where A is as in the statement of Theorem 5. ∎

3.7 Proof of Theorem 6

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 2, we immediately have

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

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

$$n \log n - O(n \log \log n) \leq H(\sigma|\sigma(G)) \leq n \log n - n + O(\log n). \tag{30}$$

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 [11]):

$$|\mathrm{Adm}_B(g)| = \frac{|\Gamma_B(g)|}{|\mathrm{Aut}(g)|}. \tag{31}$$

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 12. *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:

$$P(G = g) = \frac{\prod_{d \geq m} \prod_{v \,:\, \deg_g(v)=d} K_g(v) \prod_{j=1}^{k_g(v)} \binom{m}{j} (m + d_1(v) + \cdots d_{j-1}(v))^{d_j(v)}}{\prod_{i=1}^{n-1} (2mi)^m}. \tag{32}$$

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

$$H(\sigma|\sigma(G)) = \mathbb{E}[\log |\Gamma_{\mathrm{UDAG}(G)}(G)|] + O(n). \tag{33}$$

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 13. Now, by Lemma 12, 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 (31). This completes the proof. ∎

Remark 1. *Note that Lemma 14 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,

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

We now use Theorem 3 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

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

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.

3.8 Labeled graph compression algorithm (Proof of Theorem 9)

We will state our algorithm under the assumption that the input $G \sim \mathcal{PA}(m; n)$. We will then remark how to extend it to a broader class of graphs.

Given $G \sim \mathcal{PA}(m; n)$, we compress it using arithmetic coding as follows:

1. We start with the interval $[0, 1]$.
2. For $t = 2$ to n, do the following:
   (a) Sort the choices made by vertex t in increasing order (note that this is trivial to do, because we have the labels). For each vertex $v < t$, we can associate a number $\ell_t(v)$ (which is just the probability, conditioned on G_{t-1}, that vertex t chooses vertex v in a given step), given by $\ell_t(v) = \frac{\deg_{t-1}(v)}{2m(t-1)}$. We can efficiently compute all of these quantities, which we will use to encode the *multiset* of choices made by vertex t. Let us denote the choices made by vertex t by $v_1 < ... < v_{k_t}$, with respective multiplicities $m_1, ..., m_{k_t}$ satisfying $\sum_{j=1}^{k_t} m_j = m$. Then the conditional probability that t made those choices is given by

$$\Pr[t \to \vec{v}|G_{t-1}] = \binom{m}{m_1 \cdots m_{k_t}} \prod_{j=1}^{k_t} \ell_t(v_j)^{m_j}. \tag{34}$$

   So, at time t, we will compute a new interval that encodes the multiset of choices made by t. Consider the set of sorted sequences with m elements coming from $[t-1] = \{1, ..., t-1\}$. We can lexicographically order these sequences, and to each sequence we can assign an interval length equal to (34). Thus, we can compute the left endpoint of the new interval as the sum of the interval lengths for all sorted m-tuples less than the observed one. This sum can be computed in time $O(t^m)$.

   Having the left endpoint and the length of the new subinterval, we then rescale as appropriate, and increment t.

3. At the end of the loop, we are left with a subinterval $[a, b)$, whose length is equal to the probability of the graph G. To see this, let $\vec{v}(t)$ denote the set of vertices chosen by vertex t, for each t. Then we can write

$$\Pr[\mathcal{PA}(m; n) = G] = \prod_{t=2}^{n} \Pr[t \to \vec{v}(t)|G_{t-1}].$$

   Note, then, that the subinterval $[a, b)$ is the result of scaling by $\Pr[t \to \vec{v}(t)|G_{t-1}]$ in the tth step, for $t = 2$ to n. Thus, the length of the subinterval is equal to the probability of the graph, as desired.

   We take the first -$\lfloor \log(b-a) \rfloor$ bits of the binary expansion of the midpoint of this interval, which is known to be sufficient for uniquely identifying it. The expected code length is then $H(\mathcal{PA}(m; n)) + O(1)$, which completes the algorithm.

Note that the worst-case running time (in terms of arithmetic operations) is $\sum_{t=1}^{n} O(t^m) = O(n^{m+1})$, as a result of the evaluations of the left endpoint of the subinterval in the encoding of the multiset of choices of each vertex, and all operations are on integers representable with at most $O(n \log n)$ bits. The result is a polynomial-time algorithm in n for fixed m.

The above algorithm is trivially extended to more general multigraphs by replacing m in the t-loop with the out-degree of vertex t.

