Adv. Appl. Prob. **50,** 918–943 (2018)
*doi:*10.1017/apr.2018.41

DISPARITY OF CLUSTERING COEFFICIENTS IN THE HOLME–KIM NETWORK MODEL

R. I. OLIVEIRA,* *IMPA*
R. RIBEIRO ** *** AND
R. SANCHIS,** **** *Universidade Federal de Minas Gerais*

Abstract

The Holme–Kim random graph process is a variant of the Barabási–Álbert scale-free graph that was designed to exhibit clustering. In this paper we show that whether the model does indeed exhibit clustering depends on how we define the clustering coefficient. In fact, we find that the local clustering coefficient typically remains positive whereas global clustering tends to 0 at a slow rate. These and other results are proven via martingale techniques, such as Freedman's concentration inequality combined with a bootstrapping argument.

Keywords: Random graph; complex network; concentration bound; clustering coefficient; preferential attachment

2010 Mathematics Subject Classification: Primary 05C82
Secondary 60K40; 68R10

1. Introduction

The theory of random graphs traces its beginnings to the seminal work of Erdös and Rényi [8]. For about 40 years, most papers in this area were concerned with homogeneous models, where vertices are statistically indistinguishable. The original Erdös–Rényi model and random regular graphs both fall into this category; see [2] and [11].

More recently, inhomogeneous random graphs have also gained prominence, as large-scale networks of social, biological, and technological origins tend to be very far from homogeneous. Research on models of such 'complex networks' has attracted much interest in statistical physics, systems biology, sociology, and computer science, as well as in mathematics. We will not attempt to survey this huge area of research, but the interested reader is directed to [13] for a broad survey of the nonrigorous literature and to [6], [7], and [15] for compendia of rigorous results for many different models.

Two important complex network features are especially important for this paper. The first one is the *power-law*, or *scale-free*, degree distribution. It is believed that many real-life networks have constant average degree (relative to network size) but a highly skewed degree distribution, where the fraction of nodes of degree k behaves roughly as $k^{-\beta}$ for some $\beta > 1$. This contrasts with the Poisson degree distribution of sparse Erdös–Rényi graphs. A well-known scale-free

Received 27 October 2016; revision received 7 December 2017.
* Postal address: IMPA, Estrada Da. Castorina, 110 CEP 22460-320 Rio de Janeiro, RJ, Brazil.
Email address: rimfo@impa.br
** Postal address: Departamento de Matemática, Universidade Federal de Minas Gerais, Av. Antônio Carlos 6627 C.P. 702 CEP 30123-970 Belo Horizonte-MG, Brazil.
*** Email address: rodrigo-matematica@ufmg.br
**** Email address: rsanchis@mat.ufmg.br

model is the seminal Barabási–Álbert preferential attachment random graph model; see [1] and [4]. In this case $\beta = 3$ is the power-law exponent, but any $2 < \beta \leq 3$ may be achieved by minor modifications of the process; see [5].

Another important feature of certain social and metabolic networks is so-called *clustering*, meaning that 'friends of friends tend to be friends'. That is, if a node v is connected to both w and u then u and w are likely to be connected. This trait is not captured by the Barabási–Álbert model, but the small-world model of Strogatz and Watts [16], which is also extremely well known, does have this property (but has no power law).

1.1. The Holme–Kim model

In this paper we provide a rigorous analysis of a specific nonhomogeneous random graph model, whose motivation was to combine scale-freeness and clustering. This model was introduced in 2001 by Holme and Kim [10]. The Holme–Kim (HK) model describes a random sequence of graphs $\{G_t\}_{t\in\mathbb{N}}$ that we formally define in Section 3. Here we provide an informal description of this evolution. Fix two parameters $p \in [0, 1]$ and a positive integer $m > 1$, and start with a graph G_1. For $t > 1$, the evolution from G_{t-1} to G_t consists of the addition of a new vertex v_t and m new edges between v_t and the vertices already in G_{t-1}. These m edges are added sequentially in the following fashion.

- The first edge is always attached following the preferential attachment (PA) mechanism, that is, it connects to a previously existing node w with probability proportional to the degree of w in G_{t-1}.
- Each of the $m - 1$ remaining edges makes a random decision on how to attach.
    - With probability p, the edge is attached according to the triad formation (TF) mechanism. Let w' be the node of G_{t-1} to which the immediately preceding edge was attached. Then the current edge connects to a neighbor w of w' chosen with probability proportional to number of edges between w and w'.
    - With probability $1-p$ ($p \in [0, 1]$ fixed), the edge follows the same PA mechanism as the first edge (with fresh random choices). That is to say, the random choice of vertex w' to connect to is made independently of all other choices made up to this point.

The $p = 0$ case of this process, where only preferential attachment steps are performed, is essentially the Barabási–Álbert model [1]. The triad formation steps, on the other hand, are reminiscent of the copying model by Kumar *et al.* [12]. Holme and Kim argued on the basis of simulations and nonrigorous analysis that their model has the properties of scale-freeness and positive clustering. Still regarding the HK model, Ostroumova *et al.* [14] argued that it has vanishing clustering coefficient, although the argument lacks mathematical rigor.

Our rigorous results partly confirm their findings. The power-law degree distribution can be checked by known methods. On the other hand, we show that there are aspects of the clustering phenomenon (or lack thereof) that were not made evident in [10]. We will see that the question of whether the HK model has clustering admits two different answers depending on how we define clustering. In fact, we will argue that this same phenomenon should hold for a wide variety of network models.

1.2. Two distinct clustering coefficients

We start by providing some graph-theoretic definitions, valid for any graph G. We state these somewhat informally below and give precise definitions in Section 2.

- A *triangle* in G is a set of three vertices that are mutually connected by edges.
- A *cherry* (or path of length two) in G is a set of three vertices $\{u, v, w\}$ of G where u and w are both adjacent to v (we call v the middle vertex). Observe that a triangle contains three cherries.
- The *local clustering coefficient* $\mathcal{C}_G^{\text{loc}}$ measures the average, over vertices v of G, of the fraction of pairs of neighbors of v that are connected by edges. That is, if each pair of neighbors of v counts as a 'potential triangle involving v', then $\mathcal{C}_G^{\text{loc}}$ is the average over the vertices v of the ratio of 'actual triangles involving v' to 'potential triangles involving v'.
- The *global clustering coefficient* $\mathcal{C}_G^{\text{glo}}$ measures the fraction of all cherries $u \sim v \sim w$ in G that also satisfies $u \sim w$. If we count each cherry $u \sim v \sim w$ as a 'potential triangle' in G then $\mathcal{C}_G^{\text{glo}}$ measures the ratio of actual to potential triangles in the whole of G, multiplied by a factor of three so that $0 \le \mathcal{C}_G^{\text{glo}} \le 1$ (again, note that each actual triangle contains three cherries, if vertex labels are taken into account).

Bollobás and Riordan [3] observed that $\mathcal{C}_G^{\text{loc}}$ and $\mathcal{C}_G^{\text{glo}}$ are used interchangeably in the nonrigorous literature. They warned that:

> In more balanced graphs the definitions will give more similar values, but they will still differ by at least a constant factor much of the time [3].

In fact, more extreme differences are possible for nonregular graphs.

Example 1.1. Building a graph G consisting of an edge e and $n-2$ other vertices connected to the two endpoints of e, it is straightforward to see that $\mathcal{C}_G^{\text{loc}} = 1 - 2/n + 4/n(n-1)$. On the other hand, it is straightforward to see that $\mathcal{C}_G^{\text{glo}} = 3(n-2)/(n-2+(n-1)(n-2)) = 3/n$.

1.3. Our results

The main results in this paper show that such a disparity between local and global clustering does indeed occur in the specific case of the HK model, albeit in a less extreme form than Example 1.1 suggests. We enunciate that formally below. For a formal definition of the clustering coefficients involved at the statement of the two first theorems, we refer the reader to Section 2.4.

Theorem 1.1. (Positive local clustering for HK.) *Let* $\{G_t\}_{t\ge 0}$ *be the sequence of random graphs generated by the HK model with parameters* $m \ge 2$ *and* $p \in (0, 1)$. *Then the local clustering coefficients* $\mathcal{C}_{G_t}^{\text{loc}}$ *of the graphs* G_t *satisfy*

$$\lim_{t\to+\infty} \mathbb{P}\left(\mathcal{C}_{G_t}^{\text{loc}} \ge \frac{1-(1-p)^{m-1}}{(m+2)m(m-1)}\right) = 1.$$

Theorem 1.2. (Vanishing global clustering for HK.) *Let* $\{G_t\}_{t\ge 0}$ *be as in Theorem 1.1. Then the global clustering coefficients* $\mathcal{C}_{G_t}^{\text{glo}}$ *satisfy*

$$\lim_{t\to+\infty} \mathbb{P}\left(\frac{3(m-1)p/(m^2+m+1)}{\log t} \le \mathcal{C}_{G_t}^{\text{glo}} \le \frac{3m(m-1)/(m^2+m-1)}{\log t}\right) = 1.$$

Thus, for large t, one of the two clustering coefficients is typically far from 0, whereas the other one goes to 0 in probability, albeit at a slow rate. This shows that the remark by Bollobás and Riordan is very relevant in the analysis of at least one network model.

For completeness, we will also check that the HK model is scale free with power-law exponent $\beta = 3$; see Appendix A. The proof follows from standard methods in the literature.

Theorem 1.3. (The power law for HK.) *Let $\{G_t\}_{t\geq 0}$ be as in the previous theorem. Also let $N_t(d)$ be the number of vertices of degree d in G_t and set*

$$D_t(d) := \frac{\mathbb{E}N_t(d)}{t}.$$

Then

$$\lim_{t\to\infty} D_t(d) = \frac{2(m+1)m}{(d+2)(d+1)d}. \qquad \mathbb{P}(|N_t(d) - D_t(d)t| \geq 16dc\sqrt{t}) \leq (t+1)^{d-m}\mathrm{e}^{-c^2}.$$

1.4. Heuristics and a seemingly general phenomenon

The disparity between $\mathcal{C}_G^{\mathrm{loc}}$ and $\mathcal{C}_G^{\mathrm{glo}}$ should be a general phenomenon for large scale-free graph models with many (but not too many)triangles. This will transpire from the following heuristic analysis of the HK case with $p \in (0, 1)$.

To begin, it is not difficult to understand why Theorem 1.1 should hold. By Theorem 1.3, there is a positive fraction of nodes with degree m. Moreover, a positive fraction of these vertices are contained in at least one triangle because of the TF steps. We can make a more general observation.

