Math. Struct. in Comp. Science (2012), *vol.* 22, *pp.* 373–407. © Cambridge University Press 2012
doi:10.1017/S0960129511000570 First published online 6 March 2012

The small-community phenomenon in networks†

ANGSHENG LI‡ and PAN PENG§

‡*State Key Laboratory of Computer Science, Institute of Software,*
Chinese Academy of Sciences, P.O. Box 8718, Beijing, 100190, P.R. China
Email: angsheng@ios.ac.cn
§*State Key Laboratory of Computer Science, Institute of Software,*
Chinese Academy of Sciences, P.O. Box 8718, Beijing, 100190, P.R. China
and
School of Information Science and Engineering,
Graduate University of China Academy of Sciences, P.R. China
Email: pengpan@ios.ac.cn.

Received 21 April 2011; revised 12 August 2011

We investigate several geometric models of networks that simultaneously have some nice global properties, including the small-diameter property, the *small-community phenomenon*, which is defined to capture the common experience that (almost) everyone in society also belongs to some meaningful *small* communities, and the power law degree distribution, for which our result significantly strengthens those given in van den Esker (2008) and Jordan (2010). These results, together with our previous work in Li and Peng (2011), build a mathematical foundation for the study of both communities and the small-community phenomenon in various networks.
In the proof of the power law degree distribution, we develop the method of *alternating concentration analysis* to build a concentration inequality by alternately and iteratively applying both the sub- and super-martingale inequalities, which seems to be a powerful technique with further potential applications.

1. Introduction

With the availability of massive datasets of many real-world networks, we can make quantitative observations and studies of the underlying dynamic mechanisms and many interesting phenomena occurring in large-scale networks. Some properties such as sparseness, high-clustering, hierarchical structure, the power law degree distribution and small diameters appear in a wide range of networks, ranging from internet graphs and collaboration graphs to PPI (Protein–Protein Interaction) networks. Modelling these interesting properties and phenomena not only provides us with a good way to understand better how these networks evolve, and why these global phenomena occur through local

† This research was partially supported by NSFC distinguished young investigator award number 60325206, and its matching fund from the Hundred-Talent Program of the Chinese Academy of Sciences. Both authors are partially supported by the Grand Project 'Network Algorithms and Digital Information' of the Institute of Software, Chinese Academy of Sciences.
§ Corresponding author.

growth rules, but also gives us insights into the development of new technologies, or even cancer drugs.

A typical network always exhibits several properties at the same time. For example, in a web graph, the nodes are web-pages and directed edges are hyperlinks between the pages, the number of nodes with in-degree k is proportional to $k^{-\beta}$ for some constant β, that is, the in-degree sequence obeys the power law degree distribution (see, for example, Albert *et al.* (1999) and Kleinberg *et al.* (1999)). It has also been observed that web graphs have a small average distance (Albert *et al.* 1999; Broder *et al.* 2000). (In this paper, when there is no risk of confusion, we will use 'small' to mean that the quantity is a polylogarithmic function of the number of graph nodes.) Furthermore, the most community-like subgraphs in large web graphs turn out to have size about 100, which seems to be a general property of many real large networks (Leskovec *et al.* 2008; 2010). The three properties mentioned above are by no means specific to technological networks, and also appear in a wide range of social networks, such as the friendship network of LiveJournal.

The first two properties, that is, the power law degree distribution and the small-diameter property, have been explored extensively in recent decades. However, to our knowledge, the third property – that good communities in large-scale networks have small sizes – is still largely unexplored since there has been no mathematical definition of what we mean by good communities in networks, and this has motivated us to carry out a mathematical study of the common experiences, observations and experiments related to the small-community phenomenon in networks.

The authors of the current paper have proposed a mathematical definition for communities based on the concept of conductance and have given a definition of the *small-community phenomenon* in networks. We then conjectured that small communities are ubiquitous in various networks (Li and Peng 2011). Intuitively, a given network is said to have the small-community phenomenon if almost every node in the network is contained in some good community of small size (see Section 2 for the formal definition). We have found theoretical evidence for our conjecture in that some classical network models, including Kleinberg's small-world model (Kleinberg 2000) and the Ravasz–Barabási Hierarchical model (Ravasz and Barabási 2003), do indeed exhibit the small-community phenomenon, though there are also models that do not.

