On Symmetry of Uniform and Preferential Attachment Graphs*

Abram Magner† Svante Janson‡ Giorgos Kollias§
Wojciech Szpankowski¶

Submitted: May 30, 2014; Accepted: Aug 13, 2014; Published: Aug 28, 2014
Mathematics Subject Classifications: 05C80, 05C82

Abstract

Motivated by the problem of graph structure compression under realistic source models, we study the symmetry behavior of preferential and uniform attachment graphs. These are two dynamic models of network growth in which new nodes attach to a constant number m of existing ones according to some attachment scheme. We prove symmetry results for $m = 1$ and 2, and we conjecture that for $m \geqslant 3$, both models yield asymmetry with high probability. We provide new empirical evidence in terms of graph defect. We also prove that vertex defects in the uniform attachment model grow at most logarithmically with graph size, then use this to prove a weak asymmetry result for all values of m in the uniform attachment model. Finally, we introduce a natural variation of the two models that incorporates preference of new nodes for nodes of a similar age, and we show that the change introduces symmetry for all values of m.

1 Introduction

Study of the asymptotic behavior of the symmetries of random graphs, originally motivated by combinatorial problems, has relatively recently found a new application in the

*This work was supported by NSF Center for Science of Information (CSoI) Grant CCF-0939370, NSA Grant H98230-11-1-0141, and in addition NSF Grants DMS-0800568, and CCF-0830140, and the MNSW grant DEC-2013/09/B/ST6/02258.

†Department of Computer Science, Purdue University, Indiana, U.S.A. anmagner@purdue.edu

‡Department of Mathematics, Uppsala University, Uppsala, Sweden. S. Janson partly supported by the Knut and Alice Wallenberg Foundation. svante@math.uu.se

§Thomas J. Watson Research Center, Yorktown Heights, New York, U.S.A. gkollias@us.ibm.com

¶Department of Computer Science, Purdue University, Indiana, U.S.A.; also with the Faculty of Electronics, Telecommunications and Informatics, Gdańsk University of Technology, Poland. spa@cs.purdue.edu

problem of compression of graph structures. The basic problem can be formulated as follows: given a probability distribution on labeled graphs, determine an encoding of graph structures (that is, *unlabeled* graphs) so as to minimize expected description length.

Choi and Szpankowski (2012) 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.

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 Rashevsky (1955) and Trucco (1956), 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. Furthermore, tools developed here will allow us to study and compare topological information of nodes (i.e., by how many bits a graph view of one node differs from another).

The present paper is a first step toward the goal of extending graph structure compression results to other random graph models. In particular, many real-world graphs, such as biological and social networks, exhibit a power law degree distribution (see Durrett (2006)). To explain this phenomenon, Albert and Barabási (2002) proposed the *preferential attachment* mechanism, in which a graph is built one vertex at a time, and each new vertex t attaches to a given old vertex v with probability proportional to the current degree of v. Thus, we study a variant of a preferential attachment model. The primary problem appears to be difficult, so we also study a closely related model in which attachment is uniform, in the hope that the proof techniques used there may be generalized. In both uniform and preferential attachment models, we prove that when each new vertex chooses only one previous vertex as a neighbor, there is symmetry with high probability, and when each new vertex makes two choices, there is a positive probability of symmetry. In addition, we determine the asymptotic behavior of a quantity known as the *defect* of a vertex, introduced by Kim et al. (2002), which measures the extent to which the neighborhood of the vertex contributes to asymmetry of the graph. We then use this to prove a weak asymmetry result for the uniform attachment case.

We also introduce the *sliding window* model, a dynamic model in which new vertices choose neighbors from within windows of expected size uniformly bounded above by a constant, the purpose being to exhibit a "natural" mechanism that, coupled with a quite general attachment scheme that includes preferential and uniform attachment as special cases, results in symmetry with asymptotically positive probability.

Study of the asymptotic behavior of the automorphism group of a random graph started with Erdős and Rényi (1963), wherein Erdős-Rényi graphs with constant connec-

MOST WIEDZY Downloaded from mostwiedzy.pl

tion probability were shown to be 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 $n \to \infty$. A similar question motivated the investigation of symmetry properties of random regular graphs by Bollobás (1982) and Kim et al. (2002). In the latter paper, the authors precisely characterized the range for which Erdős-Rényi graphs are asymmetric by proving concentration results for random variables defined in terms of vertex defect. They then proved an asymmetry result for random regular graphs using the previous result.

For general models, symmetry and asymmetry results can be nontrivial to prove, due to the non-monotonicity of the properties considered. Furthermore, the particular models considered here present more difficulties not seen in the Erdős-Rényi case: there is significant dependence between edge events, and graph sparseness makes derivation of concentration results difficult.

The rest of the paper is organized as follows: in Section 2, we formally state the models and the main problem; we then state the main results, along with a discussion of their significance. We also present some empirical validation of the symmetry results, as well as evidence in support of the asymmetry conjecture. Finally, in Section 3, we give the proofs.

2 Main Results

In this section, we state the main problems, introduce the models that we consider, and formulate the main results. First, we introduce some standard graph-theoretic terminology and notation. We start with the notion of structure-preserving transformations between labeled graphs: given two graphs 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 $x, y \in V(G_1)$, there is an edge between x and y if and only if there is an edge in G_2 between $\phi(x)$ and $\phi(y)$. When such a ϕ exists, G_1 and G_2 are said to be *isomorphic*; that is, they have the same structure.

An isomorphism from the vertices of a graph G to itself is called an *automorphism* or *symmetry*. 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 $\mathrm{Aut}(G)$. Note that the image of G under any of its symmetries is G, the same labeled graph.

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. Intuitively, G is symmetric if and only if there are at least two vertices whose graph perspectives are the same; that is, their neighborhoods at any distance have the same structure.

The main problem can then be stated as follows: given a random graph process $\{G_n\}_{n \geqslant 1}$, characterize the behavior of its automorphism group for $n \to \infty$.

MOST WIEDZY Downloaded from mostwiedzy.pl

2.1 Definitions of Models

In what follows, vertices of an n-vertex graph are the elements of the set $[n] := \{1, \ldots, n\}$.

A preferential attachment model is a dynamic model of network growth in which new vertices, when they choose vertices already in the graph as neighbors, have a preference for a given vertex that is proportional to its current degree, see Albert and Barabási (2002). Thus, nodes with high degree tend to be preferred for new connections. The following definition formalizes this. (A slightly different formalization of the Barabási–Albert model is given by Bollobás and Riordan (2004).)

Definition 1 (Preferential attachment model)**.** *A preferential attachment graph $\mathcal{P}(n, m)$ on n vertices, with parameter m, is constructed as follows: at time $t = 1$, a single vertex with name 1 and attractiveness $\mathrm{att}_1(1) = 0$ is added. For each time $1 < t \leqslant n$, a vertex with name t is added, and m vertices $c_{t1}, \ldots, c_{tm}$ in $[t-1]$ are chosen with replacement such that*

$$\Pr[c_{tj} = v] = \frac{\mathrm{att}_v(t-1)}{\sum_{w=1}^{t-1} \mathrm{att}_w(t-1)} = \frac{\mathrm{att}_v(t-1)}{2m(t-2)}.$$

(Here we adopt the convention that $0/0 = 1$.) An edge between t and v is added if and only if $c_{ti} = v$ for some i. For each $v \in [t-1]$, we set $\mathrm{att}_v(t) = \mathrm{att}_v(t-1) + |\{j | c_{tj} = v\}|$. Finally, we set $\mathrm{att}_t(t) = m$.

Another way to express this is to first construct a growing multigraph, where we at each step add one new vertex and m edges from it, with the other endpoints chosen at random with replacement as above; then $\mathrm{att}_v(t)$ equals the degree of v at time t. We then reduce any set of multiple edges to a single edge to obtain the simple graph $\mathcal{P}(n, m)$.

We will also consider a variation, which we call the *uniform attachment* model, with the only change being that vertex choices are now equiprobable; that is, we now fix $\mathrm{att}_v(t-1) = 1$ for all t and $v \in [t-1]$. (For $m = 1$ this yields the well-known *random recursive tree* (see Smythe and Mahmoud (1995)).) The rationale for studying this simpler model is that solving our symmetry problems poses many of the same challenges for both models: both, for example, generate sparse graphs, which seems to rule out an approach to proving asymmetry based on defect (discussed below); furthermore, in both models, in considering a neighborhood of a vertex, one must distinguish between incoming and outgoing vertices, which complicates other possible approaches to asymmetry proofs. On the other hand, the uniform attachment model is advantageous, in that we need not deal with the dependence resulting from the preferential attachment mechanism.