Reason for positive local clustering. If a positive fraction of nodes have degree less than or equal to d (assumed constant), and a positive fraction of these nodes are contained in at least one triangle, then the local clustering coefficient $\mathcal{C}_{G_t}^{\mathrm{loc}}$ must be bounded away from 0.

We now argue that the vanishing of $\mathcal{C}_{G_t}^{\mathrm{glo}}$ should be a consequence of the power-law degree distribution. The global clustering coefficient $\mathcal{C}_{G_t}^{\mathrm{glo}}$ is essentially the ratio of the number of triangles to the number of cherries in G_t, the latter being denoted by C_t. Now we can easily show that the number of triangles in G_t grows linearly in t with high probability (w.h.p.), that is,

$$\mathcal{C}_{G_t}^{\mathrm{glo}} \approx \frac{\text{number of triangles in } G_t}{C_t} \approx \frac{t}{C_t}.$$

To estimate C_t, we note that each vertex v of degree d in G_t is the 'middle vertex' of exactly $d(d-1)/2 \approx d^2$ cherries. This means

$$\frac{C_t}{t} \approx \sum_{d=1}^{t} \frac{N_t(d)}{t} d^2 \approx \sum_{d=1}^{t} \frac{1}{d} \approx \log t,$$

noting that $N_t(d)/t \sim D_t(d) \approx d^{-3}$ by Theorem 1.3. Our reasoning is not rigorous because it requires bounds on $N_t(d)$ for very large d. However, we feel our argument is compelling enough to be true for many models. In fact, considering the case where $N_t(d)/t \approx d^{-\beta}$ for $0 < \beta \leq 3$, we are led to the following statement.

Heuristic reason for the large number of cherries. If the fraction of nodes of degree d in G_t is $\approx d^{-\beta}$ for some $0 < \beta \leq 3$, the number of cherries C_t is superlinear in t. More precisely, we expect $C_t/t \approx t^{3-\beta}$ for $0 < \beta < 3$ and $C_t/t \approx \log t$ for $\beta = 3$.

https://doi.org/10.1017/apr.2018.41 Published online by Cambridge University Press

The power-law range $0 < \beta \leq 3$ corresponds to most models of large networks in the literature. Likewise, we believe that the disparity between $\mathcal{C}^{\text{glo}}_{G_t}$ and $\mathcal{C}^{\text{loc}}_{G_t}$ should hold for all 'natural' random graph sequences with many triangles and power-law degree distribution with exponent $0 < \beta \leq 3$. We formulate the general message as follows.

Heuristic disparity between local and global clustering. Achieving positive local clustering is 'easy': just introduce a density of triangles in sparse areas of the graph. On the other hand, if the number of triangles in G_t grows linearly with time, and the fraction of nodes of degree d in G_t is $\approx d^{-\beta}$ for some $0 < \beta \leq 3$, then one can expect a vanishingly small global clustering coefficient.

1.5. Main technical ideas

At a high level, our proofs follow standard ideas from rigorous papers on complex networks. For instance, suppose we want to keep track of the number of nodes of degree d at time t for $d = m, m+1, \ldots, D$. Letting $N_t = (N_t(m), \ldots, N_t(D))$, the basic strategy adopted in several papers is to find a deterministic matrix $\mathcal{M}_{t-1}$ and a deterministic vector r_{t-1}, both measurable with respect to $G_0, \ldots, G_{t-1}$, such that

$$N_t = \mathcal{M}_{t-1}\, N_t + r_{t-1} + \varepsilon_t,$$

where $\mathbb{E}[\varepsilon_t \mid G_0, \ldots, G_t] \approx 0$. This can be seen as a 'noisy version' of the deterministic recursion $N_t = \mathcal{M}_{t-1} N_{t-1} + r_{t-1}$ with ε_t the 'noise' term. We can then study the recursion and use martingale techniques (especially the Azuma–Höffding inequality) to prove that N_t concentrates around the solution of the deterministic recursion. Our own proof of the power-law degree distribution follows this outline, and is only slightly different from the one in [6].

Once the degree sequence is analyzed, Theorem 1.1 is then a matter of observing that a density of vertices of degree m will be contained in at least one triangle, due to a TF step. On the other hand, the analysis of global clustering is more difficult, due to the need to estimate the number of cherries C_t. Justifying the heuristic calculation above would require strong control of the degree distribution up to very large values of d. We opt instead to write a 'noisy recursion' for C_t itself. However, the increments in this noisy recursion can be quite large, and the Azuma–Höffding inequality is not enough to control the process. We use instead Freedman's concentration inequality, which involves the quadratic variation, but even that is delicate because the variation might 'blow up' in certain unlikely events. In the end, we use a kind of 'bootstrap' argument, whereby a preliminary estimate of C_t is fed back into the martingale calculation to give sharper control of the predictable terms and the variation. The outcome is a weak law of large numbers for C_t, stated in the following theorem.

Theorem 1.4. (The weak law of large numbers for C_t.) *Let C_t be the number of cherries in G_t. Then*

$$\frac{C_t}{t \log t} \xrightarrow{\mathbb{P}} \binom{m+1}{2},$$

where '$\xrightarrow{\mathbb{P}}$' denotes convergence in probability.

Overall, our martingale analysis of C_t is our main technical contribution.

1.6. Organization

The remainder of the paper is organized as follows. In Section 2 we review some standard notation, introduce the relevant graph-theoretic concepts, and record martingale inequalities. In Section 3 we present a formal definition of the model. In Section 4 we prove technical estimates for the degree distribution which will be useful throughout the paper. Sections 5 and 6 are devoted to proving the bounds for the local and the global clustering coefficients, respectively. In Section 7 we present a comparative explanation for the distinct behavior of the clustering coefficients. The proof of the power-law distribution is left to Appendix A since it follows well-known martingale arguments.

2. Preliminaries

2.1. Set and multiset notation

Denote $\mathbb{N} = \{0, 1, 2, 3, \dots\}$. For $n \in \mathbb{N}\backslash\{0\}$, $[n] := \{1, 2, \dots, n\}$. Given a set S, let $|S|$ denote its cardinality, and $\binom{S}{k}$ denote the collection of all subsets of S of size $k \in \mathbb{N}$. The binomial coefficient $\binom{n}{k}$ is the number of elements in $\binom{[n]}{k}$.

A multiset M consists of a base set M_0 and, for each $s \in M_0$, a multiplicity $m_s \in \mathbb{N}\backslash\{0\}$. We say that a multiset M is contained in set S (and write $M \subset S$) if $M_0 \subset S$, and we say $s \in M$ if $s \in M_0$.

2.2. Basic graph theory

Recall that a graph $G = (V_G, E_G)$ consists of a set V_G of vertices and a multiset $E_G \subset \binom{V_G}{2}$ of edges. This implies that there might be multiple 'parallel' edges between any given $v, w \in V_G$. Given G and $v, w \in V_G$, we say that v and w are neighbors, and write $v \sim_G w$, if $\{v, w\} \in E_G$. We also write $\Gamma_G(v) = \{w \in V : w \sim_G v\}$ for the neighborhood of $v \in V_G$ and $e(\Gamma_G(v))$ for the number of edges between the neighbors of v. Since we allow multiple edges, for all vertices $v, w \in V_G$, we let $e_G(v, w)$ be the number of edges between v and w. In this case, we may define the *degree* of v in terms of $e_G(v, w)$, writing it as $d_G(w) = \sum_{w \in \Gamma_G(v)} e_G(v, w)$, that is, the total of edges where one of its ends is v.

2.3. Triangles and cherries

A triangle in a graph $G = (V_G, E_G)$ is a subset $\{u, v, w\} \in \binom{V_G}{3}$ with $u \sim_G v \sim_G w \sim_G u$. We denote the number of triangles contained in a graph G by Δ_G. For a fixed vertex $v \in V_G$, we denote the number of triangles sharing at least the common vertex v by $\Delta_G(v)$.

A cherry, or path of length two, is an element $(v, \{u, w\}) \in V_G \times \binom{V_G}{2}$ with $u \sim_G v \sim_G w$, or a pair $(v, u) \in V_G^2$ with $e_G(u, v) \geq 2$. We let C_G denote the number of cherries in G, counted according to edge multiplicities, that is,

$$C_G := \sum_{v \in V_G} \binom{d_G(v)}{2}.$$

2.4. Clustering coefficients

We assume that $G = (V_G, E_G)$ is a graph where $|V_G| > 0$ and all vertices have degree at least 2.

Definition 2.1. (*Local clustering coefficient in v.*) Given a vertex $v \in G$, the local clustering coefficient at v is

$$\mathcal{C}_G(v) = \frac{\Delta_G(v)}{\binom{d_G(v)}{2}}.$$

Note that $0 \leq \mathcal{C}_G(v) \leq 1$ always, since there can be at most one triangle formed by v and a pair of its neighbors. In probabilistic terms, $\mathcal{C}_G(v)$ measures the probability that a pair of

https://doi.org/10.1017/apr.2018.41 Published online by Cambridge University Press

random neighbors of v form an edge, that is, how likely it is that 'two friends of v are also each other's friends'.

We define the two coefficients for the graph G as follows.

Definition 2.2. (*Local and global clustering coefficients.*) The local clustering coefficient of G is defined as

$$\mathcal{C}_G^{\text{loc}} := \frac{\sum_{v \in V_G} \mathcal{C}_G(v)}{|V_G|},$$

whereas the global coefficient is

$$\mathcal{C}_G^{\text{glo}} := 3 \times \frac{\Delta_G}{C_G}.$$

Observation 2.1. For our purposes there is no significant difference between taking or not taking multiple edges into account in the above definitions. The key point is that our model allows at most m edges between two vertices. Thus, all bounds considering multiple edges may be carried over to the other case (and vice versa) at the cost of losing constant factors.

2.5. Sequences of graphs

We will often consider the sequence of graphs $\{G_t\}_t$ defined by the HK model, which is indexed by a discrete-time parameter $t \geq 0$. When considering this sequence, we replace the subscript G_t with t in our graph notation so that $V_t := V_{G_t}$, $E_t := E_{G_t}$, and so on. Given a sequence of numerical values $\{x_t\}_{t \geq 0}$ depending on t, we let $\Delta x_t := x_t - x_{t-1}$.

2.6. Asymptotics

We will use the Landau $o/O/\Theta$ notation at several points of our discussion. This always presupposes some asymptotics as a parameter n or t of interest goes to $+\infty$. The parameter will be clear from the context.