There are also other reasons for us to make this conjecture. First, we have all experienced the fact that everyone in a society belongs to some small meaningful subgroups such as classmates, friends or relatives. Second, existing empirical studies provide evidence that large communities are rare in large networks, and good communities are small. In addition to the direct evidence given in Leskovec *et al.* (2008; 2010) and Groh and Rappel (2009), there is also some indirect evidence. For example, Lang (2005) showed that spectral graph partitioning fails to generate highly unbalanced cuts for many large-scale social networks, and Kurucz and Benczúr (2010) pointed out that this failure may be caused by an abundance of small dense communities. In summary, we have reasons to conjecture that small communities are ubiquitous, at least in many large social networks, which raises a number of new problems for both the theory and applications of the small-community phenomenon in networks.

We are interested in *evolving* models that have all three of these 'good' properties:

— the power law degree distribution;
— the small-diameter property; and
— the small-community phenomenon,

all of which are found in typical web graphs and large social networks. Models with one or two of these properties are easily constructed in a natural way. In particular, the power law degree distribution arises from the preferential attachment scheme; the small-diameter property follows from a broad class of graph processes (Bollobás 2001); and the small-community property may be generated by *homophily* – the tendency for close or similar individuals to associate with each other, which is commonly observed as a reason for two people establishing a relationship with each other.

However, when we try to define a model that encompasses all three properties, we often come across conflicts that are hard to reconcile. Roughly speaking, the first two properties usually result from some expander-like graphs, while the small-community property corresponds to highly structured graphs, which seem to be, to some extent, like anti-expander graphs (Li and Peng 2011). However, as Li and Peng (2011) shows, the Ravasz–Barabási Hierarchical model (Ravasz and Barabási 2003) satisfies all three requirements. On the other hand, the Ravasz–Barabási Hierarchical model has a very unnatural growth rule, which can only capture very special networks.

Another good candidate may be the Geometric Preferential Attachment (GPA) model introduced in Flaxman *et al.* (2007a; 2007b), where the motivation was to model networks with a power law degree distribution and small expansion. This model is defined on a unit-area spherical surface S, for which a natural distance can be introduced. Flaxman *et al.* combined the rich-get-richer effect and the concept of homophily in a simple way such that every newcomer only chooses neighbours from those existing vertices that are close to them using the preferential attachment scheme, and then proved that the power law distribution occurs when the parameters in the model satisfy certain conditions. We showed in Li and Peng (2011) that good communities exist for every node in a model under these conditions. However, the resulting communities, and the diameter, are relatively large.

In the current paper, we will first study a base model, which is a GPA model with *additive fitness*. We will generalise the result of Jordan (2010) by showing that, under appropriate conditions, the base model has both a power law degree distribution and the small-community phenomenon. However, the diameter of the model is large in this situation. To resolve this problem, we try to incorporate a simple growth rule into our base model that leads to the small-diameter property and does not change the degree sequence too much. The rule we try is the uniform recursive tree: that is, each time a new vertex is generated, it chooses a neighbour uniformly at random from existing vertices. It is well known that such a simple process results in a graph of diameter and maximum degree of order $\Theta(\ln n)$, where n is the number of generated vertices (Smythe and Mahmoud 1995). We give two alternative ways to incorporate this rule, and although the resulting models are similar, their structures are different. The first is a *hybrid model*, which can be viewed as a composition of two independent parts: *a local graph*, which has

https://doi.org/10.1017/S0960129511000570 Published online by Cambridge University Press

the power law degree distribution, and *a global graph*, which may connect vertices that are far away. The hybrid model as a whole has both the small-diameter property and a small-community structure. The second model is a *self-loop model* in which we treat the additive fitness in our base model as the number of self-loops attached with the new vertex. This gives a new interpretation for the use of fitness in preferential attachment schemes. With some further development, the self-loop model is shown to have all three of our good properties.