We also study another practical variant of the attachment model called the sliding window model that we define next.

Definition 2 (Sliding window model)**.** *The sliding window model with random window size works as follows: at time 1, vertex 1 is added. At time $t > 1$, vertex t is added, and a window size W_t, taking values in $\{2, \ldots, t-1\}$, is chosen according to the distribution function F_t, independent of anything else. Then, m vertices are chosen with replacement from the set $[t - W_t, t-1]$ (which we call the* window *of vertex t), with the distribution*

MOST WIEDZY Downloaded from mostwiedzy.pl

of each choice c_{tj} determined by the ratio of the attractiveness of any node in the window to the total attractiveness of the window. That is,

$$\Pr[c_{tj} = v | W_t = w] = \frac{\mathrm{att}_v(t-1)}{\sum_{k=t-w}^{t-1} \mathrm{att}_k(t-1)}.$$

Here, $\mathrm{att}_v(x)$ *denotes the attractiveness of vertex v at time x. In the preferential attachment sliding window model,* $\mathrm{att}_v(x)$ *is given by*

$$\mathrm{att}_v(x) = m + \sum_{k=1}^{x-1} \sum_{j=1}^{m} [c_{kj} = v].$$

In the uniform attachment version, $\mathrm{att}_v(x) = 1$.

2.2 Statement of Results

The first result characterizes the expected vertex defect for the uniform attachment model. Vertex, permutation, and graph defect were introduced by Kim et al. (2002) in order to prove asymmetry for Erdős-Rényi graphs. The definitions are as follows.

Definition 3 (Defect)**.** *Fix a graph G on n vertices. Given a permutation $\pi \in S_n$ and $u \in [n]$, we define the defect of u with respect to π to be*

$$D_\pi(u) = |N(\pi(u)) \triangle \pi(N(u))|,$$

where $N(x)$ denotes the set of neighbors of vertex x, and $\triangle$ denotes the symmetric difference of two sets. We define the defect of π to be

$$D_\pi(G) = \max_{u \in [n]} D_\pi(u).$$

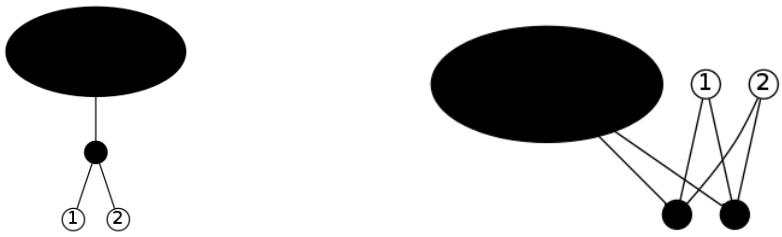
Finally, we define the defect of G to be

$$D(G) = \min_{\pi \neq \mathrm{ID}} D(\pi).$$

Some simple consequences of these definitions are as follows: for a graph G on n vertices and a permutation $\pi \in [n]$, π is an automorphism of G if and only if $D_\pi(G) = 0$, which is equivalent to non-existence of a vertex $u \in [n]$ such that $D_\pi(u) \neq 0$; G has nontrivial symmetries if and only if $D(G) = 0$. Thus, vertex defect measures the extent to which a given vertex's neighborhood structure breaks the symmetry of π, permutation defect measures the number of edges adjacent to any particular vertex that need to be modified in order to make π a symmetry of G, and graph defect measures the number of edges adjacent to any vertex that need to be modified in order to introduce a nontrivial symmetry into G.

For a nontrivial permutation $\pi \in S_n$ and any $u \in [n]$, define $\omega(\pi, u) = \min\{u, \pi(u)\}$. We also define $\omega(\pi)$ to be the minimum vertex not fixed by π (where we say that π *fixes* a vertex w if $\pi(w) = w$).

MOST WIEDZY Downloaded from mostwiedzy.pl



(a) (b)

Figure 1: The shapes on which we focus for the proof of Theorem 3.

Theorem 1 (Expected defect for a vertex)**.** *Fix $m \in \mathbb{N}$ in the uniform attachment model. For any n sufficiently large, $\pi \neq \mathrm{ID}, \pi \in S_n$, and $u \in [n]$ not fixed by π,*

$$\log\left(\frac{n}{\max\{\omega(\pi,u)+2,(2m+2)\}}\right) \leqslant \mathbb{E}[D_\pi(u)] \leqslant 1+4m\left(2+\log\left(\frac{n}{\omega(\pi,u)}\right)\right).$$

This theorem is significant for two reasons: it plays a central role in the proof of Theorem 2, and it gives an indication that an approach to an asymmetry proof via defects, as used in the setting of Erdős-Rényi graphs (see Kim et al. (2002)), may not be fruitful. A key difference between the Erdős-Rényi model and the uniform and preferential attachment ones is that the expected defect in the former is $\Theta(np(1-p))$ for $p, 1-p \gg \frac{\log n}{n}$, which is essential for the proof technique used for that model to work.

The previous theorem can be used to derive a *weak asymmetry* result for the uniform attachment model as follows: for a given sequence of permutations $\pi_n \neq \mathrm{ID}$, to show that $\pi_n \notin \mathrm{Aut}(G_n)$ with high probability, it is sufficient to exhibit a sequence of vertices u_n such that $\lim_{n\to\infty} \Pr[D_{\pi_n}(u_n) = 0] = 1$. In particular, we can choose $u_n = \omega(\pi_n)$, the minimum non-fixed vertex of π_n. We prove the following result.

Theorem 2 (Probability of vertex defect being **0**)**.** *Fix $m \geqslant 1$ and consider a sequence of graphs in the uniform attachment model $G_n \sim \mathcal{U}(n,m)$. Let $\{\pi_n\}_{n=1}^\infty, \pi_n \in S_n - \{\mathrm{ID}\}$, and, for each n, let $u_n = \omega(\pi_n)$. Then*

$$\Pr[D_{\pi_n}(u_n) = 0] \xrightarrow{n\to\infty} 0,$$

so that the asymptotic probability that $\pi_n \in \mathrm{Aut}(G_n)$ is 0.

We remark that we call this a weak asymmetry result because it is a statement about which permutations are *not* in the automorphism group of G_n: any given sequence of permutations (or small sets of permutations) is asymptotically not likely to be in the automorphism group of a growing uniform attachment graph. Thus, this result has the flavor of an asymmetry result.

Observe that Theorem 2 does not prove asymmetry of a uniform attachment model. For this we would need to prove that the *graph defect* $D(G) > 0$ whp. However, we are

MOST WIEDZY Downloaded from mostwiedzy.pl

able to make some statements about symmetry/asymmetry of this model. We discuss it next.

In the case $m = 1$, both the uniform and preferential attachment process yield trees. In such trees, we count the number of vertices that are parents of precisely two nodes, both leaves (which some authors have called a *cherry*), as illustrated in Figure 1(a). The two children may be swapped, thereby forming a symmetry. We find, via a Pólya urn approach, that the number of cherries is positive with high probability, which gives symmetry with high probability in both models. Furthermore, as a result of the methods that we apply, we are able to compute the asymptotic fraction of simple cherries, which turns out to be approximately 4/105 and 1/24 in the preferential and uniform cases, respectively (see Section 3.3).

The case $m = 2$ is midway between the high-probability symmetry of the $m = 1$ case and the conjectured asymmetry of the $m \geqslant 3$ case. Examining the asymptotic probability of two vertices making the same choices and being unchosen by subsequent vertices yields the following results as shown in Figure 1(b).

Theorem 3 (Symmetry results for $\boldsymbol{m = 1, 2}$)**.** *Fix $m = 1, 2$, and let $G_n \sim \mathcal{U}(n, m)$ or $G_n \sim \mathcal{P}(n, m)$. Then there exists a constant $C > 0$ such that, for n sufficiently large,*

$$\Pr[|\mathrm{Aut}(G_n)| > 1] > C.$$

In the result for $m = 1$, for both models, we can strengthen the statement to symmetry with high probability (that is, the statement is true for all $C < 1$).

We conjecture that for $m = 2$, in both models, $\Pr[|\mathrm{Aut}(G_n)| > 1]$ converges to a constant strictly less than 1.

The result for $m = 2$ is particularly interesting in light of the fact that empirical investigations of the symmetries of $\mathcal{U}(n, 2)$ graphs with insufficiently many samples (cf. Figure 2) may lead to the incorrect conclusion that there is asymmetry with high probability in this case.

For fixed $m \geqslant 3$, we propose the following conjecture.