2.7. Martingale concentration inequalities

Recall that a supermartingale $(M_n, \mathcal{F}_n)_{n \in \mathbb{N}}$ consists of a filtration $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots$ and a family $\{M_n\}_{n \in \mathbb{N}}$ of integrable random variables where, for each $n \in \mathbb{N}$, M_n is $\mathcal{F}_n$-measurable and $\mathbb{E}[M_{n+1} \mid \mathcal{F}_n] \leq M_n$. If $(-M_n, \mathcal{F}_n)_{n \in \mathbb{N}}$ is also a supermartingale, we say that $(M_n, \mathcal{F}_n)_{n \in \mathbb{N}}$ is a martingale. We recall two well-known concentration inequalities for martingales.

Theorem 2.1. (Azuma–Höffding inequality; see [6].) *Let $(M_n, \mathcal{F}_n)_{n \geq 1}$ be a (super)martingale satisfying*

$$|M_{i+1} - M_i| \leq a_i.$$

Then, for all $\lambda > 0$ we have

$$\mathbb{P}(M_n - M_0 > \lambda) \leq \exp\biggl(-\frac{\lambda^2}{\sum_{i=1}^n a_i^2}\biggr).$$

Theorem 2.2. (Freedman's inequality; see [9].) *Let $(M_n, \mathcal{F}_n)_{n \geq 1}$ be a (super)martingale. Write*

$$V_n := \sum_{k=1}^{n-1} \mathbb{E}[(M_{k+1} - M_k)^2 \mid \mathcal{F}_k]$$

and suppose that $M_0 = 0$ and

$$|M_{k+1} - M_k| \leq R \quad \textit{for all } k.$$

Then, for all $\lambda > 0$ we have

$$\mathbb{P}(M_n \geq \lambda, V_n \leq \sigma^2 \text{ for some } n) \leq \exp\left(-\frac{\lambda^2}{2\sigma^2 + 2R\lambda/3}\right).$$

3. Formal definition of the process

In this section we provide a more formal definition of the HK process (compare with Section 1.1).

The model has two parameters: a positive integer number $m \geq 2$ and a real number $p \in [0, 1]$. It produces a graph sequence $\{G_t\}_{t\geq 1}$ which is obtained inductively according to the growth rule we describe below.

Initial state. The initial graph G_1, which will be taken as the graph with vertex set $V_1 = \{1\}$ and a single edge, is a self-loop.

Evolution. For $t > 1$, obtain G_{t+1} from G_t adding to it a vertex $t + 1$ and m edges between $t + 1$ and vertices $Y_{t+1}^{(i)} \in V_t$, $1 \leq i \leq m$. These vertices are chosen as follows. Let $\mathcal{F}_t$ be the σ-field generated by all random choices made in our construction up to time t. Assume we are given independent and identically distributed random variables $(\xi_{t+1}^{(i)})$ independent from $\mathcal{F}_t$. We define

$$\mathbb{P}(Y_{t+1}^{(1)} = u \mid \mathcal{F}_t) = \frac{d_t(u)}{2mt},$$

which means the first choice of vertex is always made using the preferential attachment mechanism. The next $m - 1$ choices $Y_{t+1}^{(i)}$, $2 \leq i \leq m$, are made as follows. Let $\mathcal{F}_t^{(i-1)}$ be the σ-field generated by $\mathcal{F}_t$ and all subsequent random choices made in choosing $Y_{t+1}^{(j)}$ for $1 \leq j \leq i - 1$. Then

$$\mathbb{P}(Y_{t+1}^{(i)} = u \mid \xi_{t+1}^{(i)} = x, \mathcal{F}_t^{(i-1)}) = \begin{cases} \dfrac{d_t(u)}{2mt} & \text{if } x = 0, \\ \dfrac{e_t(Y_{t+1}^{(i-1)}, u)}{d_t(Y_{t+1}^{(i-1)})} & \text{if } x = 1 \text{ and } u \in \Gamma_t(Y_{t+1}^{(i-1)}), \\ 0 & \text{otherwise.} \end{cases}$$

In other words, for each choice of the $m - 1$ end points, we flip an independent coin of parameter p and decide according to the outcome which mechanism we use to choose the end point. With probability p, we use the triad formation mechanism, that is, we choose the end point among the neighbors of the previously chosen vertex $Y_{t+1}^{(i-1)}$. With probability $1 - p$, we make a fresh choice from V_t using the preferential attachment mechanism. In this sense, if $\xi_t^{(i)} = 1$ we say that we have taken a TF-step. Otherwise, we say that a PA-step was performed.

4. Technical estimates for vertex degrees

In this section we collect several results on vertex degrees. In Subsection 4.1 we describe the probability of degree increments in a single step. In Subsection 4.2 we obtain upper bounds on all degrees. Some of these results are fairly technical and may be skipped in a first reading.

4.1. Degree increments

We begin with the following simple lemma.

Lemma 4.1. *For all $k \in \{1, \dots, m\}$, there exist positive constants $c_{m,p,k}$ and $\tilde{c}_{m,p,k}$ such that*

$$\left|\mathbb{P}(\Delta d_t(v) = k \mid G_t) - c_{m,p,k}\frac{d_t^k(v)}{t}\right| \le \tilde{c}_{m,p,k}\frac{d_t^{k+1}(v)}{t^{k+1}}.$$

In particular, for $k = 1$ we have $c_{m,p,1} = \frac{1}{2}$.

Before proving the lemma, we start by proving the following claim involving the random variables $Y_t^{(i)}$.

Claim 4.1. *For all $i \in \{0, 1, 2, \dots, m\}$,*

$$\mathbb{P}(Y_{t+1}^{(i)} = v \mid G_t) = \frac{d_t(v)}{2mt}. \tag{4.1}$$

Proof. The proof follows by induction on i. For $i = 1$ we have nothing to do. So, suppose the claim holds for all choices before $i - 1$. Then

$$\begin{aligned}\mathbb{P}(Y_{t+1}^{(i)} = v \mid G_t) &= \mathbb{P}(Y_{t+1}^{(i)} = v, Y_{t+1}^{(i-1)} = v \mid G_t) + \mathbb{P}(Y_{t+1}^{(i)} = v, Y_{t+1}^{(i-1)} \ne v \mid G_t)\\ &= \frac{(1-p)}{4m^2}\frac{d_t^2(v)}{t^2} + \mathbb{P}(Y_{t+1}^{(i)} = v, Y_{t+1}^{(i-1)} \ne v \mid G_t).\end{aligned} \tag{4.2}$$

For the first term on the right-hand side the only way we can choose v again is following a PA-step and then choosing v according to a preferential attachment rule. Thus,

$$\mathbb{P}(Y_{t+1}^{(i)} = v, Y_{t+1}^{(i-1)} = v \mid G_t) = \frac{(1-p)}{4m^2}\frac{d_t^2(v)}{t^2}.$$

For the second term, we divide it in two sets; whether the vertex chosen at the previous choice is a neighbor of v or not. Thus,

$$\begin{aligned}&\mathbb{P}(Y_{t+1}^{(i)} = v, Y_{t+1}^{(i-1)} \ne v \mid G_t)\\ &\quad= \sum_{u \notin \Gamma_{G_t}(v)} \mathbb{P}(Y_{t+1}^{(i)} = v, Y_{t+1}^{(i-1)} = u \mid G_t) + \sum_{u \in \Gamma_{G_t}(v)} \mathbb{P}(Y_{t+1}^{(i)} = v, Y_{t+1}^{(i-1)} = u \mid G_t)\\ &\quad= (1-p)\frac{d_t(v)}{2mt}\left(\sum_{u \notin \Gamma_{G_t}(v)} \frac{d_t(u)}{2mt}\right) + p\left(\sum_{u \in \Gamma_{G_t}(v)} \frac{e_{G_t}(u, v)}{d_t(u)}\frac{d_t(u)}{2mt}\right)\\ &\quad\quad+ \frac{(1-p)}{2m}\frac{d_t(v)}{t}\left(\sum_{u \in \Gamma_{G_t}(v)} \frac{d_t(u)}{2mt}\right)\\ &\quad= \frac{d_t(v)}{2mt} - \frac{(1-p)}{4m^2}\frac{d_t^2(v)}{t^2},\end{aligned}$$

where we used our inductive hypothesis and the fact that

$$\sum_{u \in \Gamma_{G_t}(v)} \frac{d_t(u)}{2mt} + \sum_{u \notin \Gamma_{G_t}(v)} \frac{d_t(u)}{2mt} = 1 - \frac{d_t(v)}{2mt}.$$

Returning to (4.2), we prove the claim. □

https://doi.org/10.1017/apr.2018.41 Published online by Cambridge University Press

Proof of Lemma 4.1. We will show the particular case of $k = 1$ since we have a particular interest in the value of $c_{m,p,1}$ and point out how to obtain the other cases. We begin by noting that the process of choosing is by definition a homogeneous Markovian process. This means that in order to evaluate the probability of a vertex increasing its degree by exactly one, the $k = 1$ case, we just need to know the probabilities of transition. In this way, we use the notation $\mathbb{P}_t$ to denote the measure conditioned on G_t and proceed to the computation of the probabilities of transition. We start with the most challenging one, that is,

$$\begin{aligned}\mathbb{P}_t(&Y_{t+1}^{(i+1)} = v \mid Y_{t+1}^{(i)} \neq v) \\ &= \sum_{u \in \Gamma_{G_t}(v)} \mathbb{P}_t(Y_{t+1}^{(i+1)} = v \mid Y_{t+1}^{(i)} = u)\mathbb{P}_t(Y_{t+1}^{(i)} = u \mid Y_{t+1}^{(i)} \neq v) \\ &\quad + \sum_{u \notin \Gamma_{G_t}(v)} \mathbb{P}_t(Y_{t+1}^{(i+1)} = v \mid Y_{t+1}^{(i)} = u)\mathbb{P}_t(Y_{t+1}^{(i)} = u \mid Y_{t+1}^{(i)} \neq v). \end{aligned} \tag{4.3}$$

When $u \in \Gamma_{G_t}(v)$, we can choose v by taking either a PA-step or a TP-step. This implies that

$$\mathbb{P}_t(Y_{t+1}^{(i+1)} = v \mid Y_{t+1}^{(i)} = u) = (1-p)\frac{d_t(v)}{2mt} + p\frac{e_t(u,v)}{d_t(u)}, \tag{4.4}$$

but when $u \notin \Gamma_{G_t}(v)$, the only way we can choose v is following a PA-step, which implies that

$$\mathbb{P}_t(Y_{t+1}^{(i+1)} = v \mid Y_{t+1}^{(i)} = u) = (1-p)\frac{d_t(v)}{2mt}. \tag{4.5}$$

We also note that the following equation holds since $u \neq v$ and Claim 4.1:

$$\mathbb{P}_t(Y_{t+1}^{(i)} = u \mid Y_{t+1}^{(i)} \neq v) = \frac{\mathbb{P}_t(Y_{t+1}^{(i)} = u)}{\mathbb{P}_t(Y_{t+1}^{(i)} \neq v)} = \frac{d_t(u)}{2mt\mathbb{P}_t(Y_{t+1}^{(i)} \neq v)}. \tag{4.6}$$

Claim 4.1 also implies that

$$\frac{1}{\mathbb{P}_t(Y_{t+1}^{(i)} \neq v)} = \frac{1}{1 - d_t(v)/2mt} = 1 + \sum_{n=1}^{\infty}\left(\frac{d_t(v)}{2mt}\right)^n. \tag{4.7}$$

Combining (4.4)–(4.6), we obtain

$$\begin{aligned}\mathbb{P}_t(Y_{t+1}^{(i+1)} = v \mid Y_{t+1}^{(i)} \neq v) &= (1-p)\frac{d_t(v)}{2mt} + p\frac{d_t(v)}{2mt}\left(1 + \sum_{n=1}^{\infty}\left(\frac{d_t(v)}{2mt}\right)^n\right) \\ &= \frac{d_t(v)}{2mt} + O\left(\frac{d_t^2(v)}{t^2}\right). \end{aligned} \tag{4.8}$$

If we chose v at the previous choice, the only way we select it again is following a PA-step, meaning that

$$\mathbb{P}_t(Y_{t+1}^{(i+1)} = v \mid Y_{t+1}^{(i)} = v) = (1-p)\frac{d_t(v)}{2mt}. \tag{4.9}$$

From these two probabilities of transition we are able to obtain the remaining ones.