The methodology we use to show the power law degree distribution has some independent interest. The proof technique is inspired by the work in Jordan (2010), which investigated the *asymptotic behaviour* of the degree sequence of the base model (see Section 2). In our proof of the concentration inequalities, there are subtle restrictions on the parameters for which deeper mathematics is needed. Rather than using the coupling techniques used in Flaxman *et al.* (2007a) and van den Esker (2008), we use the submartingale and supermartingale concentration inequalities recursively (Chung and Lu 2006) to give a better bound at each step, and this results in a sharp bound for the desired quantity.

Other related work

Avin studied a random distance graph that incorporates both the Erdös–Rényi graph and the random geometric graph (Avin 2008). This graph was shown to have several good properties, for example, the small-diameter property and a high clustering coefficient.

A hybrid model composed of a power law graph and a grid-like local graph has been studied by several groups of researchers (Chung and Lu 2004; Kurucz and Benczúr 2010; Fraigniaud and Giakkoupis 2009).

Clusters or communities based on the concept of conductance were studied in Kannan *et al.* (2004) and Leskovec *et al.* (2008; 2010), where spectral algorithms and other approximation algorithms were used to detect good clusters or communities.

Structure of the paper

We will introduce our definition of the small-community phenomenon in Section 2, where we will also define our models and state the main results of the paper.

In Section 3, we introduce some useful tools and basic facts, as preparation for Sections 4–6, where we show that the models have the desired properties. In Section 7, we discuss the effect of parameter choice on the properties of our proposed models, and, finally, we present some brief conclusions in Section 8.

2. Basic definitions, the model and main results

2.1. *The small-community phenomenon*

In a graph $G = (V, E)$, we use $\deg_G(v)$ to denote the degree of a node $v \in V$. The volume of a subset of $S \subseteq V$ is defined to be the sum of degrees of the vertices in it, that is,

$$\mathrm{vol}(S) = \sum_{v \in S} \deg_G(v).$$

Our definition of communities is inspired by Leskovec *et al.* (2008), which used conductance as a measure of the goodness of a community. In Li and Peng (2011), we introduced the concept of an (α, β, γ)-community based on the conductance and size of a set of nodes. The conductance $\Phi(S)$ of S is the ratio between the number of edges coming out of S and either its volume or the volume of its complement $\bar{S}$, whichever is smaller, that is,

$$\Phi(S) = \frac{|e(S, \bar{S})|}{\min\{\mathrm{vol}(S), \mathrm{vol}(\bar{S})\}},$$

where $e(S, T)$ denotes the set of edges with one endpoint in S and the other in T.

We can now define the (α, β, γ)-community as follows.

Definition 2.1. Given a graph $G = (V, E)$ with $|V| = n$, a connected set $S \subset V$ with $|S| = \omega(1)$ is a strong (α, β)-community if

$$\Phi(S) \leqslant \frac{\alpha}{|S|^{\beta}}. \tag{1}$$

Moreover, if $|S| = O((\ln n)^{\gamma})$, we say that S is a strong (α, β, γ)-community.

Note that in the above definition we require that the size of a community is not too small (that is, $|S| = \omega(1)$). This requirement helps us avoid the trivial case in our definition (when $|S|$ is constant, it can always be treated as a proper community by choosing large α). In fact, a meaningful community in society can never be too small because of a lack of requisite variety or other group function (Allen 2004).

The following definition characterises the property that almost everyone in the network belongs to some small community (from now on, when we write *almost every* we will mean $1 - o_n(1)$, where n is the number of vertices in G).

Definition 2.2. A network (model) G is said to exhibit the small-community phenomenon if almost every node belongs to some (α, β, γ)-community, where $\alpha, \beta, \gamma > 0$ are some global constants.