Conjecture 1 (Asymmetry conjecture)**.** *Fix $m \geqslant 3$ and let either $G_n \sim \mathcal{U}(n, m)$ or $G_n \sim \mathcal{P}(n, m)$. Then*

$$\Pr[|\mathrm{Aut}(G_n)| > 1] \xrightarrow{n \to \infty} 0.$$

That is, graphs drawn according to the specified distributions are asymmetric with high probability.

Empirical evidence in support of this conjecture abounds. For instance, MacArthur and Anderson (2006) give plots of number of automorphisms as n increases for sampled graphs, which show initial increase and then swift decay to 1. We contribute defect-based evidence here.

Figure 2 shows growth of a graph defect estimate as n, the number of vertices of the sampled graphs, grows large, for a few values of m. As only a small subset of permutations could be sampled due to time and space constraints, the pictured defect estimates only

MOST WIEDZY Downloaded from mostwiedzy.pl

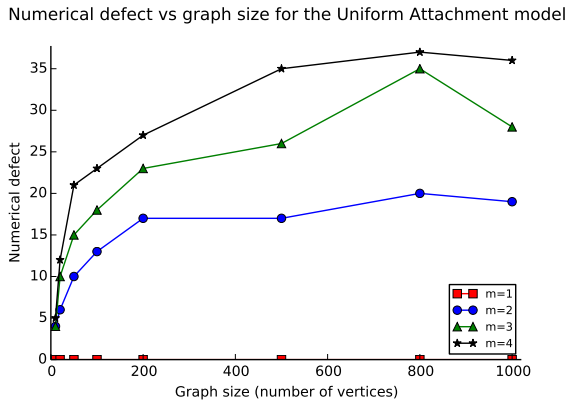


Figure 2: Plots showing minimum/maximum vertex defects for certain classes of permutations.

give upper bounds on the true defects. For $m = 1$, the estimate quickly drops to 0, due to the presence of automorphisms that are swaps of two vertices, as the proof of Theorem 3 indicates. For $m = 2$, the estimate grows away from 0, but this does not give a complete picture of the situation in this case: it fails to capture the phenomenon of symmetry with asymptotically nonzero (but quite small) probability predicted by Theorem 3. For $m \geqslant 3$, the graph defect exhibits logarithmic growth, which is in keeping with the statement of Theorem 1. Furthermore, since the defect grows away from 0, the evidence is in keeping with the asymmetry conjecture (though the small permutation sample size prevents us from claiming it as strong evidence of the conjecture).

We also give some weak supporting evidence in the form of a theorem about probability of automorphism group membership for sequences of permutations, that of Theorem 2.

Finally, we discuss the sliding window model, that could naturally capture the behavior of dynamic networks in which new nodes are very unlikely to attach to old ones, but whose attachment policy is otherwise quite general. (For example, one may think of a social network whose nodes are people admitted to a university and whose edges represent friendships formed after admission, and then, except with small probability, nodes will choose neighbors in a window of bounded size.) The next result deals with symmetry in the sliding window model. If windows are restricted to be of expected length less than a constant bound, then considering the event that nodes $n-1$ and n form a cherry shows that symmetry results with nonzero probability.

Theorem 4 (Symmetry results for sliding window model)**.** *In the sliding window model with random window size, for any m, if there exists a constant w such that $\mathbb{E}[W_i] \leqslant w$ for all i, then the probability of symmetry is asymptotically positive. If there exists w such that, for all i, $W_i \leqslant w$ with probability 1, then a graph drawn according to this distribution is symmetric with high probability.*

MOST WIEDZY Downloaded from mostwiedzy.pl

3 Proofs

In this section, we fix some useful notation, then give proofs of the theorems. Given two vertices u and v, we write $E[u, v]$ for the event that there is an (undirected) edge between u and v. For vertex $u \in [n]$ and permutation π, we can write the defect $D_\pi(u)$ as:

$$D_\pi(u) = \sum_{v=1}^{n} B_{u,\pi}(v),$$

where we define $B_{u,\pi}(v)$ to be 1 if $v \in N(\pi(u)) \triangle \pi(N(u))$ and 0 otherwise. We can express each such indicator in terms of edge events:

$$B_{u,\pi}(v) = (E[v, \pi(u)] \cap \neg E[\pi^{-1}(v), u]) \cup (\neg E[v, \pi(u)] \cap E[\pi^{-1}(v), u]).$$

Note that the two conjunctions are disjoint.

3.1 Proof of Theorem 1

We now assume uniform attachment model. First, we state some useful lemmas about probabilities of edge events. We omit the simple proofs.

Lemma 1. *For all i, q, j, r such that $i < j$ and $q < r$, if $j < r$, then* $\Pr[E[i, j]] > \Pr[E[q, r]]$.

Lemma 2. *For all $i < j$,*

$$\frac{1}{j} \leqslant \Pr[E[i, j]] \leqslant \frac{2m}{j}.$$

Lemma 3. *For all i, q, j, r such that $i < j$, $q < r$, $r \geqslant 2m + 1$, and either $i \neq q$ or $j \neq r$,*

$$\Pr[E[i, j] \cap \neg E[q, r]] \geqslant \Pr[E[i, j] \cap E[q, r]].$$

Lemma 4. *For all $x \geqslant 1$, $\sum_{i=1}^{x} \Pr[E[i, x]] \leqslant 2m$.*

Lemma 5 (Harmonic Sum Log Sandwich)**.** *For all n and $j \in \mathbb{Z}$ such that $1 \leqslant j \leqslant n$,*

$$\log \frac{n}{j} \leqslant \sum_{i=j}^{n} \frac{1}{i} \leqslant \frac{1}{j} + \log \frac{n}{j}.$$

Now we move on to the proof of the main result. Throughout, we assume that $u < \pi(u)$; the case $u > \pi(u)$ follows from this by noting that $D_{\pi^{-1}}(\pi(u)) = D_\pi(u)$. First, we derive the lower bound. We start by lower bounding the probability of event $B_{u,\pi}(i)$ by the probability of an edge. For any vertex i such that $\pi^{-1}(i) \geqslant 2m + 1$ and $i \neq u, \pi(u)$ (so all but a constant number of them), we have

$$\Pr[B_{u,\pi}(i)] = \Pr[E[i, \pi(u)] \cap \neg E[\pi^{-1}(i), u]] + \Pr[\neg E[i, \pi(u)] \cap E[\pi^{-1}(i), u]]$$

MOST WIEDZY Downloaded from mostwiedzy.pl

$$\begin{aligned}
&\overset{(a)}{\geqslant} \Pr[E[i,\pi(u)] \cap E[\pi^{-1}(i),u]] + \Pr[E[\pi^{-1}(i),u] \cap \neg E[i,\pi(u)]] \\
&= \Pr[E[\pi^{-1}(i),u]].
\end{aligned}$$

Here, (a) is a result of Lemma 3 since $\max\{\pi^{-1}(i),u\} \geqslant 2m+1$. Hence,

$$\begin{aligned}
\sum_{i=1}^{n} \Pr[B_{u,\pi}(i)] &\geqslant \sum_{\pi^{-1}(i)=\max\{u+1,(2m+1)\}}^{n} \Pr[B_{u,\pi}(i)] \\
&\overset{(a)}{\geqslant} \sum_{\pi^{-1}(u)\neq\pi^{-1}(i)=\max\{u+1,(2m+1)\}}^{n} \Pr[E[\pi^{-1}(i),u]] \\
&\overset{(b)}{\geqslant} \sum_{\pi^{-1}(u)\neq\pi^{-1}(i)=\max\{u+1,(2m+1)\}}^{n} \frac{1}{\pi^{-1}(i)} \\
&\overset{(c)}{\geqslant} \log\left(\frac{n}{\max\{u+2,(2m+2)\}}\right) = \log\left(\frac{n}{\max\{\omega(\pi,u)+2,2m+2\}}\right),
\end{aligned}$$

where (a) is a consequence of the previous inequality, (b) is an invocation of Lemma 2, and (c) is a result of Lemma 5 and the observation that, if $\pi^{-1}(u) \geqslant \max\{u+1,2m+1\}$, then its contribution to the sum is $\frac{1}{\pi^{-1}(u)} \leqslant \frac{1}{\max\{u+1,2m+1)\}}$. This completes the proof of the lower bound.