To compute the probability of $\{\Delta d_t(v) = 1\}$ given G_t, we may split it in m possible ways to increase $d_t(v)$ by exactly one. For each of these we have an index $i \in \{1, \ldots, m\}$ that

https://doi.org/10.1017/apr.2018.41 Published online by Cambridge University Press

indicates v has been chosen at the ith step and avoided at the other $m-1$ choices. This means that each of these m ways has a probability similar to

$$\left(1-\frac{d_t(v)}{2mt}-O\left(\frac{d_t^2(v)}{t^2}\right)\right)^{i-1}\left(\frac{d_t(v)}{2mt}+O\left(\frac{d_t^2(v)}{t^2}\right)\right)\left(1-\frac{d_t(v)}{2mt}-O\left(\frac{d_t^2(v)}{t^2}\right)\right)^{m-i}.$$

Expanding the products and summing on $i \in \{1, 2, \dots, m\}$, we find a positive constant $\tilde{c}_{m,p,1}$ which does not depend on v or t and satisfies

$$\left|\mathbb{P}(\Delta d_t(v) = 1 \mid G_t) - \frac{d_t(v)}{2t}\right| \le \tilde{c}_{m,p,1}\frac{d_t^2(v)}{t^2}.$$

For the sake of simplicity, we write the above statement as

$$\mathbb{P}(\Delta d_t(v) = 1 \mid G_t) = \frac{d_t(v)}{2t} + O\left(\frac{d_t^2(v)}{t^2}\right).$$

The $k > 1$ cases are obtained in the same way, considering the $\binom{m}{k}$ ways of increasing $d_t(v)$ by k. □

4.2. Upper bounds on vertex degrees

To control the number of cherries C_t, we will need upper bounds on vertex degrees. The bound is obtained applying the Azuma–Hoffding inequality and Theorem 2.1 to the degree of each vertex, which is a martingale after normalizing by

$$\phi(t) := \prod_{s=1}^{t-1}\left(1+\frac{1}{2s}\right).$$

The following fact about $\phi(t)$ will be useful: there exist positive constants b_1 and b_2 such that

$$b_1\sqrt{t} \le \phi(t) \le b_2\sqrt{t} \quad \text{for all } t.$$

Proposition 4.1. *For each vertex j, the sequence $(X_t^{(j)})_{t\ge j}$ defined as*

$$X_t^{(j)} := \frac{d_t(j)}{\phi(t)}$$

is a martingale.

Proof. Since the vertex j will remain fixed throughout the proof, we will simply write X_t instead of $X_t^{(j)}$.

Observe that we can write $d_{t+1}(j)$ as

$$d_{t+1}(j) = d_t(j) + \sum_{k=1}^{m} \mathbf{1}(Y_{t+1}^{(k)} = j),$$

where $\mathbf{1}(\cdot)$ is the indicator function on the event $(\cdot)$. In addition, (4.1), proved in Lemma 4.1, ensures that for all $k \in 1, \dots, m$, we have

$$\mathbb{P}(Y_{t+1}^{(k)} = j \mid G_t) = \frac{d_t(j)}{2mt}. \tag{4.10}$$

https://doi.org/10.1017/apr.2018.41 Published online by Cambridge University Press

Thus, the follow equivalence relation holds:

$$\mathbb{E}[d_{t+1}(j) \mid G_t] = \left(1 + \frac{1}{2t}\right) d_t(j). \tag{4.11}$$

Then, dividing the above equation by $\phi(t+1)$, the desired result follows. □

Once we have Proposition 4.1 we are able to obtain an upper bound for $d_t(j)$.

Theorem 4.1. *There is a positive constant b_3 such that for all vertices j,*

$$\mathbb{P}(d_t(j) \geq b_3 \sqrt{t} \log(t)) \leq t^{-100}.$$

Proof. The proof is essentially applying Azuma's inequality to the martingale we obtained in Proposition 4.1. Again we will write it as X_t.

Applying Azuma's inequality demands controlling the X_t's variation, which satisfies the following upper bound:

$$|\Delta X_s| = \left| \frac{d_{s+1}(j) - (1 + 1/2s) d_s(j)}{\phi(s+1)} \right| \leq \frac{2m}{\phi(s+1)} \leq \frac{b_4}{\sqrt{s}}. \tag{4.12}$$

Thus,

$$\sum_{s=j+1}^{t} |\Delta X_s|^2 \leq b_5 \log(t).$$

We must note that none of the above constants depend on j. Then Azuma's inequality yields

$$\mathbb{P}(|X_t - X_0| > \lambda) \leq 2 \exp\left(-\frac{\lambda^2}{b_5 \log(t)} \right).$$

Choosing $\lambda = 10\sqrt{b_5} \log(t)$ and recalling $X_t = d_t(j)/\phi(t)$, we obtain

$$\mathbb{P}\left(\left| d_t(j) - \frac{m\phi(t)}{\phi(j)} \right| > 10\sqrt{b_5} \phi(t) \log(t) \right) \leq t^{-100}.$$

The reader may wonder why we need such a large exponent in the above inequality. For our purposes, it will be useful to have the polynomial decay with a large exponent so that we can make use of a union bound several times and still have all the bounds going to 0 polynomially fast. We state other theorems in this way throughout the paper.

Finally, using the fact that $b_1\sqrt{t} \leq \phi(t+1) \leq b_2\sqrt{t}$, we have

$$\mathbb{P}\left(\left| d_t(j) - \frac{m\phi(t)}{\phi(j)} \right| > 10\sqrt{b_5} b_2 \sqrt{t} \log(t) \right) \leq t^{-100},$$

implying the desired result. □

An immediate consequence of Theorem 4.1 is an upper bound for the maximum degree of G_t

Corollary 4.1. (Upper bound to the maximum degree.) *There exists a positive constant b_1 such that*

$$\mathbb{P}(d_{\max}(G_t) \geq b_1 \sqrt{t} \log(t)) \leq t^{-99}.$$

https://doi.org/10.1017/apr.2018.41 Published online by Cambridge University Press

Proof. The event involving $d_{\max}(G_t)$ may be written as

$$\{d_{\max}(G_t) \geq b_1\sqrt{t}\log(t)\} = \bigcup_{j\leq t}\{d_t(j) \geq b_1\sqrt{t}\log(t)\}.$$

Using a union bound and applying Theorem 4.1 completes the proof. □

The next three lemmas are of a technical nature. Their statements will become clearer in the proof of the upper bound for C_t.

Lemma 4.2. *There are positive constants b_6 and b_7 such that for all vertex j and all time $t_0 \leq t$, we have*

$$\mathbb{P}\left(d_{t_0}(j) > b_4\sqrt{\frac{t_0}{t}}d_t(j) + b_5\sqrt{t_0}\log(t)\right) \leq t^{-100}.$$

Proof. For each vertex j and $t_0 \leq t$, consider the sequence of random variables $(Z_s)_{s\geq 0}$ defined as $Z_s = X_{t_0+s}$, which is adaptable to the filtration $\mathcal{F}_s := G_{t_0+t}$.

Concerning the variation of the Z_s, using (4.12) we have the upper bound

$$|\Delta Z_s| = |\Delta X_{t_0+s}| \leq \frac{b_4}{\sqrt{t_0+s}}.$$

Thus,

$$\sum_{s=0}^{t-t_0}|\Delta Z_s|^2 \leq b_5\log(t).$$

Applying Azuma's inequality, we obtain

$$\mathbb{P}(|Z_{t-t_0} - Z_0| \geq \lambda) \leq 2\exp\left(-\frac{\lambda^2}{b_5\log(t)}\right). \tag{4.13}$$

However, the definition of Z_s and the fact that $\phi(t) = \Theta(\sqrt{t})$ means that the inclusion of events

$$\left\{d_{t_0}(j) > b_2\sqrt{t_0}\lambda + \frac{b_2\sqrt{t_0}}{b_1\sqrt{t}}d_t(j)\right\} \subset \{|Z_t - Z_0| \geq \lambda\}$$

hold, which, combined with (4.13), proves the lemma if we choose $\lambda = 10\sqrt{b_5}\log(t)$. □

Lemma 4.3. *There is a positive constant b_8 such that*

$$\mathbb{P}\left(\bigcup_{j=1}^{t}\bigcup_{t_0=j}^{t}\{d_{t_0}(j) > b_8\sqrt{t_0}\log(t)\}\right) \leq 2t^{-98}.$$

Proof. This lemma is a consequence of Theorem 4.1 and Lemma 4.2, which state, respectively,

$$\mathbb{P}(d_t(j) \geq b_3\sqrt{t}\log(t)) \leq t^{-100}, \qquad \mathbb{P}\left(d_{t_0}(j) > b_4\sqrt{\frac{t_0}{t}}d_t(j) + b_5\sqrt{t_0}\log(t)\right) \leq t^{-100},$$

in which the constants b_3, b_4, and b_5 do not depend on the vertex j nor do the times t_0 and t.

https://doi.org/10.1017/apr.2018.41 Published online by Cambridge University Press

Now, for each $t_0 \le t$ and vertex j, consider the events

$$A_{t_0,j} := \{d_{t_0}(j) > b_8\sqrt{t_0}\log(t)\}, \qquad B_{t_0,j} := \left\{d_{t_0}(j) > b_4\sqrt{\frac{t_0}{t}}d_t(j) + b_5\sqrt{t_0}\log(t)\right\},$$
$$C_{t,j} := \{d_t(j) \ge b_3\sqrt{t}\log(t)\}.$$