2.2. *The geometric model*

The base model we will use is a geometric preferential attachment model with additive fitness. Such a model has been studied in van den Esker (2008) and Jordan (2010) – see also Flaxman *et al.* (2007a; 2007b). We assume that a self-loop counts as degree 1. The model is defined on a unit-area spherical surface S (that is, the radius of the sphere is $1/2\sqrt{\pi}$). Let n be the number of vertices we are going to generate. Let $\xi > 0$ be an arbitrary constant and m, r and $\delta = \xi m$ be parameters, which may depend on n (note that this is the essential difference compared with the cases studied in Jordan (2010)). Intuitively speaking, m is the number of edges we are going to add in each step; r is the distance restriction on the two endpoints of an edge; and δ is the additive fitness. Let $B_R(v)$ denote the spherical cap of radius R around v in S, that is,

$$B_R(v) = \{u \in S : \|u - v\| \leqslant R\},$$

where $\|\cdot\|$ denotes the angular distance on S. Let $A_R = \mathrm{area}(B_R(v))$ be the area of the spherical cap of radius R, which is independent of v.

The base model

We start the process from a graph G_1, which is composed of a uniformly generated (from S) node x_1 with $2m$ self-loops. At each time $t+1$ for $t > 0$, if $G_t = (V_t, E_t)$, we first generate a new node x_{t+1} uniformly at random from S, and then connect it to some existing vertices or itself. Specifically, if there is no node in $B_r(x_{t+1})$, we add $2m$ self-loops to x_{t+1}. And if $B_r(x_{t+1}) \cap V_t \neq \emptyset$, we choose m contacts (with replacement) independently from $B_r(x_{t+1})$ for the newcomer such that for any i with $1 \leqslant i \leqslant m$, the probability that some vertex $v \in B_r(x_{t+1})$ is chosen as the ith contact is defined by

$$\Pr\left[y_i^{t+1} = v\right] = \frac{\deg_t(v) + \delta}{\sum_{w \in B_r(x_{t+1}) \cap V_t} (\deg_t(w) + \delta)}. \tag{2}$$

Remark 2.1. A self-loop parameter $\alpha > 2$ was introduced in van den Esker (2008) (and Flaxman *et al.* (2007a; 2007b)) to avoid a technical problem when proving the power law degree distribution. In their settings, a node $v \in B_r(x_{t+1})$ is chosen as the contact with probability

$$\frac{\deg_t(v) + \delta}{\max\left\{\sum_{w \in B_r(x_{t+1}) \cap V_t} (\deg_t(w) + \delta), \alpha(m + \delta/2) A_r t\right\}}, \tag{3}$$

where $\delta > -m$. The case of $\alpha = 0$ was left open in those papers, but Jordan (2010) has since investigated the asymptotic behaviour of the degree sequence for this case. In that study, $m, r, \delta > 0$ are constants that do not depend on n, which converges to infinity. However, in our situation, we need a strong concentration result, so the parameters may depend on n. We will give such a result when $\alpha = 0$ and $\delta > 0$, which strengthens the results in van den Esker (2008) and Jordan (2010), and partially answers the open question in Flaxman *et al.* (2007a; 2007b).

We can show that when $\delta = \xi m > 0$ and $r = r_0 = n^{-\frac{1}{2}} (\ln n)^{c_0}$, where $c_0 = c_0(\xi)$ is large and may depend on ξ, the base model has the power law degree distribution and the small-community phenomenon but does not have the small-diameter property. To incorporate the missing property without changing the other two properties too much, we will introduce some operations that, essentially, generate a uniform recursive tree. We give two different operations such that both of the resulting variants of the base model have the three properties to some extent.

The hybrid model

In this model, every edge has an attribute that indicates whether it is a *local-edge* or a *long-edge*, that is, whether the two endpoints of the edge are local- or long-contacts of each other. A local- (or long-) edge contributes to the local- (or long-) degree of both of its endpoints. We start from G_1^{H}, which is the same as G_1 in the above, and let the self-loops of x_1 be local-edges. At each step $t+1$ for $t \geqslant 1$, we form G_{t+1}^{H} from G_t^{H}, by choosing a new vertex x_{t+1} uniformly at random from S. We first choose for the newcomer m local-contacts $y_i, 1 \leqslant i \leqslant m$, independently at random, as in the base model, with $\deg_t(v)$

in Equation (2) denoting the *local-degree* of v at time t. Then we choose for x_{t+1} another long contact z uniformly from $x_1, \ldots, x_t$.