Now we prove the upper bound. We start by upper bounding the probability of $B_{u,\pi}(i)$:

$$\begin{aligned}
\Pr[B_{u,\pi}(i)] &= \Pr[E[i,\pi(u)] \cap \neg E[\pi^{-1}(i),u]] + \Pr[\neg E[i,\pi(u)] \cap E[\pi^{-1}(i),u]] \\
&\overset{(a)}{\leqslant} \Pr[E[i,\pi(u)]] + \Pr[E[\pi^{-1}(i),u]],
\end{aligned}$$

where (a) is a consequence of two applications of monotonicity of probabilities. Hence

$$\begin{aligned}
\sum_{i=1}^{n} \Pr[B_{u,\pi}(i)] &\overset{(a)}{\leqslant} \sum_{i=1}^{n} \Pr[E[i,\pi(u)]] + \sum_{\pi^{-1}(i)=1}^{n} \Pr[E[\pi^{-1}(i),u]] \\
&\overset{(b)}{\leqslant} 1 + 2\sum_{\pi^{-1}(i)=1}^{n} \Pr[E[\pi^{-1}(i),u]] \\
&\leqslant 1 + 2\sum_{\pi^{-1}(i)=1}^{u} \Pr[E[\pi^{-1}(i),u]] + 2\sum_{\pi^{-1}(i)=u}^{n} \Pr[E[\pi^{-1}(i),u]] \\
&\overset{(c)}{\leqslant} 1 + 4m + 2\sum_{\pi^{-1}(i)=u}^{n} \Pr[E[\pi^{-1}(i),u]],
\end{aligned}$$

where (a) follows from the previous inequality and (c) from Lemma 4. The justification for (b) is slightly more complicated than for the other two steps: it follows from the fact that

$$\Pr[E[i,\pi(u)]] \leqslant \Pr[E[i,u]], \qquad i \neq u,$$

MOST WIEDZY Downloaded from mostwiedzy.pl

which can be seen as follows: for $u \neq i < \pi(u)$, it follows from Lemma 1. If $i = \pi(u)$, then the left-hand side is 0, so the inequality holds. Finally, if $i > \pi(u) > u$, then the two probabilities are equal, due to the uniformity of the attachment process. For the case $i = u$, the inequality fails, and we instead upper bound that term by 1.

Thus, we can upper bound some more:

$$\sum_{\pi^{-1}(i)=u}^{n} \Pr[E[\pi^{-1}(i), u]] \overset{(a)}{\leqslant} 2m \sum_{\pi^{-1}(i)=u}^{n} \frac{1}{\pi^{-1}(i)} \overset{(b)}{\leqslant} 2m \left(1 + \log \frac{n}{u}\right),$$

where (a) follows from Lemma 2 and (b) from Lemma 5. This completes the proof.

3.2 Proof of Theorem 2

We will start by splitting into three cases: $u_n = O(1)$, $u_n \to \infty$, or neither is true (that is, $\limsup u_n = \infty$ and $\liminf u_n = O(1)$). The first we will handle using the second moment method. For the second, we use various upper bounds. The third can be handled by splitting the sequence into two subsequences, each of which can be handled via the previous two arguments.

3.2.1 $u_n = O(1)$

In the case where $u_n = O(1)$, applying Theorem 1 shows that $\mathbb{E}[D_{\pi_n}(u_n)] = \Theta(\log n)$, so that it grows to infinity. Thus, we will use the second moment method to show that $\Pr[D_{\pi_n}(u_n) = 0] \to 0$. We start with

$$\Pr[D_{\pi_n}(u_n) = 0] \leqslant \frac{\mathbf{Var}\ [D_{\pi_n}(u_n)]}{(\mathbb{E}[D_{\pi_n}(u_n)])^2}.$$

Recall that we can write $D_{\pi_n}(u_n)$ as a sum of indicators, so that the numerator becomes

$$\begin{aligned}
\mathbf{Var}\ [D_{\pi_n}(u_n)] &= \mathbf{Var}\ \left[\sum_{j=1}^{n} B_{u_n,\pi_n}(j)\right] \\
&= \sum_{j=1}^{n} \mathbf{Var}\ [B_{u_n,\pi_n}(j)] + 2 \sum_{1 \leqslant i < j \leqslant n} \mathrm{Cov}[B_{u_n,\pi_n}(i), B_{u_n,\pi_n}(j)] \\
&\overset{(a)}{\leqslant} \sum_{j=1}^{n} \mathbb{E}[B_{u_n,\pi_n}(j)] + 2 \sum_{1 \leqslant i < j \leqslant n} \mathrm{Cov}[B_{u_n,\pi_n}(i), B_{u_n,\pi_n}(j)] \\
&\overset{(b)}{\leqslant} \mathbb{E}[D_{\pi_n}(u_n)] + 2 \sum_{1 \leqslant i < j \leqslant n} \mathrm{Cov}[B_{u_n,\pi_n}(i), B_{u_n,\pi_n}(j)],
\end{aligned}$$

where (a) is because the value of an indicator function is either 0 or 1, and (b) is by linearity of expectation. We show that the sum of covariance terms is small enough

MOST WIEDZY Downloaded from mostwiedzy.pl

leading to $\mathbf{Var}\,[D_{\pi_n}(u_n)] = O(\mathbb{E}[D_{\pi_n}(u_n)])$. The following inequality will be useful for arbitrary h and k:

$$\begin{aligned}
\mathrm{Cov}[B_{u_n,\pi_n}(h), B_{u_n,\pi_n}(k)] &= \Pr[B_{u_n,\pi_n}(h) \cap B_{u_n,\pi_n}(k)] - \Pr[B_{u_n,\pi_n}(h)]\Pr[B_{u_n,\pi_n}(k)] && (1)\\
&\leqslant \Pr[B_{u_n,\pi_n}(h) \cap B_{u_n,\pi_n}(k)] && (2)\\
&\leqslant \Pr[B_{u_n,\pi_n}(h)]. && (3)
\end{aligned}$$

First, we consider a few subsets of terms for which elementary probability inequalities suffice. After dealing with these subsets, we will be left with a subset such that the ordering of vertices involved in the relevant edge events is sufficiently constrained for us to bound the term contributions using edge independence and dependence quantification properties.

Consider the contribution of pairs $i < j$ such that at least one of $\pi_n^{-1}(i)$, $\pi_n^{-1}(j)$ is at most u_n. It is bounded above by

$$\begin{aligned}
&\sum_{\pi_n^{-1}(i)=1}^{u_n} \sum_{j=1}^{n} \max\{\mathrm{Cov}[B_{u_n,\pi_n}(i), B_{u_n,\pi_n}(j)], 0\}\\
&+ \sum_{\pi_n^{-1}(j)=1}^{u_n} \sum_{i=1}^{n} \max\{\mathrm{Cov}[B_{u_n,\pi_n}(i), B_{u_n,\pi_n}(j)], 0\}\\
&\leqslant \sum_{\pi^{-1}(i)=1}^{u_n} \sum_{j=1}^{n} \Pr[B_{u_n,\pi_n}(j)] + \sum_{\pi^{-1}(j)=1}^{u_n} \sum_{i=1}^{n} \Pr[B_{u_n,\pi_n}(i)]\\
&= 2\sum_{i=1}^{u_n} \mathbb{E}[D_{\pi_n}(u_n)] = 2u_n\mathbb{E}[D_{\pi_n}(u_n)] = O(\mathbb{E}[D_{\pi_n}(u_n)]).
\end{aligned}$$

Here, the first upper bound is by considering all possible values for $\pi_n^{-1}(i) \leqslant u_n$ and $\pi_n^{-1}(j) \leqslant u_n$, respectively. For each such value, the other element of the pair may take values in a (possibly strict) subset of $[n]$. The upper bound results from summing over all values in $[n]$ of the second parameter and adding only non-negative terms. The second upper bound is an application of (3). The rest follows from linearity of expectation. In the same way, we get an analogous bound for the case where either $\pi_n^{-1}(i)$ or $\pi_n^{-1}(j)$ is equal to $\pi_n(u_n)$ and for the case where $i = u_n$.

Now, consider the contribution of pairs $i < j$ such that i and j are neighbors in the disjoint cycle representation of π_n and such that both $\pi_n^{-1}(i)$ and $\pi_n^{-1}(j)$ are at least u_n+1. The contribution of such terms is at most

$$\sum_{i=1}^{n} \max\{\mathrm{Cov}[B_{u_n,\pi_n}(i), B_{u_n,\pi_n}(\pi_n(i))], 0\} \leqslant \sum_{i=1}^{n} \Pr[B_{u_n,\pi_n}(i)] \leqslant \mathbb{E}[D_{\pi_n}(u_n)].$$

We next consider the case in which at least one of $i, j \geqslant \pi_n(u_n)$ and i and j are not neighbors in the disjoint cycle representation of π_n. When $i = \pi_n(u_n)$, $B_{u_n,\pi_n}(i) = 0$ with probability 1, so that such covariance terms are 0. The same holds for $j = \pi_n(u_n)$. Thus,

MOST WIEDZY Downloaded from mostwiedzy.pl

we are left with the case in which at least one of $i, j > \pi_n(u_n)$. In this case, the covariance is 0, as a simple consequence of edge independence properties. In particular, we claim that $B_{u_n,\pi_n}(i)$ and $B_{u_n,\pi_n}(j)$ are independent in this situation. We have one of

$$u_n < \pi_n(u_n) < i < j \tag{4}$$

or

$$u_n < i < \pi_n(u_n) < j. \tag{5}$$

For simplicity, we only explain (4). The other case is similar.