3.9 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 15. *For any* v, w,

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

In particular,

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

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

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

Note that

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

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 [9] (it also is a corollary of Lemma 15):

Lemma 16 (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 17. *For $s < t$ we have*

$$P\left[|\deg_t(s) - \mathbb{E}[\deg_t(s)]| > y\right] \le \exp\left(-\frac{y^2}{O(t^{1/2+\epsilon_1}/s^{1/2})}\right) + \exp(-\text{poly}(t)) \tag{35}$$

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 [8].

Lemma 18 (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*

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

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 17.

Proof of Lemma 17. 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 $s + \tau$, $\tau = 1, ..., t - s$, 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 18 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 16, we can upper bound $P(B)$:

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

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 $\tau \in \{0, ..., t-s\}$, $\mathcal{F}_\tau$ to be the σ-field generated by the event B^C and the connection choices of the vertices $s+1, ..., s+\tau$. 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

$$\begin{aligned} X_{\tau+1} - X_\tau =& \sum_{x=s+1}^{s+\tau+1} [\deg(x \to s)|A] + \sum_{x=s+\tau+2}^{t} \mathbb{E}[\deg(x \to s)|\mathcal{F}_{\tau+1}] \\ &- (\sum_{x=s+1}^{s+\tau} [\deg(x \to s)|A] + \sum_{x=s+\tau+1}^{t} \mathbb{E}[\deg(x \to s)|\mathcal{F}_\tau]) \\ =& [\deg((s+\tau+1) \to s)|A] - \mathbb{E}[\deg(s+\tau+1) \to s|\mathcal{F}_\tau] \qquad (36) \\ &+ \sum_{x=s+\tau+2}^{t} (\mathbb{E}[\deg(x \to s)|\mathcal{F}_{\tau+1}] - \mathbb{E}[\deg(x \to s)|\mathcal{F}_\tau]). \end{aligned}$$

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 (36). 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

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

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 (36). The first two terms may be easily upper bounded by $2m$:

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

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 18, we find that

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

for any $y \le 2V/O(m)$, 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 [15].

Lemma 19 (Expected value of $N_{t,d}$). *We have, for $t \ge 1$ and $1 \le d \le t$ and for any fixed $m \ge 1$,*

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

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

$$\bar{N}_{t,d} = O\left(\frac{t}{d(d+1)(d+2)}\right) = O\left(\frac{t}{d^3}\right). \tag{37}$$

Proof. We will prove the claimed upper bound by induction on the number of edge connection choices made so far in the graph (e.g., after vertex t has made all of its choices, this number is mt).

Let us define $\bar{M}_{\tau,d}$ to be the expected number of vertices with degree d in the graph after τ vertex choices have been made in the graph. Note that $\bar{M}_{\tau,d} = \bar{N}_{\tau/m,d}$ whenever τ is divisible by m. Thus, to prove our desired result, it is sufficient to prove that

$$\bar{M}_{\tau,d} = O\left(\frac{\{\tau\}_m}{d(d+1)(d+2)}\right) \tag{38}$$

for $\tau \to \infty$ and $d \geq (\tau/m)^{1/15}$ (for convenience, we denote by $\{\tau\}_m$ the largest integer $\leq \tau$ that is divisible by m). The base case is provided by Lemma 19.

Next, note that $\bar{M}_{\tau,d}$ satisfies the following recurrence:

$$\begin{aligned}\bar{M}_{\tau,d} &\leq \bar{M}_{\tau-1,d}\left(1 - \frac{d-m}{2\{\tau\}_m}\right) + \bar{M}_{\tau-1,d-1}\frac{d-1}{2\{\tau\}_m} - \bar{M}_{\tau-1,d}\frac{d-m}{2\{\tau\}_m} \\ &= \bar{M}_{\tau-1,d}\left(1 - \frac{d-m}{\{\tau\}_m}\right) + \bar{M}_{\tau-1,d-1}\frac{d-1}{2\{\tau\}_m}. \end{aligned} \tag{39}$$

This is because an m-tuple that has degree d after choice τ either had degree d after choice $\tau - 1$ and wasn't chosen by the τth choice, or had degree $d-1$ and was chosen by choice τ. Moreover, any m-tuple with degree d at time $\tau - 1$ that *was* chosen by choice τ no longer has degree d. The upper bound is a result of the specific details of our model but may be generalized.