This model can be seen to be composed of two parts: a local power law graph and a global uniform recursive tree. These can be generated in two phases. First, we generate the local power law graph following the rules used in the base geometric model. We then generate a recursive tree as follows: taking each $t \geqslant 1$ in turn, x_{t+1} connects a long-contact chosen uniformly at random from $x_1, \ldots, x_t$.

The independence of the local and global parts of the hybrid model conforms to our intuition that local and long contacts are formed by different mechanisms. Previous studies on such a model have usually had a global power law graph and a local grid-like graph (see, for example, Chung and Lu (2004)), which is comparable to ours.

The self-loop model

In this model, every new node is created with δ *flexible self-loops*, which may be eliminated in later steps. We now generate x_1 uniformly at random from S and add $2m + \delta$ self-loops to it, with $\delta \geqslant 2$ loops marked flexible. This is the start graph G_1^{S}. At each step $t + 1$ for $t \geqslant 1$, we form G_{t+1}^{S} from G_t^{S}, by choosing a new vertex x_{t+1} uniformly at random from S, and then adding δ flexible self-loops to it. We then choose m contacts $y_i, 1 \leqslant i \leqslant m$ independently at random, as in the base model, with $\deg_t(v)$ in Equation (2) denoting *the number of non-flexible edges* incident to v at time t. Then we choose for x_{t+1} another contact z uniformly from the set of existing nodes containing one or more flexible self-loops (this set cannot be empty because x_t is a member of it), and delete one flexible self-loop from both x_{t+1} and z. The newly added edge (x_{t+1}, z) is marked flexible. Note that the edge-rewiring keeps the degree of the vertices unchanged, which facilitates the analysis of its degree distribution.

This model can also be seen to be composed of two parts: a flexible part and a non-flexible part, which can be generated in several phases. We first generate the non-flexible part following the growth rules of the base model. We then add δ flexible self-loops to each vertex. Then, taking each $t \geqslant 1$ in turn, x_{t+1} connects a contact z chosen uniformly at random from $x_1, \cdots, x_t$, containing flexible self-loop(s), then a flexible self-loop of x_{t+1} and z is deleted and a new flexible edge (x_{t+1}, z) is added.

We can give a plausible explanation of the self-loops emerging in this model. It is a common observation in the social sciences that members of society do not just have direct relationships with others, but some implicitly asymmetric ‘parasocial’ interactions with celebrities, virtual characters, and so on. In these relationships, one of the parties knows a great deal about the other, but the other knows very little if anything about the first party (Horton and Wohl 1956). Such relationships are seldom reflected in commonly used friendship networks, which are mainly concerned with symmetric two-sided friendships. Our model incorporates the parasocial relationships as self-loops, and the edge-rewiring may be roughly interpreted as the long-distance relationship being established at the expense of its parasocial connections.

2.3. *Main results*

Our main results are that the two models have rather good properties. We will assume that $\delta = m\xi$, where $\xi > 0$ is some constant and $r_0 = n^{-1/2}(\ln n)^{c_0}$ for some large constant c_0, which may depend on ξ.

For $r \geqslant r_0$, it is obvious that the diameter of the base model is

$$\Omega(1/r) = \Omega\big(n^{1/2}(\ln n)^{-c_0}\big)$$

(see Section 4), which is large, while the small diameters of the uniform recursive trees imply the small-diameter property in our two generalised models.

Theorem 2.1 (Small-diameter property).