Both $\pi_n^{-1}(i)$ and $\pi_n^{-1}(j)$ are greater than u_n (this from the fact that u_n is the least non-fixed vertex, and both i and j are greater than u_n). Thus, the variable $B_{u_n,\pi_n}(i)$ involves only the choices at times i and $\pi_n^{-1}(i)$, and $B_{u_n,\pi_n}(j)$ involves only the choices at times j and $\pi_n^{-1}(j)$. By the assumption that i and j aren't cycle neighbors, these two sets of choices are disjoint, so that they are independent, so that the covariance is 0, as claimed.

With the easy cases out of the way, we can focus on the contribution of pairs such that $\pi_n^{-1}(i), \pi_n^{-1}(j) > u_n$, i and j are not neighbors in the disjoint cycle representation of π_n, $i, j < \pi_n(u_n)$, and $\pi_n^{-1}(i) \neq \pi_n(u_n) \neq \pi_n^{-1}(j)$. We proceed by expressing an arbitrary covariance term in terms of edge events.

$$\mathrm{Cov}[B_{u_n,\pi_n}(i), B_{u_n,\pi_n}(j)] = \Delta_{1,1}(i,j) + \Delta_{1,2}(i,j) + \Delta_{2,1}(i,j) + \Delta_{2,2}(i,j),$$

where

$$\Delta_{k,l}(i,j) = \Pr[F_k(i) \cap F_l(j)] - \Pr[F_k(i)]\Pr[F_l(j)],$$

where we define

$$\begin{aligned} F_1(x) &= E[x, \pi_n(u_n)] \cap \neg E[\pi_n^{-1}(x), u_n] \\ F_2(x) &= \neg E[x, \pi_n(u_n)] \cap E[\pi_n^{-1}(x), u_n]. \end{aligned}$$

To write these more succinctly and to simplify a later definition, we define

$$\rho_1(k) = \begin{cases} 0 & k = 1 \\ -1 & k = 2, \end{cases}$$

$$\rho_2(k) = \begin{cases} 1 & k = 1 \\ 0 & k = 2 \end{cases}.$$

Then define $\rho_1'(k) = -(\rho_1(k) + 1)$ and $\rho_2'(k) = -\rho_2(k) + 1$. With this notation, we can write each F_k as

$$F_k(x) = E[\pi_n^{\rho_1(k)}(x), \pi_n^{\rho_2(k)}(u_n)] \cap \neg E[\pi_n^{\rho_1'(k)}(x), \pi_n^{\rho_2'(k)}(u_n)].$$

To proceed, we need some lemmas about probabilities of conjunctions and negations of edge events. We again omit the simple proofs.

MOST WIEDZY Downloaded from mostwiedzy.pl

Lemma 6 (Edge Event Probabilities)**.** *For all $i, q < j$, $i \neq q$,*

$$\begin{aligned}
&\Pr[E[i,j]] = 1 - \left(\frac{j-2}{j-1}\right)^m \\
&\Pr[E[i,j] \cap \neg E[q,j]] = \left(1 - \left(\frac{j-3}{j-2}\right)^m\right)\left(\frac{j-2}{j-1}\right)^m \\
&\Pr[E[i,j] \cap \neg E[q,j]] = \Pr[E[i,j]]\Pr[\neg E[q,j]](1 + O(j^{-1})) \\
&\Pr[\neg E[i,j] \cap \neg E[q,j]] = \Pr[\neg E[i,j]]\Pr[\neg E[q,j]](1 + O(j^{-2})).
\end{aligned}$$

Lemma 7 (Edge Negation Inequality)**.** *For $j \geqslant 2m+1$,*

$$\Pr[\neg E[i,j]] \geqslant \Pr[E[i,j]].$$

Lemma 8 (Edge Conjunction Inequalities)**.** *For all i, q, j, r such that $i < j$ and $q < r$, and either $i \neq q$ or $j \neq r$,*

$$\Pr[E[i,j] \cap E[q,r]] \leqslant \Pr[E[i,j]]\Pr[E[q,r]]$$

and

$$\Pr[E[i,j] \cap \neg E[q,r]] \geqslant \Pr[E[i,j]]\Pr[\neg E[q,r]].$$

Lemma 9. *For all i, q, j, r such that $i < j$, $q < r$, $r \geqslant 2m+1$, and either $i \neq q$ or $j \neq r$,*

$$\Pr[E[i,j] \cap \neg E[q,r]] \geqslant \Pr[E[i,j] \cap E[q,r]].$$

Now, the probability of each $F_k(x)$ can be lower bounded by the product of the involved edge probabilities, by Lemma 8, which gives an upper bound on each $\Delta_{k,l}(i,j)$. We call the resulting expression $\Delta'_{k,l}(i,j)$; more specifically, $\Delta'_{k,l}(i,j)$ is given by

$$\begin{aligned}
\Delta'_{k,l}(i,j) = &\Pr[F_k(i) \cap F_l(j)] \\
&- \Pr[E[\pi_n^{\rho_1(k)}(i), \pi_n^{\rho_2(k)}(u_n)]]\Pr[\neg E[\pi_n^{\rho'_1(k)}(i), \pi_n^{\rho'_2(k)}(u_n)]] \\
&\cdot \Pr[E[\pi_n^{\rho_1(l)}(j), \pi_n^{\rho_2(l)}(u_n)]]\Pr[\neg E[\pi_n^{\rho'_1(l)}(j), \pi_n^{\rho'_2(l)}(u_n)]]
\end{aligned}$$

Now we show that $\Delta'_{1,1}(i,j) \leqslant 0$.

$$\begin{aligned}
\Pr[F_1(i) \cap F_1(j)] &\overset{(a)}{=} \Pr[E[i, \pi_n(u_n)] \cap E[j, \pi_n(u_n)]] \cdot \Pr[\neg E[\pi_n^{-1}(i), u_n] \cap \neg E[\pi_n^{-1}(j), u_n]] \\
&\overset{(b)}{\leqslant} \Pr[E[i, \pi_n(u_n)]]\Pr[E[j, \pi_n(u_n)]]\Pr[\neg E[\pi_n^{-1}(i), u_n]]\Pr[\neg E[\pi_n^{-1}(j), u_n]].
\end{aligned}$$

Here, (a) is an application of edge independence properties: we have $i < j < \pi_n(u_n)$ and $u_n < \pi_n^{-1}(i), \pi_n^{-1}(j)$, and neither of $\pi_n^{-1}(i), \pi_n^{-1}(j)$ is equal to $\pi_n(u_n)$. The inequality, (b), is due to applying edge independence to split the negated events and applying Lemma 8 to split the other conjunction. It follows from this inequality that $\Delta'_{1,1}(i,j) \leqslant 0$.

In a similar manner, we handle $\Delta'_{1,2}(i,j)$:

$$\Delta'_{1,2}(i,j) = \Pr[E[i, \pi_n(u_n)] \cap \neg E[j, \pi_n(u_n)]]\Pr[\neg E[\pi_n^{-1}(i), u_n]]\Pr[E[\pi_n^{-1}(j), u_n]]$$

MOST WIEDZY Downloaded from mostwiedzy.pl

$$
\begin{aligned}
&- \Pr[E[i,\pi_n(u_n)]]\Pr[\neg E[j,\pi_n(u_n)]]\Pr[\neg E[\pi_n^{-1}(i),u_n]]\Pr[E[\pi_n^{-1}(j),u_n]]\\
&\leqslant \Pr[E[i,\pi_n(u_n)]\cap\neg E[j,\pi_n(u_n)]] - \Pr[E[i,\pi_n(u_n)]]\Pr[\neg E[j,\pi_n(u_n)]]\\
&= \Pr[E[i,\pi_n(u_n)]]\Pr[\neg E[j,\pi_n(u_n)]](1+O(\pi_n(u_n)^{-1})-1)\\
&\leqslant \frac{2m}{\pi_n(u_n)}O(\pi_n(u_n)^{-1}) = O\left(\frac{1}{(\pi_n(u_n))^2}\right).
\end{aligned}
$$

The first equality is by applying edge independence properties. The inequality is by factoring out the probabilities involving u_n and upper bounding them by 1. The final two steps are by applying Lemmas 6 and 2.