Next, we apply the inductive hypothesis, resulting in

$$\bar{M}_{\tau,d} \leq \frac{C\{\tau-1\}_m}{d(d+1)(d+2)}\left(1 - \frac{d-m}{\{\tau\}_m}\right) + \frac{C\{\tau\}_m}{(d-1)d(d+1)}\frac{d-1}{2\{\tau\}_m} \tag{40}$$

$$\leq \frac{C\{\tau-1\}_m}{d(d+1)(d+2)}\left(1 - \frac{d-m}{\{\tau\}_m}\right) + \frac{C}{2d(d+1)}, \tag{41}$$

for some positive constant $C(m) = C$. This can be rearranged to yield

$$\bar{M}_{\tau,d} \leq \frac{C\{\tau-1\}_m}{d(d+1)(d+2)} + \frac{C}{2d(d+1)} - \frac{C\{\tau-1\}_m(d-m)}{d(d+1)(d+2)\{\tau\}_m}. \tag{42}$$

To continue, we split into two cases: either $\{\tau-1\}_m = \{\tau\}_m$ or $\{\tau-1\}_m = \tau - m = \{\tau\}_m - m$. In the first case, (42) becomes

$$\bar{M}_{\tau,d} \leq \frac{C\{\tau\}_m}{d(d+1)(d+2)} + \frac{C}{2d(d+1)} - \frac{C}{(d+1)(d+2)} + \frac{Cm}{d(d+1)(d+2)}.$$

Now, provided that τ is large enough, and since d is $\Omega(\tau^{1/15})$, the sum of the last three factors is negative, so that

$$\bar{M}_{\tau,d} \leq \frac{C\{\tau\}_m}{d(d+1)(d+2)},$$

as desired.

Now we handle the second case (where $\tau - 1_m = \{\tau\}_m - m$):

$$\begin{aligned}\bar{M}_{\tau,d} &\le \frac{C\{\tau\}_m - Cm}{d(d+1)(d+2)} + \frac{C}{2d(d+1)} - \frac{C(\{\tau\}_m - m)(d-m)}{d(d+1)(d+2)\{\tau\}_m} \\ &= \frac{C\{\tau\}_m}{d(d+1)(d+2)} + \frac{C}{2d(d+1)} - \frac{Cd}{d(d+1)(d+2)} + \frac{Cm(d-m)}{d(d+1)(d+2)\{\tau\}_m} \\ &\le \frac{C\{\tau\}_m}{d(d+1)(d+2)} + \frac{C}{2d(d+1)} - \frac{Cd}{d(d+1)(d+2)} + \frac{Cmd}{d(d+1)(d+2)\{\tau\}_m}.\end{aligned}$$

We then proceed exactly as in the previous case, which completes the proof. ■

The next result, a corollary of Lemma 15, gives an upper bound on the probability that two given vertices are adjacent.

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 $v, w \ge \epsilon n$ 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

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

When $d \le d_0 = 8m\sqrt{v/w}\log(3v/w)$ the above sum is clearly smaller $d_0/2 = 4m\sqrt{1/vw}\log(3v/w)$. If $d \ge d_0$ one can use Lemma 15 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 21. *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

$$\begin{aligned}P(\deg_n(s) = \deg_n(s')) = &\sum_{\underline{d},\underline{d}',\underline{\underline{d}}'} P(\deg_n(s) = \deg_n(s') | \underline{\deg}(s) = \underline{d}, \underline{\deg}(s') = \underline{d}', \underline{\underline{\deg}}(s') = \underline{\underline{d}}') \\ &\qquad \times P(\underline{\deg}(s) = \underline{d}, \underline{\deg}(s') = \underline{d}', \underline{\underline{\deg}}(s') = \underline{\underline{d}}') \\ = &\sum_{\underline{d},\underline{d}',\underline{\underline{d}}'} P(\underline{\underline{\deg}}(s) = \underline{d}' + \underline{\underline{d}}' - \underline{d} | \underline{\deg}(s) = \underline{d}, \underline{\deg}(s') = \underline{d}', \underline{\underline{\deg}}(s') = \underline{\underline{d}}') \\ &\qquad \times P(\underline{\deg}(s) = \underline{d}, \underline{\deg}(s') = \underline{d}', \underline{\underline{\deg}}(s') = \underline{\underline{d}}')\,. \end{aligned} \tag{43}$$

Observe that due to Lemma 17, 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

$$\left|\frac{\underline{d}n'}{2mn} - \underline{\underline{\deg}}(s)\right| = \left|0.5m\underline{d}n^{-0.4} - \underline{\underline{\deg}}(s)\right| \le \left(\frac{\underline{d}n'}{2mn}\right)^{0.6} \le n^{0.08}\,. \tag{44}$$

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

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

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

$$\frac{m \deg_{v-1}(s)}{2m(v-1)} = \frac{\underline{d} + O(\underline{d}n^{-0.4})}{2(n - O(n'))} = \frac{\underline{d}}{2n}\Big(1 + O(n^{-0.4})\Big). \tag{45}$$

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 $O(k^4 n^{-.044}) = O(n^{-.004})$. ∎