Now we obtain an upper bound for $\mathbb{P}(A_{t_0,j})$ using the bounds we obtained for the probabilities of $B_{t_0,j}$ and $C_{t,j}$. Thus,

$$\begin{aligned}\mathbb{P}(A_{t_0,j}) &= \mathbb{P}(A_{t_0,j} \cap B_{t_0,j}) + \mathbb{P}(A_{t_0,j} \cap B^c_{t_0,j})\\ &\le \mathbb{P}(B_{t_0,j}) + \mathbb{P}(A_{t_0,j} \cap B^c_{t_0,j} \cap C_{t,j}) + \mathbb{P}(A_{t_0,j} \cap B^c_{t_0,j} \cap C^c_{t,j})\\ &\le \mathbb{P}(B_{t_0,j}) + \mathbb{P}(C_{t_0,j}) + \mathbb{P}(A_{t_0,j} \cap B^c_{t_0,j} \cap C^c_{t,j}).\end{aligned}$$

However, note that we have the inclusion of events

$$B^c_{t_0,j} \cap C^c_{t,j} \subset \{d_{t_0}(j) \le (b_4b_3 + b_5)\sqrt{t_0}\log(t)\}.$$

Thus, choosing $b_8 = 2(b_4b_3 + b_5)$, we have $A_{t_0,j} \cap B^c_{t_0,j} \cap C^c_{t,j} = \varnothing$, which allows us to conclude that

$$\mathbb{P}(A_{t_0,j}) \le 2t^{-100}.$$

Finally, a union bound over t_0 followed by a union bound over j completes the proof. □

5. Positive local clustering

In this section we prove Theorem 1.1, which says that the local clustering coefficient is bounded away from 0 w.h.p.

Proof of Theorem 1.1. We need a lower bound for

$$\mathcal{C}^{\text{loc}}_{G_t} := \frac{1}{t}\sum_{v\in G_t}\mathcal{C}_{G_t}(v).$$

Let v_m be a vertex in G_t whose degree is m. Observe that each TF-step we took when v_m was added increases $e(\Gamma_{G_t}(v_m))$ by 1. So denote by T_v the number of TF-steps taken at the moment of creation of vertex v. Since all the choices of steps are made independently, T_v follows a binomial distribution with parameters $m-1$ and p. Now, for every vertex we add to the graph, put a blue label on it if $T_v \ge 1$. The probability of labeling a vertex is $1-(1-p)^{m-1}$ and we denote it by p_{b} in order to simplify our notation.

By Theorem 1.3, with probability at least $1 - t^{-100}$, we have

$$N_t(m) \ge b_1t - b_2\sqrt{t\log(t)},$$

where $b_1 = 2/(m+2)$ and $b_2 = 160m$. Thus, the number of vertices in G_t of degree m which were labeled, $N^{\text{b}}_t(m)$ is bounded from below by a binomial random variable B_t with parameters $b_1t - b_2\sqrt{t\log(t)}$ and p_{b}. But, about B_t we have, for all $\delta > 0$,

$$\mathbb{P}\big(B_t \le \tfrac{1}{4}\mathbb{E}[B_t]\big) \le (1 - p_{\text{b}} + p_{\text{b}}\mathrm{e}^{-\delta})^{b_1t-b_2\sqrt{t\log(t)}}\exp\big(\tfrac{1}{4}\delta p_{\text{b}}(b_1t - b_2\sqrt{t\log(t)})\big)$$

and choosing δ properly, we conclude that w.h.p.

$$N^{\text{b}}_t(m) \ge \tfrac{1}{4}p_{\text{b}}(b_1t - b_2\sqrt{t\log(t)}). \tag{5.1}$$

Finally, note that each blue vertex of degree m has $\mathcal{C}_{G_t}(v) > 2/m(m-1)$. Combining this with (5.1), we have

$$\begin{aligned}\mathcal{C}_{G_t}^{\mathrm{loc}} &\geq \frac{1}{t} \sum_{v \in N_t^{\mathrm{b}}(m)} \mathcal{C}_{G_t}(v) \\ &> \frac{t^{-1} N_t^{\mathrm{b}}(m) 2}{m(m-1)} \\ &\geq \frac{t^{-1} p_{\mathrm{b}}(b_1 t - b_2 \sqrt{t \log(t)})}{2m(m-1)} \\ &\to \frac{1-(1-p)^{m-1}}{(m+2)m(m-1)} \\ &> 0 \quad \text{as } t \to +\infty,\end{aligned}$$

proving the theorem. □

6. Vanishing global clustering

This section is devoted to the proof of Theorem 1.2 which states that the global clustering of G_t goes to 0 at $1/\log t$ speed. Since the proof depends on estimates for the number of cherries C_t, we first derive the necessary bounds and finally combine our results at the end of the section.

6.1. Preliminary estimates for number of cherries

Let

$$\tilde{C}_t := \sum_{j=1}^{t} d_t^2(j)$$

denote the sum of the squares of the degrees in G_t. We will prove bounds for $\tilde{C}_t$ instead of proving them directly for C_t. Since $C_t = \frac{1}{2}\tilde{C}_t - mt$, the results obtained for $\tilde{C}_t$ directly extend to C_t.

Lemma 6.1. *There is a positive constant B_3 such that*

$$\mathbb{E}[(\tilde{C}_{s+1} - \tilde{C}_s)^2 \mid G_s] \leq B_3 \frac{d_{\max}(G_s)\tilde{C}_s}{s}.$$

Proof. We start the proof by noting that for all vertices j, we have

$$d_{s+1}^2(j) - d_s^2(j) \leq 2m d_s(j) + m^2$$

deterministically. From this remark, it follows that

$$\tilde{C}_{s+1} - \tilde{C}_s \leq 2m \sum_{j=1}^{s} d_s(j)\, \mathbf{1}(\Delta d_s(j) \geq 1) + 2m^2.$$

Since all vertices have degree at least m, we have $m^2 \leq m \sum_{j=1}^{s} d_s(j)\, \mathbf{1}(\Delta d_s(j) \geq 1)$; thus,

$$\tilde{C}_{s+1} - \tilde{C}_s \leq 4m \sum_{j=1}^{s} d_s(j)\, \mathbf{1}(\Delta d_s(j) \geq 1).$$

https://doi.org/10.1017/apr.2018.41 Published online by Cambridge University Press

Applying the Cauchy–Schwarz inequality to the above inequality, we obtain

$$\begin{aligned}(\tilde{C}_{s+1}-\tilde{C}_s)^2 &\le 16m^2\Big(\sum_{j=1}^{s} d_s(j)\,\mathbf{1}(\Delta d_s(j)\ge 1)\,\mathbf{1}(\Delta d_s(j)\ge 1)\Big)^2\\ &\le 16m^2\Big(\sum_{j=1}^{s} d_s^2(j)\,\mathbf{1}(\Delta d_s(j)\ge 1)\Big)\Big(\sum_{j=1}^{s}\mathbf{1}(\Delta d_s(j)\ge 1)\Big)\\ &\le 16m^3\sum_{j=1}^{s} d_s^2(j)\,\mathbf{1}(\Delta d_s(j)\ge 1).\end{aligned}$$

Recalling that

$$\mathbb{P}(\Delta d_s(j)\ge 1 \mid G_s) \le \frac{d_s(j)}{2s},$$

we have

$$\mathbb{E}[(\tilde{C}_{s+1}-\tilde{C}_s)^2 \mid G_s] \le B_3\sum_{j=1}^{s}\frac{d_s^3(j)}{s} \le B_3\frac{d_{\max}(G_s)\tilde{C}_s}{s},$$

concluding the proof. □

Theorem 6.1. (Upper bound for C_t.) *There is a positive constant B_1 such that*

$$\mathbb{P}(C_t \ge B_1 t\log^2(t)) \le t^{-98}.$$

Proof. We show the result for $\tilde{C}_t$, which is greater than C_t. To do this, we need to determine $\mathbb{E}[d_{t+1}^2(j) \mid G_t]$.

As in the proof of Proposition 4.1, write

$$d_{t+1}(j) = d_t(j) + \sum_{k=1}^{m}\mathbf{1}(Y_{t+1}^{(k)} = j)$$

and denote $\sum_{k=1}^m \mathbf{1}(Y_{t+1}^{(k)} = j)$ by $\Delta d_t(j)$. Thus,

$$d_{t+1}^2(j) = d_t^2(j)\Big(1+\frac{\Delta d_t(j)}{d_t(j)}\Big)^2 = d_t^2(j) + 2d_t(j)\Delta d_t(j) + (\Delta d_t(j))^2.$$

Combining the above equation with (4.10) and (4.11), we obtain

$$\mathbb{E}[d_{t+1}^2(j) \mid G_t] = d_t^2(j) + \frac{d_t^2(j)}{t} + \mathbb{E}[(\Delta d_t(j))^2 \mid G_t].$$

Dividing the above equation by $t+1$, we have

$$\mathbb{E}\Big[\frac{d_{t+1}^2(j)}{t+1}\,\Big|\, G_t\Big] = \frac{d_t^2(j)}{t} + \frac{\mathbb{E}[(\Delta d_t(j))^2 \mid G_t]}{t+1},$$

which implies

$$\mathbb{E}\Big[\frac{\tilde{C}_{t+1}}{t+1}\,\Big|\, G_t\Big] = \frac{\tilde{C}_t}{t} + \frac{m^2}{t+1} + \sum_{j=1}^{t}\frac{\mathbb{E}[(\Delta d_t(j))^2 \mid G_t]}{t+1}. \tag{6.1}$$

https://doi.org/10.1017/apr.2018.41 Published online by Cambridge University Press

It is straightforward to see that $\Delta d_t(j) \le (\Delta d_t(j))^2 \le m\Delta d_t(j)$, which implies

$$\mathbb{E}[(\Delta d_t(j))^2 \mid G_t] = \Theta\left(\frac{d_t(j)}{t}\right).$$

Thus, (6.1) may be written as

$$\mathbb{E}\left[\frac{\tilde{C}_{t+1}}{t+1} \,\middle|\, G_t\right] = \frac{\tilde{C}_t}{t} + \Theta\left(\frac{1}{t}\right). \tag{6.2}$$

Now define

$$X_t := \frac{\tilde{C}_{t+1}}{t+1}.$$

Equation (6.2) states that X_t is a martingale up to a term of magnitude $\Theta(1/t)$. In order to apply martingale concentration inequalities, we decompose X_t as in Doob's decomposition theorem. Then X_t can be written as $X_t = M_t + A_t$, in which M_t is a martingale and A_t is a predictable process. By (6.2), we have

$$A_t = \sum_{s=2}^{t} \mathbb{E}[X_s \mid G_{s-1}] - X_{s-1} = \sum_{s=2}^{t} \Theta\left(\frac{1}{s}\right),$$

that is, $A_t = \Theta(\log(t))$ almost surely.