The other cases are handled analogously. Thus, each covariance term in the set considered is upper bounded by $\frac{C}{\pi_n(u_n)^2}$, where C is a constant. Since there are at most $C'\pi_n(u_n)^2$ such pairs, the total contribution is at most a constant. This completes the proof for $u_n = O(1)$.

3.2.2 $u_n \to \infty$

For $u_n \to \infty$, we will upper bound the probability of the event

$$A = \bigcap_{i=1}^{u_n-1}[B_{u_n,\pi_n}(i)=0],$$

which is a superset of $[D_{\pi_n}(u_n)=0]$. The condition that $u_n \to \infty$ gives us enough fixed vertices for a somewhat simple proof. Observe that the event is equivalent to u_n and $\pi_n(u_n)$ having precisely the same neighbors in $[u_n-1]$.

First, condition on the set N of distinct choices made by vertex $\pi_n(u_n)$ and let $N_1 = N\cap[u_n-1]$. Then

$$|N_1| \leqslant |N| \leqslant m.$$

Now, if A occurs, then all m of the choices $c_{u_n,j}$ at time u_n lie in N_1. This happens with probability at most

$$\left(\frac{|N_1|}{u_n-1}\right)^m \leqslant \left(\frac{m}{u_n-1}\right)^m = O(u_n^{-m}).$$

Since $u_n \to \infty$, this tends to 0, which completes the proof for this case.

3.2.3 u_n is neither bounded nor tending to ∞

Our goal is to show that, for any $\epsilon > 0$, there exists N such that, for all $n \geqslant N$,

$$\Pr[D_{\pi_n}(u_n)=0] < \epsilon.$$

In the case where u_n is neither bounded nor tending to ∞, we start by choosing an appropriate $C = C(\epsilon)$ and partitioning the sequence of permutations into two subsequences: those for which $u_n \leqslant C$ and those for which $u_n > C$.

For the former subsequence (call its sequence of probabilities a_n), we can apply the result for $u_n = O(1)$ to conclude that there is some N such that, if $n \geqslant N$, then $a_n < \epsilon$.

MOST WIEDZY Downloaded from mostwiedzy.pl

For the latter (call its sequence of probabilities b_n), we can apply the argument used in the proof for $u_n \to \infty$ to conclude that there is a large enough $C(\epsilon)$ such that $b_n < \epsilon$ for all n. This completes the proof of Theorem 2.

3.3 Proof of Theorem 3

Case $m = 1$.
To prove symmetry with high probability, we give a positive lower bound on the asymptotic number of cherries which holds with high probability. This we do by formulating the preferential and uniform attachment processes as urn models, in which one type of ball corresponds to a *simple* cherry (i.e., a node with exactly two children, both leaves). We then use Theorem 3.21 of Janson (2004) (see also (Athreya and Ney, 1972, Section V.9.3)) to determine the asymptotic fraction of such balls, which turns out to be positive.

We now review the aforementioned Theorem 3.21. In the general setting, we have an urn model with q types of balls, and for each non-negative integer time t, we have a (random) vector of ball type counts: $X_t = (X_{t,1}, \ldots, X_{t,q})$, where $\{X_t\}_{t=0}^{\infty}$ forms a Markov chain. Furthermore, each ball type i gets a fixed weight (also called *activity*) a_i as well as a random q-dimensional vector $\xi_i = (\xi_{i,j})_{j=1}^{q}$ which will be used to form the transition matrix. At time t, a ball is drawn with probability proportional to its weight and current count:

$$\Pr[\text{Type } j \text{ is drawn at time } t] = \frac{a_j X_{t-1,j}}{\sum_{k=1}^{q} a_k X_{t-1,k}}.$$

Next, the ball is returned to the urn, and a vector $w_t = (w_{t,r})_{r=1}^{q}$ is drawn according to the distribution of ξ_j (i.e., for any r and x, $\Pr[w_{t,r} = x] = \Pr[\xi_{j,r} = x]$), independent of everything else. Finally, for each $k \in [q]$, $w_{t,k}$ balls of type k are added to the urn. This results in X_t. The main object which is defined in order to study the asymptotic behavior of the number of balls of each type is the $q \times q$ matrix A, whose entries are given by

$$A_{i,j} = a_j \mathbb{E}[\xi_{j,i}].$$

We call A the *transition matrix* of the model. For us, the types of balls will correspond to different subtrees, and, roughly speaking, the weights will come from the degrees of the vertices in the subtrees (which is how we will encode the preferential attachment process).

Then Theorem 3.21 states that, assuming several hypotheses and given that there are always balls with nonzero weight left, we have the following asymptotic result:

$$\frac{X_t}{t} \xrightarrow{a.s.} \lambda_1 v_1,$$

where λ_1 and v_1 are the principal eigenpair of A. From this, we can read off the asymptotic linear growth of the number of any of the types of balls.

Conveniently, the paper gives Lemma 2.1, which states sufficient conditions for the hypotheses of Theorem 3.21 to hold: we need only check that

MOST WIEDZY Downloaded from mostwiedzy.pl

- A is *irreducible*; that is, we define a partial order $i \succ j$, where $i \succ j$ if and only if there exists some $n \geqslant 0$ for which $(A^n)_{j,i} > 0$ (that is, starting with ball of type i, it is possible for a ball of type j to appear in the future). Then we say that A is irreducible if, for all i and j, we have both $i \succ j$ and $j \succ i$.

- Almost surely, for all $j \neq i$,

$$\xi_{i,j} \geqslant 0, \qquad\qquad \xi_{i,i} \geqslant -1.$$

- For all i and j,

$$\mathbb{E}[\xi_{i,j}^2] < \infty.$$

- We have, for every i,

$$\sum_j \mathbb{E}[\xi_{i,j}] \geqslant 0$$

  and, for some i,

$$\sum_j \mathbb{E}[\xi_{i,j}] > 0.$$

We focus first on the preferential case, then explain how the proof changes for uniform attachment.

For preferential attachment, we have the following types of balls:

- SC (simple cherry): a node with precisely two children, both of whom are leaves. This is modeled by a single ball of type SC.

- PC (protocherry): a node with precisely one child, which is a leaf. This is modeled by a single ball of type PC.

- L (leaf): a leaf which is not part of a simple cherry or a protocherry. This is modeled by a single ball of type L.

- N (none of the above): a node with three or more children or with grandchildren. A single such node with degree k is modeled by k balls of type N.

Since these ball types will also correspond to coordinates in entries of a matrix, it will be convenient to associate SC with 1, PC with 2, L with 3, and N with 4. With each type of ball, there is an associated weight, which comes from considering the number of half edges in each situation (each edge in the subtree corresponding to a given ball counts as two half edges; finally, there is a single half edge which comes from the root vertex's connection to its parent):

$$a_{SC} = 5 \qquad\qquad a_{PC} = 3 \qquad\qquad a_L = 1 \qquad\qquad a_N = 1.$$

We can then formulate the process as follows. We start with a tree with 4 nodes, in order to avoid complications arising from the root. Each such tree corresponds to a well defined

MOST WIEDZY Downloaded from mostwiedzy.pl

initial vector of ball counts. At a given time step, we pick a ball randomly, where the probability of each type is in proportion to the current number of balls of that type and its weight: if we denote the type of the jth choice by T_j and the number of balls of type x at time j by $N_j(x)$,

$$\Pr[T_j = x] = \frac{a_x N_j(x)}{\sum_y a_y N_j(y)},$$

where the sum is over the types. After drawing a ball, we discard it and add new balls to the urn in the following manner, based on T_j.

- $$SC \implies \begin{cases} 4N + 3L & \text{with probability } 3/5 \\ 3N + 1L + 1PC & \text{with probability } 2/5 \end{cases}$$

- $$PC \implies \begin{cases} 1SC & \text{with probability } 2/3 \\ 2N + 1PC & \text{with probability } 1/3 \end{cases}$$

- $$L \implies 1PC$$

- $$N \implies 1L + 2N.$$

The transitions for SC can be seen as follows. Drawing a ball of type SC means, in terms of the preferential attachment process, that either the root node of a simple cherry or one of its two children is chosen by the new vertex. The root node has degree 3, and each child has degree 1, so that the total degree of the subgraph is 5. Thus, with probability 3/5, the root node is chosen, which destroys the simple cherry. It is replaced with 4 balls of type N (corresponding to a non-simple cherry whose root node has degree 4), along with 3 balls of type L (corresponding to the three leaves). With probability 2/5, one of the two leaves of the simple cherry is chosen. This destroys the simple cherry, replacing it with three balls of type N (because the root of the simple cherry still has degree 3), and the chosen child becomes a protocherry, which adds a single ball of type PC. The other child becomes a leaf, which adds a ball of type L.

The other transitions follow by similar reasoning. Note that, in each case, the degree of each node in the new subtree must somehow be encoded in the number and types of balls.

This yields, in the language of Janson (2004), the following transition matrix:

$$A = \begin{pmatrix} -5 & 2 & 0 & 0 \\ 2 & 2 & 1 & 0 \\ 11 & 0 & -1 & 1 \\ 18 & 2 & 0 & 1 \end{pmatrix}.$$

MOST WIEDZY Downloaded from mostwiedzy.pl

We demonstrate how A was obtained from the transitions and the weights by explicitly computing the first column, which corresponds to the balls of type SC. We need the vector $\mathbb{E}[\xi_1]$, which gives the expected number of balls of each type which are added to the urn after SC is drawn (and before it is discarded). Since SC is discarded with probability 1, we have $\mathbb{E}[\xi_{1,1}] = -1$. The expected number of PC balls added is $\mathbb{E}[\xi_{1,2}] = 0\cdot\frac{3}{5}+1\cdot\frac{2}{5} = 2/5$. For L, it is $\mathbb{E}[\xi_{1,3}] = 3\cdot\frac{3}{5}+1\cdot\frac{2}{5} = 11/5$. Finally, for N, it is $\mathbb{E}[\xi_{1,4}] = 4\cdot\frac{3}{5}+3\cdot\frac{2}{5} = 18/5$. Multiplying by the weight a_{SC} for simple cherries, we get the claimed first column of A. The other columns are computed analogously. Now, we can apply Lemma 2.1 of Janson (2004) to show that the hypotheses of Theorem 3.21 there are satisfied: the process is easily seen to be irreducible, since, with positive probability, SC yields N, which yields L, which yields PC, which yields SC. The other hypotheses of Lemma 2.1 are similarly trivial to check. From this, we conclude that all of the hypotheses of Theorem 3.21, in addition to impossibility of essential extinction, hold. Thus, it remains to determine the dominant eigenvalue, i.e., the largest real eigenvalue, and the corresponding eigenvector of A. We get, from standard calculations, that the largest real eigenvalue of A is $\lambda_1 = 2$ (the other eigenvalues are $-1, -3, -5$), and that, after normalization, the corresponding eigenvector is

$$v_1^T = \tfrac{1}{105}(2, 7, 24, 50) \approx (0.019048, 0.06667, 0.22857, 0.47619).$$

Multiplying by λ_1, we get from (Janson, 2004, Theorem 3.12) that, asymptotically, the number of simple cherries is with high probability approximately $\frac{4n}{105}$, which is certainly greater than 0. Thus, we have symmetry with high probability.

A similar proof (with the same types of balls, but with a slightly different interpretation N, to be explained below) yields the same result (with different constants) for the uniform attachment model. Because the attachment process pays attention to number of nodes in a subgraph, rather than to degrees of vertices, a ball of type N still corresponds to a root node with three or more children or with grandchildren, but now a single such node corresponds to a single ball of type N. Its descendants are represented by other balls.

The weights change, again to account for the number of vertices in each type of subtree. Thus, for example, the weight of SC is 3. In this case, the weights of the balls are $(3, 2, 1, 1)$.

Finally, the transitions change. Namely, instead of adding multiple copies of N in any given transition, we only add one. Thus, the transition matrix is

$$A = \begin{pmatrix} -3 & 1 & 0 & 0 \\ 2 & -1 & 1 & 0 \\ 5 & 0 & -1 & 1 \\ 3 & 1 & 0 & 0 \end{pmatrix}$$

with dominant eigenvalue 1 and corresponding normalized eigenvector $v_1^T = \frac{1}{24}(1, 4, 6, 7)$.

Case $m = 2$

We will show that, with positive probability, in both models, there is at least one *diamond* (i.e., a pair of nodes that choose the same parents and that are not chosen by any subsequent nodes) as shown in Figure 1(b). The details are technically more intricate than

MOST WIEDZY Downloaded from mostwiedzy.pl

in the $m = 1$ case, and the argument there does not work in this case, because node n must choose, from a set of size $\Theta(n^2)$ (pairs of vertices), one of $O(n)$ elements (previously chosen pairs). We thus rely on a birthday paradox-style argument to show that there is a positive probability of two vertices making the same choices, then condition on the lexicographically smallest such pair to complete the proof.

Let $A(u, v)$ be the event that vertices u and v choose the same pair of parents, and let $B(u, v)$ be the event that u and v are both unchosen. Now, define $N(k)$ to be the number of pairs u, v of vertices such that $k < u < v$ and $A(u, v)$ and $B(u, v)$ simultaneously hold. Define $N_A(k)$ and $N_B(k)$ analogously for pairs for which events A and B hold, respectively. Finally, denote by $S_{>x}$ the set $\{k \in [n] | k > x\}$. We then aim to prove that $\Pr[N(0) > 0] > C > 0$ for some constant C and n large enough. We do this by proving

$$\Pr[N(n/2) > 0] > C > 0,$$

in light of the fact that

$$\Pr[N(n/2) > 0] \leqslant \Pr[N(0) > 0].$$

For any x, we have

$$\Pr[N(x) > 0] = \Pr[N(x) > 0 \cap N_A(x) > 0] = \Pr[N(x) > 0 | N_A(x) > 0] \cdot \Pr[N_A(x) > 0],$$

where the first equality is from the fact that $[N(x) > 0] \subseteq [N(x)_A > 0]$. The goal now is bound the remaining probabilities below by positive constants. We do this in the next two lemmas, which hold for both uniform and preferential attachment graphs. We will prove them in the uniform case, then explain the modifications needed to extend them to the preferential case.

Lemma 10 (Probability of two vertices picking the same pair)**.** *There exists a positive constant C such that*

$$\Pr[N_A(n/2) > 0] > C$$

for all n sufficiently large.

Proof. To show this, we will instead compute $\Pr[N_A(n/2) = 0]$ and bound it above by a constant less than 1. The condition $N_A(n/2) = 0$ means that all vertices $> n/2$ choose distinct pairs. This is given by

$$\begin{aligned}
\Pr[N_A(n/2) = 0] &\leqslant \prod_{k=1}^{n/2} \left(1 - \frac{k-1}{(\frac{n}{2} + k - 1)^2}\right) \\
&\leqslant \prod_{k=1}^{n/2} \left(1 - \frac{k-1}{n^2}\right) \leqslant \prod_{k=n/4+1}^{n/2} \left(1 - \frac{k-1}{n^2}\right) \\
&\leqslant \prod_{k=n/4+1}^{n/2} \left(1 - \frac{n}{4n^2}\right) = \left(1 - \frac{(1/4)}{n}\right)^{\frac{n}{4}} \xrightarrow{n\to\infty} e^{-\frac{1}{16}} < 1.
\end{aligned}$$

In the preferential attachment case, the proof is similar, except that we apply the fact that the attractiveness of any vertex $v < t$ at time t is at least $\frac{m}{2m(t-2)} = \frac{1}{2(t-2)}$. $\square$

MOST WIEDZY Downloaded from mostwiedzy.pl

Lemma 11 (Conditional probability of two vertices with the same neighborhood)**.** *There exists a positive constant C such that*

$$\Pr[N(n/2) > 0 | N_A(n/2) > 0] > C$$

for all n sufficiently large.

Proof. We condition on the lexicographically smallest pair X from $S_{>n/2}$ such that $A(X)$ holds. Let $D(u,v)$ be the event that the pair (u,v) is the smallest pair from $S_{>n/2}$ for which A holds. Then

$$\begin{aligned}
&\Pr[N(n/2) > 0 | N_A(n/2) > 0] \\
&\qquad = \sum_{u<v\in S_{>n/2}} \Pr[N(n/2) > 0 | D(u,v)] \Pr[D(u,v) | N_A(n/2) > 0] \\
&\qquad \geqslant \sum_{u<v\in S_{>n/2}} \Pr[B(u,v) | D(u,v)] \Pr[D(u,v) | N_A(n/2) > 0] \\
&\qquad \geqslant C \sum_{u<v\in S_{>n/2}} \Pr[D(u,v) | N_A(n/2) > 0] = C.
\end{aligned}$$