The remainder of the proof is devoted to controlling the martingale component of the X_t using Freedman's inequality (Theorem 2.2). Once again, by Doob's decomposition theorem, we have

$$M_t := X_0 + \sum_{s=2}^{t} X_s - \mathbb{E}[X_s \mid G_{s-1}].$$

Observe that $M_{t+1} = M_t + X_{t+1} - \mathbb{E}[X_{t+1} \mid G_t]$; thus,

$$\begin{aligned} |\Delta M_s| &= |X_{s+1} - \mathbb{E}[X_{s+1} \mid G_s]| \\ &\le |X_{s+1} - X_s| + \frac{b_9}{s} \\ &\le \left|\frac{\tilde{C}_{s+1} - (1+1/s)\tilde{C}_s}{s+1}\right| + \frac{b_9}{s} \\ &\le b_{10}\frac{d_{\max}(G_s)}{s} + b_{11}\frac{\tilde{C}_s}{s^2} + \frac{b_9}{s}, \end{aligned}$$

since $\Delta\tilde{C}_s$ attains its maximum when the vertices of maximum degree in G_s receive at least a new edge at time $s+1$. Furthermore, since $d_{\max}(G_s) \le ms$ and $\tilde{C}_s \le m^2 s^2$, there exists a constant b_{12} such that $\max_{s\le t}|\Delta M_s| \le b_{12}$ almost surely.

Combining

$$|\Delta M_s| \le \left|\frac{\tilde{C}_{s+1} - (1+1/s)\tilde{C}_s}{s+1}\right| + \frac{b_9}{s}$$

with the Cauchy–Schwarz inequality and Lemma 6.1, we obtain positive constants b_{13}, b_{14}, and b_{15} such that

$$\begin{aligned} \mathbb{E}[(\Delta M_s)^2 \mid G_s] &\le b_{13}\frac{\mathbb{E}[(\Delta\tilde{C}_s)^2 \mid G_s]}{s^2} + b_{14}\frac{\tilde{C}_s^2}{s^4} + \frac{b_{15}}{s^2} \quad \text{(Cauchy–Schwarz)} \\ &\le b_{16}\frac{d_{\max}(G_s)\tilde{C}_s}{s^3} + b_{14}\frac{\tilde{C}_s^2}{s^4} + \frac{b_{15}}{s^2} \quad \text{(Lemma 6.1)}. \end{aligned} \tag{6.3}$$

https://doi.org/10.1017/apr.2018.41 Published online by Cambridge University Press

Now define V_t as

$$V_t := \sum_{s=2}^{t} \mathbb{E}[(\Delta M_s)^2 \mid G_s]$$

and define a bad set as

$$B_t := \bigcup_{j=1}^{t} \bigcup_{t_0=j}^{t} \{d_{t_0}(j) > b_8 \sqrt{t_0} \log(t)\},$$

observing that Lemma 4.3 guarantees $\mathbb{P}(B_t) \le 2t^{-98}$.

Also note that $\tilde{C}_s \le b_{17} d_{\max}(G_s) s$ almost surely and in B_t^c, we have $d_{\max}(G_s) \le b_8 \sqrt{s} \log(t)$ for all $s \le t$. Then outside B_t, we have

$$\begin{aligned} V_t &\le \sum_{s=2}^{t} b_{16} \frac{d_{\max}(G_s)\tilde{C}_s}{s^3} + b_{14} \frac{\tilde{C}_s^2}{s^4} + \frac{b_{15}}{s^2} \quad \text{(from (6.3))} \\ &\le \sum_{s=2}^{t} \frac{b_{16} b_8^2 b_{17} s^2 \log^2(t)}{s^3} + \frac{b_{14} b_{17}^2 s^3 \log^2(t)}{s^4} + \frac{b_{15}}{s^2} \\ &\le b_{18} \log^3(t). \end{aligned} \tag{6.4}$$

So, by Freedman's inequality, we obtain

$$\mathbb{P}(M_t > \lambda, V_t \le b_{18} \log^3(t)) \le \exp\biggl(-\frac{\lambda^2}{2b_{18} \log^3(t) + 2b_{12}\lambda/3}\biggr).$$

Therefore, if $\lambda = b_{19} \log^2(t)$ with large enough b_{19}, we obtain

$$\mathbb{P}(M_t > b_{19} \log^2(t), V_t \le b_{18} \log^3(t)) \le t^{-100}.$$

Inequality (6.4) guarantees the following inclusion of events:

$$B_t^c \subset \{V_t \le b_{18} \log^3(t)\}.$$

Also,

$$\{X_t \ge b_{21} \log^2(t)\} \subset \{M_t \ge (b_{21} - b_{20}) \log^2(t)\}$$

since $A_t \le b_{20} \log(t)$ and $M_t \ge X_t - b_{20} \log(t)$.

Finally,

$$\begin{aligned} &\mathbb{P}(M_t > b_{19} \log^2(t)) \\ &\quad = \mathbb{P}(M_t > b_{19} \log^2(t), V_t \le b_{18} \log^3(t)) + \mathbb{P}(M_t > b_{19} \log^2(t), V_t > b_{18} \log^3(t)) \\ &\quad \le t^{-100} + \mathbb{P}(B_t) \end{aligned}$$

and

$$\mathbb{P}(M_t > b_{19} \log^2(t)) \le 3t^{-98},$$

proving the theorem. □

https://doi.org/10.1017/apr.2018.41 Published online by Cambridge University Press

We note that from (6.1), we may extract the recurrence

$$\mathbb{E}[\tilde{C}_t] = \left(1 + \frac{1}{t-1}\right)\mathbb{E}[\tilde{C}_{t-1}] + c_0,$$

in which c_0 is a positive constant depending on m and p only. Expanding it, we obtain

$$\mathbb{E}[\tilde{C}_t] = \prod_{s=1}^{t-1}\left(1 + \frac{1}{s}\right)\mathbb{E}[\tilde{C}_1] + c_0 \sum_{s=1}^{t-1}\prod_{r=s}^{t-1}\left(1 + \frac{1}{r}\right),$$

which implies $\mathbb{E}[\tilde{C}_t] = \Theta(t \log t)$. This means the upper bound for $\tilde{C}_t$ given by Theorem 6.1 is exactly $\mathbb{E}[\tilde{C}_t]\log(t)$.

6.2. The bootstrap argument

Obtaining bounds for C_t requires some control of its quadratic variation, which requires bounds for the maximum degree and C_t, as in Lemma 6.1. Applying some deterministic bounds and upper bounds on the maximum degree, we are able to derive an upper bound for C_t, which is of order $\mathbb{E}[C_t]\log t$. To improve this bound and obtain the correct order, we proceed as in the proof of Theorem 6.1, but making use of the preliminary estimate just discussed. This is what we call *the bootstrap argument*.

The result we obtain is enunciated in Theorem 1.4 and consists of a weak law of large numbers, which states that C_t divided by $t \log t$ actually converges in probability to a constant depending only on m.

Proof of Theorem 1.4. In the proof of Theorem 6.1, we decomposed the process $X_t = \tilde{C}_t/t$ into two components: M_t and A_t. The first part of the proof will be dedicated to showing that $M_t = o(\log(t))$ w.h.p. Then we show that $A_t = (m^2 + m)\log(t)$ also w.h.p.

We repeat the proof given for Theorem 6.1, but this time we change our definition of a bad set to

$$B_t = \bigcup_{s=\log^{1/2}(t)}^{t} \{\tilde{C}_s \geq b_{20} s \log^2(s)\}. \tag{6.5}$$

By Theorem 6.1 and a union bound, $\mathbb{P}(B_t) \leq \log^{-97/2}(t)$. Observe that an upper bound for $\tilde{C}_s$ yields an upper bound for $d_{\max}(G_s)$, since

$$d^2_{\max}(G_s) \leq \tilde{C}_s \quad \Longrightarrow \quad d_{\max}(G_s) \leq \sqrt{\tilde{C}_s} \quad \Longrightarrow \quad d_{\max}(G_s) \leq \sqrt{s}\log(s)$$

when $\tilde{C}_s \leq s \log^2(s)$.

Using (6.3), we have, in B_t^c,

$$\begin{aligned} V_t &\leq \sum_{s=1}^{t-1} b_{16}\frac{d_{\max}(G_s)\tilde{C}_s}{s^3} + b_{14}\frac{\tilde{C}_s^{\,2}}{s^4} + \frac{b_{14}}{s^2} \\ &\leq \sum_{s=1}^{\log^{1/2}(t)-1} b'_{16} + \sum_{s=\log^{1/2}(t)}^{t-1} b_{17}\frac{\sqrt{s}\log(s)s\log^2(s)}{s^3} + b_{18}\frac{s^2\log^4(s)}{s^4} + \frac{b_{14}}{s^2} \\ &\leq b_{19}\log^{1/2}(t), \end{aligned} \tag{6.6}$$

https://doi.org/10.1017/apr.2018.41 Published online by Cambridge University Press

since $d_{\max}(G_s) \le ms$ and $\tilde{C}_s \le 2m^2 s^2$ for all s and in B_t^c, $d_{\max}(G_s) \le \sqrt{b_{20} s}\log(s)$ and $\tilde{C}_s \le b_{20} s \log^2(s)$ for all $s \ge \log^{1/2}(t)$. Then, by Freedman's inequality,

$$\mathbb{P}(M_t \ge \log^{1/4+\delta}(t),\, V_t \le b_{19}\log^{1/2}(t)) = o(1). \tag{6.7}$$

Recall (6.1). Now, from Lemma 4.1, we recall that for all $k \in \{1, \dots, m\}$,

$$\mathbb{P}(Y_{t+1}^{(k)} = v \mid G_t) = \frac{d_t(v)}{2mt}.$$

Furthermore, there exists a positive constant $\tilde{c}_{m,p}$ depending only on m and p such that for all time t and vertex v, we have

$$\mathbb{P}(Y_{t+1}^{(k)} = v, Y_{t+1}^{(j)} = v \mid G_t) \le \tilde{c}_{m,p}\frac{d_t^2(v)}{t^2}.$$

Thus, there exists another constant $\tilde{c}_{m,p}$ with the same conditions as above such that

$$\left|\mathbb{E}[(\Delta d_t(v))^2 \mid G_t] - \frac{d_t(v)}{2t}\right| \le \tilde{c}_{m,p}\frac{d_t^2(v)}{t^2}, \tag{6.8}$$

which, in turn, implies that

$$\left|\mathbb{E}\left[\frac{\tilde{C}_{t+1}}{t+1}\,\middle|\, G_t\right] - \frac{\tilde{C}_t}{t} - \frac{m^2+m}{t+1}\right| \le \tilde{c}_{m,p}\frac{\tilde{C}_t}{t^3},$$

and, consequently, there exists a positive constant b_{11} such that

$$\left|A_t - \sum_{s=2}^{t}\frac{m^2+m}{s+1}\right| \le \sum_{s=2}^{t} b_{11}\frac{\tilde{C}_s}{s^3}.$$

Then, by the definition of the event B_t from (6.5), we deduce that, in B_t^c,

$$\sum_{s=2}^{t} b_{11}\frac{\tilde{C}_s}{s^3} \le b_{11}\sum_{s=2}^{\log^{1/2}(t)}\frac{\tilde{C}_s}{s^3} + \sum_{s=\log^{1/2}(t)}^{t}\frac{s\log^2(s)}{s^3} \le b_{12}\log(\log(t)).$$

Thus, in B_t^c,

$$A_t = (m^2+m)\log(t) + o(\log(t)).$$

Finally, fix a small positive ε. Then

$$\begin{aligned}\mathbb{P}\left(\left|\frac{\tilde{C}_t}{t\log(t)} - m^2 - m\right| > \varepsilon\right) &= \mathbb{P}\left(\left|\frac{M_t + A_t}{\log(t)} - m^2 - m\right| > \varepsilon\right)\\ &\le \mathbb{P}\left(\left|\frac{M_t + A_t}{\log(t)} - m^2 - m\right| > \varepsilon,\, B_t^c\right) + \mathbb{P}(B_t).\end{aligned} \tag{6.9}$$

We also have $B_t^c \subset \{V_t \le b_{19}\log^{1/2}(t)\}$, which, combined to (6.7), implies that

$$\mathbb{P}\left(\left|\frac{M_t + A_t}{\log(t)} - m^2 - m\right| > \varepsilon,\, M_t \ge \log^{1/4+\delta}(t),\, B_t^c\right) = o(1).$$

https://doi.org/10.1017/apr.2018.41 Published online by Cambridge University Press

Observe that, in the event B_t^c, M_t is bounded by $\log^{1/4+\delta}(t)$, as assured by (6.6), and $A_t = (m^2+m)\log(t) + o(\log(t))$; thus,

$$\mathbb{P}\left(\left|\frac{M_t + A_t}{\log(t)} - m^2 - m\right| > \varepsilon, M_t < \log^{1/4+\delta}(t), B_t^c\right) = 0 \quad \text{for large enough } t.$$

This proves that

$$\frac{\tilde{C}_t}{t\log t} \xrightarrow{\mathbb{P}} m^2 + m.$$

Since $C_t = \tilde{C}_t/2 - mt$, as defined at the beginning of this section, we obtain

$$\frac{C_t}{t\log t} \xrightarrow{\mathbb{P}} \binom{m+1}{2},$$

which proves the theorem. □

6.3. Wrapping up

So far we have devoted our efforts to properly controlling the number of cherries in G_t. Now we combine these results with simple bounds for the number of triangles in G_t to finally obtain the exact order of the global clustering.

Proof of Theorem 1.2. We begin by pointing out that the number of triangles in G_t, Δ_{G_t}, is bounded from above by $\binom{m}{2}t$ and that every TF-step we take increases Δ_{G_t} by 1. Then

$$\Delta_{G_t} \geq Z_t = \sum_{s=1}^{t} T_s,$$

where T_s is the number of TF-steps taken at time s. Since all the choices regarding the types of step we follow are independent, $T_s \sim \mathrm{bin(m-1, p)}$ and $Z_t \sim \mathrm{bin((m-1)t, p)}$. Using Chernoff bounds, we are able to obtain

$$\mathbb{P}\left(Z_t \leq \frac{(m-1)p}{2}t\right) \leq \mathrm{e}^{-(m-1)pt/8}.$$

Combining the above inequality with the deterministic upper bound for Δ_{G_t}, we have w.h.p.

$$\frac{(m-1)p}{2}t \leq \Delta_{G_t} \leq \frac{m(m-1)}{2}t. \tag{6.10}$$

Now we proceed to bound C_t. Taking $\varepsilon = \frac{1}{2}$ in (6.9), we obtain w.h.p.

$$\left(m^2 + m - \tfrac{1}{2}\right)t\log t \leq \tilde{C}_t \leq (m^2+m+1)t\log t,$$

which implies that

$$\tfrac{1}{2}\left(m^2 + m - \tfrac{1}{2}\right)t\log(t) - mt \leq C_t \leq \tfrac{1}{2}(m^2+m+1)t\log(t) - mt,$$

which, for large enough t, may be written in the more symmetric and simpler form as

$$\tfrac{1}{2}(m^2+m-1)t\log(t) \leq C_t \leq \tfrac{1}{2}(m^2+m+1)t\log(t) \quad \text{w.h.p.}$$

Combining the above inequalities with (6.10), we prove that, also w.h.p.,

$$\frac{(m-1)p}{(m^2+m+1)\log(t)} \leq \frac{\Delta_{G_t}}{C_t} \leq \frac{m(m-1)}{(m^2+m-1)\log(t)}.$$

Multiplying the above inequalities by 3, we conclude the proof. □

7. Final comments on clustering

We conclude this paper by comparing the two clustering coefficients from a different perspective than in Section 1.4. Recall that $\mathcal{C}^{\mathrm{loc}}_{G_t}$ is an unweighted average of local clustering coefficients, that is,

$$\mathcal{C}^{\mathrm{loc}}_{G_t} := \frac{1}{t} \sum_{v \in G_t} \mathcal{C}_{G_t}(v).$$

On the other hand, $\mathcal{C}^{\mathrm{glo}}_{G_t}$ is a weighted average, where the weight of vertex v is the number of cherries that it belongs to, that is,

$$\mathcal{C}^{\mathrm{glo}}_{G_t} = 3 \times \frac{\sum_{v \in G_t} \mathcal{C}_{G_t}(v)\binom{d_t(v)}{2}}{\sum_{v \in G_t} \binom{d_t(v)}{2}}.$$

Thus, the weight of v in $\mathcal{C}^{\mathrm{glo}}_{G_t}$ is essentially proportional to the square of the degree. This skews the distribution of weights towards high-degree nodes. The clustering of the high-degree vertices is the reason why the two coefficients present such distinct behaviors.

We will show below that $\mathcal{C}_{G_t}(v)$ for a vertex v of high degree d is of order d^{-1}, which explains why $\mathcal{C}^{\mathrm{glo}}_{G_t}$ goes to 0. Recall that the random variable $e_t(\Gamma_v)$ counts the number of edges between the neighbors of v. Due to the definition of our model, we can increase $e_t(\Gamma_v)$ by only 1 if $d_t(v)$ is also increased by at least one unit. Since $e_t(\Gamma_v)$ can increase by only m units in each time step, we have

$$e_t(\Gamma_v) \le m\, d_t(v),$$

which implies an upper bound for $\mathcal{C}_{G_t}(v) \approx e_t(\Gamma_v)/d_t(v)^2$ of order $d_t^{-1}(v)$. In the next proposition we state a lower bound of the same order.

Proposition 7.1. *Let v be a vertex of G_t. Then there are positive constants, b_1 and b_2 such that*

$$\mathbb{P}\bigg(\mathcal{C}_{G_t}(v) \le \frac{b_1}{d_t(v)} \,\bigg|\, d_t(v) \ge b_2 \log(t)\bigg) \le t^{-100}.$$

This proposition does not prove our clustering estimates, but seems interesting in any case.

Proof of Proposition 7.1. Observe that if we choose v and take a TF-step thereafter, we increase $e_t(\Gamma_v)$ by 1. Then, if we look only at times in which this occurs, $e_t(\Gamma_v)$ must be greater than a binomial random variable with parameters: number of times we choose v at the first choice and p. Since all the choices concerning the type of step we take are made independently of the whole process, we just need to prove that the number of times we choose v at the first choice, denoted by $d_t^{(1)}(v)$, is proportional to $d_t(v)$ w.h.p.

Recall that $Y_s^{(1)}$ indicates the vertex chosen at time s at the first of our m choices. The random variable $d_t^{(1)}(v)$ can be written in terms of the $Y^{(1)}$ as

$$d_t^{(1)}(v) = \sum_{s=v+1}^{t} \mathbf{1}(Y_s^{(1)} = v).$$

We first claim that if $d_t(v)$ is large enough, a positive fraction of its value must come from $d_t^{(1)}(v)$.

Claim 7.1. *There exist positive constants b_1 and b_2 such that*

$$\mathbb{P}(d_t^{(1)}(v) \le b_1 d_t(v) \mid d_t(v) \ge b_2 \log(t)) \le t^{-100}.$$

Proof. To prove the claim we condition on all possible trajectories of $d_t(v)$. In this direction, let ω be an event describing when v was chosen and by how many times at each step. We have to note that ω does not record whether v was chosen by a PA-step or a TF-step. The event ω can be regarded as a vector in $\{0, 1, \dots, m\}^{t-v-1}$ such that $\omega(s) = k$ means we have chosen v k-times at time s. For each ω, let $d_\omega(v)$ be the degree of v obtained by the sequence of choices given by ω.