Here, the equalities are simply due to the law of total probability, and the first inequality is because the event $B(u,v)$ is a subset of the event $N(n/2) > 0$. The second inequality is by direct computation. Note first that $D(u,v)$ means that all lexicographically smaller pairs choose distinct pairs, and u and v choose the same pair (so that v cannot choose u). So

$$\begin{aligned}
\Pr[B(u,v)|D(u,v)] &= \prod_{j=u+1}^{v} \Pr[j \text{ avoids } u | D(u,v)] \prod_{j=v+1}^{n} \Pr[j \text{ avoids } v, u | D(u,v)] \\
&\geqslant \prod_{j=u+1}^{v-1} \left(\frac{(j-1)^2 - 2j - u}{(j-1)^2} \right) \cdot 1 \cdot \prod_{j=v+1}^{n} \left(\frac{(j-1)^2 - 4j - u}{(j-1)^2} \right) \\
&\geqslant \prod_{j=u+1}^{v-1} \left(1 - \frac{c}{n}\right) \prod_{j=v+1}^{n} \left(1 - \frac{c}{n}\right) \\
&\geqslant \left(1 - \frac{c}{n}\right)^n,
\end{aligned}$$

where $c > 0$ is some constant. Here, the first inequality results from bounding the numerators below by giving upper bounds for the number of pairs that vertex j must avoid in order to avoid u and v and for the number of pairs that j must avoid in order to pick a pair that is distinct from the choices of all vertices x such that (x,j) is lexicographically smaller than (u,v). More specifically, conditioning on $D(u,v)$ reduces the number of pairs that j may pick, but this number can be upper bounded by $(j-1)^2$, which gives the denominators. Now, in order to avoid u, the j in the first product must avoid at most $2j$ pairs (those candidate pairs with u in the first and in the second coordinate,

MOST WIEDZY Downloaded from mostwiedzy.pl

respectively). We must also subtract u because of the condition that j must avoid pairs which were picked by vertices x such that (x, j) is lexicographically smaller than (u, v). That is, j must avoid at most u other choices of pairs. Similarly, for the second product, j must avoid pairs containing either of u or v (at most $2j$ possibilities for each of these), and j avoids at most u pairs which were chosen by vertices x for which (x, j) is lexicographically smaller than (u, v). The 1 between the products is from the fact that v avoids u with probability 1, due to the conditioning by $D(u, v)$. The second inequality holds for all n sufficiently large, since $j > \frac{n}{2}$. The last inequality is because all factors are bounded above by 1. Finally, by taking n sufficiently large, the last value can be made arbitrarily close to e^{-c}. The proof in the preferential case is again similar, relying on the previously stated lower bound on vertex attractiveness. □

3.4 Proof of Theorem 4

We will prove the claim when the expected window size is bounded by a constant by examining the probability that vertices $n-1$ and n both choose m times vertex $n-2$, thereby forming a cherry. Let η_i be the event that vertex i chooses m times vertex $n-2$, so that the task is to find a lower bound for $\Pr[\eta_{n-1} \cap \eta_n]$.

For any i, by the bound $\mathbb{E}[W_i] \leqslant w$, we have $\Pr[W_i \leqslant 2w] \geqslant 1/2$. Now,

$$\begin{aligned}
\Pr[\eta_{n-1} \cap \eta_n] &\overset{(a)}{\geqslant} \Pr[\eta_{n-1} \cap \eta_n | W_{n-1} \leqslant 2w, W_n \leqslant 2w] \cdot \Pr[W_{n-1} \leqslant 2w] \Pr[W_n \leqslant 2w] \\
&\geqslant \tfrac{1}{4} \Pr[\eta_{n-1} \cap \eta_n | W_{n-1} \leqslant 2w, W_n \leqslant 2w] \\
&= \tfrac{1}{4} \Pr[\eta_n | \eta_{n-1}, W_{n-1} \leqslant 2w, W_n \leqslant 2w] \cdot \Pr[\eta_{n-1} | W_{n-1} \leqslant 2w, W_n \leqslant 2w]
\end{aligned}$$

Here, (a) is by conditioning on the event that the window sizes of vertices $n-1$ and n are bounded by $2w$, then applying the independence of the window sizes.

Now, in both the preferential and uniform attachment cases, the total attractiveness within the window $\{n-2w, \ldots, n\}$ is bounded above by a constant. Furthermore, since vertex $n-2$ makes m choices, its attractiveness to vertices $n-1$ and n is at least a positive constant. Thus, the two remaining probabilities are both bounded below by a positive constant, which completes the proof.

If the window sizes are uniformly bounded above by w with probability 1, then symmetry with high probability results, roughly because the set $[n]$ may be partitioned into subintervals, each having a positive probability of contributing a symmetry; since window sizes are bounded, once a symmetry arises, if a particular constant amount of time passes, the symmetry will never be destroyed. First, we choose an appropriate subinterval size $\zeta = 2w + 2$ and write $n = q\zeta + r$, for some integers q, r, with $0 \leqslant r < \zeta$. We then divide the interval $[r, \ldots, n]$ into q disjoint subintervals $I_1, \ldots, I_q$, each of size ζ. Let I_k be one of the intervals. Consider the permutation $\pi_k = \pi$ that swaps the middlemost two vertices (call them x_k and $x_k + 1$) of I_k and leaves all other vertices in $[n]$ fixed. The event that $\pi_k \in \mathrm{Aut}(G_n)$ contains the event η_j that x_j and x_{j+1} both choose x_{j-1} m times and are

MOST WIEDZY Downloaded from mostwiedzy.pl

unchosen. We will show that

$$\Pr\left[\neg\bigcup_{j=1}^{q}\eta_j\right] \xrightarrow{n\to\infty} 0, \tag{6}$$

from which our claim follows. We can write the probability in (6) as

$$\Pr[\neg\bigcup_{j=1}^{q}\eta_j] = \prod_{j=1}^{q}\left(1-\Pr\left[\eta_j|\bigcap_{k=1}^{j-1}\neg\eta_k\right]\right).$$

Now, the goal is to show that each factor in the product is bounded above by a constant $C<1$ (independent of j). In both the preferential and the uniform model, at any time t, the conditional probability of choosing any given vertex within the window of vertex t is bounded above and below by positive constants (dependent only on m and w). In particular, at time t, the attractiveness of any vertex v within the window of vertex t is at least m since v makes m choices, and the total attractiveness within the window is upper bounded by $w(w+1)m$ (since each vertex within the window makes m choices, and at most w vertices may choose a given vertex within the window at most m times, and there are w vertices in the window). Thus, the probability, for any choice of t, that t chooses any v within its window is at least $\frac{m}{w(w+1)m}=\frac{1}{w(w+1)}$. That is, every vertex within the window of t has a non-negligible probability of being chosen by t. Note that this bound holds regardless of previous vertex choices.

It follows from this that there is a constant $C<1$ such that

$$1-\Pr\left[\eta_j|\bigcap_{k=1}^{j-1}\neg\eta_k\right] \leqslant C<1,$$

so that

$$\Pr\left[\neg\bigcup_{j=1}^{q}\eta_j\right] \leqslant C^q \xrightarrow{n\to\infty} 0,$$

since $q\to\infty$ with n. This completes the proof.

References

R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys*, 74:47–97, 2002.

K. B. Athreya and P. E. Ney. *Branching Processes.* Springer Berlin Heidelberg, 1972.

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

B. Bollobás and O. Riordan. The diameter of a scale-free random graph. *Combinatorica*, 24(1):5–34, 2004. doi:10.1007/s00493-004-0002-2.

MOST WIEDZY Downloaded from mostwiedzy.pl

Y. Choi and W. Szpankowski. Compression of graphical structures: Fundamental limits, algorithms, and experiments. *Information Theory, IEEE Transactions on*, 58(2):620–638, 2012. doi:10.1109/TIT.2011.2173710.

R. Durrett. *Random Graph Dynamics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, New York, NY, USA, 2006.

P. Erdős and A. Rényi. Asymmetric graphs. *Acta Math. Acad. Sci. Hungar.*, 14:295–315, 1963.

S. Janson. Functional limit theorems for multitype branching processes and generalized Pólya urns. *Stochastic Processes and their Applications*, 110(2):177 – 245, 2004. doi:10.1016/j.spa.2003.12.002.

J. H. Kim, B. Sudakov, and V. H. Vu. On the asymmetry of random regular graphs and random graphs. *Random Structures & Algorithms*, 21(3–4):216–224, 2002. doi:10.1002/rsa.10054.

B. D. MacArthur and J. W. Anderson. Symmetry and self-organization in complex systems. arXiv:cond-mat/0609274, Sept. 2006.

N. Rashevsky. Life, information theory, and topology. *The Bulletin of Mathematical Biophysics*, 17(3):229–235, 1955. doi:10.1007/BF02477860.

R. T. Smythe and H. M. Mahmoud. A survey of recursive trees. *Theory Prob. Math. Statist.*, 51:1–27, 1995.

E. Trucco. A note on the information content of graphs. *The Bulletin of Mathematical Biophysics*, 18(2):129–135, 1956. doi:10.1007/BF02477836.

MOST WIEDZY Downloaded from mostwiedzy.pl