Recall (4.8) and (4.9). For any ω such that $\omega(s) = k \geq 1$, we may show, using (4.8) and (4.9), that there exists a positive constant δ depending only on m and p such that

$$\mathbb{P}(Y_s^{(1)} = v \mid \omega) \geq \delta.$$

Furthermore, given ω, the random variables $\mathbf{1}(Y_s^{(1)} = v)$ are independent. This implies that, given ω, the random variable $d_t^{(1)}(v)$ dominates stochastically another random variable following a binomial distribution of parameters $d_\omega(v)/m$ and δ. Thus, using Chernoff bounds, we can choose a small b_1 such that

$$\mathbb{P}(d_t^{(1)}(v) \leq b_1 d_\omega(v) \mid \omega) \leq \exp(-d_\omega(v)).$$

Since we are on the event $D_t := \{d_t(v) \geq b_2 \log(t)\}$, all $d_\omega(v) \geq b_2 \log(t)$ for some b_2 can be chosen in a way such that

$$\mathbb{P}(d_t^{(1)} \leq b_1 d_\omega(v) \mid \omega) \leq t^{-100} \quad \text{for all } \omega \text{ compatible with } D_t.$$

Finally, to estimate $\{d_t^{(1)}(v) \leq b_1 d_t(v)\}$, we condition on all the possible history of choices ω, that is,

$$\mathbb{P}(d_t^{(1)}(v) \leq b_1 d_t(v) \mid D_t) = \sum_{\omega} \mathbb{P}(d_t^{(1)}(v) \leq b_1 d_t(v) \mid \omega, D_t)\mathbb{P}(\omega \mid D_t) \leq t^{-100}$$

and this proves the claim. □

Returning to the proof of Proposition 7.1, as we observed at the beginning, $e_t(\Gamma_v)$ dominates a random variable $\mathrm{bin}(\mathrm{d}_t^{(1)}(\mathrm{v}), \mathrm{p})$. And, from Claim 7.1, $d_t^{(1)}(v)$ is proportional to $d_t(v)$ w.h.p. Using Chernoff bounds, we obtain the result. □

Appendix A. Power-law degree distribution

Lemma A.1. (See [6, Lemma 3.1].) *Let a_t be a sequence of positive real numbers satisfying the recurrence relation*

$$a_{t+1} = \left(1 - \frac{b_t}{t}\right)a_t + c_t.$$

Furthermore, suppose that $b_t \to b > 0$ and $c_t \to c$. Then

$$\lim_{t\to\infty} \frac{a_t}{t} = \frac{c}{1+b}.$$

Proof of Theorem 1.3. We divide the proof into two parts: (a) is the power law for the expected value of the proportion of vertices with degree d, and (b) is the concentration inequalities $N_t(d)$.

https://doi.org/10.1017/apr.2018.41 Published online by Cambridge University Press

(a) The argument is essentially the same as that of [6, Section 3.2]. The key step is to obtain a recurrence relation involving $\mathbb{E}[N_t(d)]$ which has the same form of that required by Lemma A.1. To obtain the recurrence relation, observe that $N_{t+1}(d)$ can be written as

$$N_{t+1}(d) = \sum_{v \in N_t(d)} \mathbf{1}(\Delta d_t(v) = 0) + \sum_{v \in N_t(d-1)} \mathbf{1}(\Delta d_t(v) = 1) + \cdots + \sum_{v \in N_t(d-m)} \mathbf{1}(\Delta d_t(v) = m).$$

Taking the conditional expected value with respect to G_t in the above equation, applying Lemma 4.1, and recalling that $N_t(d) \leq t$, we obtain

$$\mathbb{E}[N_{t+1}(d) \mid G_t] = N_t(d)\bigg[1 - \frac{d}{2t} + O\bigg(\frac{d^2}{t^2}\bigg)\bigg] + N_t(d-1)\bigg[\frac{d-1}{2t} + O\bigg(\frac{(d-1)^2}{t^2}\bigg)\bigg] + O\bigg(\frac{1}{t}\bigg).$$

Finally, taking the expected value on both sides and denoting $\mathbb{E}N_t(d)$ by $a_t^{(d)}$, we have

$$a_{t+1}^{(d)} = \bigg[1 - \frac{d/2 + O(d^2/t)}{t}\bigg]a_t^{(d)} + a_t^{(d-1)}\bigg[\frac{(d-1)/2 + O((d-1)^2/t)}{t}\bigg] + O\bigg(\frac{1}{t}\bigg).$$

From here the proof follows by an induction on $d \geq m$ and and application of Lemma A.1, assuming that $\mathbb{E}N_t(d-1)/t \longrightarrow D_{d-1}$, which yields

$$\frac{a_t(d)}{t} \to \frac{D_{d-1}(d-1)/2}{1 + d/2} = D_{d-1}\frac{d-1}{2+d} =: D_d$$

leading to

$$D_d = \frac{2}{2+m}\prod_{k=m+1}^{d}\frac{k-1}{k+2} = \frac{2(m+1)m}{(d+2)(d+1)d},$$

which proves part (a).

(b) The proof is in line with the proof of [6, Theorem 3.2]. For this reason we show only that the process

$$X_t^{(d)} := \frac{N_t(d) - D_d t + 16dc\sqrt{t}}{\psi_d(t)},$$

in which $\psi_d(t)$ is defined as

$$\psi_d(t) := \prod_{s=d}^{t-1}\bigg(1 - \frac{d}{2s}\bigg),$$

is a submartingale and we state an upper bound for its variation.

As in [6, Theorem 3.2], the proof follows by induction on d.

Inductive step. Suppose that for all $d' \leq d-1$, we have

$$\mathbb{P}(N_t(d') \leq D_{d'}t - 16d'c\sqrt{t}) \leq (t+1)^{d'-m}\mathrm{e}^{-c^2}. \tag{A.1}$$

Recalling that

$$\mathbb{E}[N_{t+1}(d) \mid G_t] = \bigg(1 - \frac{d}{2t} + O\bigg(\frac{d^2}{t^2}\bigg)\bigg)N_t(d) + \frac{(d-1)N_t(d-1)}{2t} + O(t^{-1}),$$

we have the recurrence relation

$$\mathbb{E}[\psi_d(t+1)X_t^{(d)} \mid G_t] \geq \left(1-\frac{d}{2t}\right)N_t(d) + \frac{(d-1)N_t(d-1)}{2t} + O(t^{-1}) + 16dc\sqrt{t} - D_d(t+1). \quad \text{(A.2)}$$

The inductive hypothesis ensures that

$$N_t(d-1) \geq D_{d-1}t - 16(d-1)c\sqrt{t}$$

with probability at least $1-(t+1)^{d-1-m}\mathrm{e}^{-c^2}$. Thus, returning to (A.2),

$$\mathbb{E}[\psi_d(t+1)X_t^{(d)} \mid G_t] \geq \left(1-\frac{d}{2t}\right)N_t(d) + \frac{(d-1)D_{d-1}}{2} - D_d(t+1) + 16dc\sqrt{t} + O(t^{-1}). \quad \text{(A.3)}$$

However, observe that for the right-hand side of the above inequality, we have

$$\begin{aligned}
&\frac{(d-1)D_{d-1}}{2} - D_d(t+1) + 16dc\sqrt{t} + O(t^{-1}) \geq \left(1-\frac{d}{2t}\right)(-D_dt + 16dc\sqrt{t})\\
&\iff \quad \frac{(d-1)D_{d-1}}{2} - D_d + O(t^{-1}) \geq \frac{dD_d}{2t} - \frac{8d^2c}{\sqrt{t}}\\
&\iff \quad \frac{(d-1)D_{d-1}}{2} + \frac{8d^2c}{\sqrt{t}} + O(t^{-1}) \geq D_d + \frac{dD_d}{2t}
\end{aligned}$$

with the last inequality holding since we have $D_d = 2(m+1)m/(d+2)(d+1)d$ and $(d-1)D_{d-1} = (d+2)D_d$. Returning to (A.3), we have just proved that $X_{t+1}^{(d)}$ is a submartingale with fail probability bounded from above by $(t+1)^{d-m}\mathrm{e}^{-c^2}$. Its variation $\Delta X_t^{(d)}$ satisfies the upper bound

$$\begin{aligned}
|\Delta X_s^{(d)}| &\leq \frac{\Delta N_s(d) + D_d + 16dcs^{-1/2} + dN_s(d)(2s)^{-1}}{\psi_d(s+1)}\\
&\leq \frac{m + 2/(d+2) + 17dcs^{-1/2} + d/2}{\psi_d(s+1)}\\
&\leq \frac{2d}{\psi_d(s+1)} + \frac{17dc}{\sqrt{s}\psi_d(s+1)},
\end{aligned}$$

since $\Delta N_s(d) \leq m$, $N_s(d) \leq s$, and $D_d \leq 2/(d+2)$ for all s and d. Thus, there is a positive constant M such that

$$|\Delta X_s^{(d)}|^2 \leq \frac{16d^2}{\psi_d^2(s+1)} + \frac{Md^2c^2}{s\psi_d^2(s+1)}. \quad \text{(A.4)}$$

The lower bound for $N_t(d)$ is proved by applying [6, Theorem 2.36] on $X_t^{(d)}$, and setting

$$\lambda = 2c\sqrt{\sum_{s=d}^{t+1}|\Delta X_s^{(d)}|^2}.$$

The upper bound is obtained the same way, but considering the process

$$^{-}X_t^{(d)} := \frac{N_t(d) - D_dt - 16dc\sqrt{t}}{\psi_d(t)},$$

which is a supermartingale. □

https://doi.org/10.1017/apr.2018.41 Published online by Cambridge University Press

References

[1] Barabási, A.-L. and Álbert, R. (1999). Emergence of scaling in random networks. *Science* **286,** 509–512.
[2] Bollobás, B. (2001). *Random Graphs*, 2nd edn. Cambridge University Press.
[3] Bollobás, B. and Riordan, O. M. (2003). Mathematical results on scale-free random graphs. In *Handbook of Graphs and Networks*, Wiley-VCH, Weinheim, pp. 1–34.
[4] Bollobás, B., Riordan, O., Spencer, J. and Tusnády, G. (2001). The degree sequence of a scale-free random graph process. *Random Structures Algorithms* **18,** 279–290.
[5] Buckley, P. G. and Osthus, D. (2004). Popularity based random graph models leading to a scale-free degree sequence. *Discrete Math.* **282,** 53–68.
[6] Chung, F. and Lu, L. (2006). *Complex Graphs and Networks*. American Mathematical Society, Providence, RI.
[7] Durrett, R. (2007). *Random Graph Dynamics*. Cambridge University Press.
[8] Erdős, P. and Rényi, A. (1959). On random graphs. I. *Publ. Math. Debrecen* **6,** 290–297.
[9] Freedman, D. A. (1975). On tail probabilities for martingales. *Ann. Prob.* **3,** 100–118.
[10] Holme, P. and Kim, B. J. (2002). Growing scale-free networks with tunable clustering. *Phys. Rev. E* **65,** 026107.
[11] Janson, S., Łuczak, T. and Rucinski, A. (2000). *Random Graphs*. John Wiley, New York.
[12] Kumar, R. *et al.* (2000). Stochastic models for the web graph. In *Proceedings of the 41st Annual Symposium on the Foundations of Computer Science,* IEEE, pp. 57–65.
[13] Newman, M. E. J. (2010). *Networks: An Introduction*. Oxford University Press.
[14] Ostroumova, L., Ryabchenko, A. and Samosvat, E. (2013). Generalized preferential attachment: Tunable power-law degree distribution and clustering coefficient. In *Algorithms and Models for the Web Graph,* Springer, Cham, pp. 185–202.
[15] van der Hofstad, R. (2009). Random graphs and complex networks. Available at http://www.win.tue.nl/∼rhofstad/NotesRGCN.html.
[16] Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature* **393,** 440–442.