(i) For any $m \geqslant 1, r > 0$, the diameter of G_n^{H} is $O(\ln n)$ with high probability.
(ii) For $m \geqslant K_1(\xi)\ln n$ and $r > 0$, the diameter of G_n^{S} is $O(\ln n)$ with high probability, where $K_1(\xi)$ is some constant depending on ξ.

Given the geometric structure of the models, it is natural to think that a group of vertices close to each other behaves like a good community. We will make this intuition rigorous by considering the *R-neighbourhood* $C_R(v)$ of a vertex v, which is the set of all vertices within a distance of at most R from v in G_n, and then show, for some appropriate r and R, that $C_R(v)$ is a good community for every v. We have the following result.

Theorem 2.2 (Small-community phenomenon). If $r = r_0$ and $m \geqslant K_2(\xi)\ln n$, where $K_2(\xi)$ is some constant depending on ξ, both G_n^{H} and G_n^{S} have the small-community phenomenon, that is, in each model, for every node $v \in V_n$, there exists with high probability some (α, β, γ)-community containing v, where α, β, γ are constants independent of n.

A simple corollary of the above theorem is that the base model G_n also displays the small-community phenomenon, which indicates that the community structure is mainly determined by the geometric structure of our model and that the effect of long edges is small because every new node can establish $m \gg \ln n$ local edges but only 1 long edge.

The power law degree distribution is a result of the preferential attachment scheme used in our base model, for which we have the following theorem.

Theorem 2.3 (Degree distribution of the base model). In the base model, if $r \geqslant r_0$, $m = O(\ln^2 n)$ and $\delta = m\xi$ for any constant $\xi > 0$, there exist constants C_k and μ, such that for all $k = k(n) \geqslant m$,

$$\mathrm{E}[d_k(t)] = C_k \frac{n}{k^{3+\xi}} + O\left(\frac{n}{(nr^2)^{\mu}}\right), \tag{4}$$

where:

— $d_k(t)$ denotes the number of vertices with degree k in the base model G_t;
— $C_k = C_k(m, \xi)$ tends to a limit $C_\infty(m, \delta)$ that only depends on m and δ as $k \to \infty$; and
— μ is some constant depending on ξ and strictly less than 1.

Theorem 2.3 already significantly strengthens the results in both van den Esker (2008) and Jordan (2010). The proof of this theorem requires a new technique of recursively bounding the concentration inequalities, which we will develop in Section 6.

We can show from Theorem 2.3 that in our generalised models, the networks satisfy a nice power law degree distribution.

Theorem 2.4 (Power law degree distribution). For $r \geqslant r_0$ and $m = O(\ln^2 n)$, the expected degree sequences of **the local graph** of the hybrid model G_n^H and **the whole graph** of the self-loop model G_n^S both follow a power law distribution with exponent $3 + \xi$. More specifically, there exist constants C_k^{H}, C_k^{S} and μ such that for all $k = k(n) \geqslant m$:

(i) In the hybrid model,

$$\mathrm{E}[d_k(n)] = C_k^{\mathrm{H}} \frac{n}{k^{3+\xi}} + O\left(\frac{n}{(nr^2)^{\mu}}\right),$$

where $d_k(t)$ denotes the number of vertices with local-degree k in G_t^{H}.

(ii) In the self-loop model,

$$\mathrm{E}[d_k(n)] = C_k^{\mathrm{S}} \frac{n}{k^{3+\xi}} + O\left(\frac{n}{(nr^2)^{\mu}}\right),$$

where $d_k(t)$ denotes the number of vertices with total degree k in G_t^{S}.

In the above statements, both C_k^{H} and C_k^{S} tend to some limits that depend only on m and δ as $k \to \infty$, and μ is some constant depending on ξ and strictly less than 1.

From the above theorems, we know that when $r = r_0 = n^{-1/2}(\ln n)^{c_0}$, the two generalised models have all three properties simultaneously to some extent (the reservation being that only the local part has the power law degree distribution in the hybrid model). One might then ask about the cases when r is too large or too small, and we will give some evidence that at least one of the three properties disappears in such cases. In particular, we have the following new phenomenon when r is large.

Theorem 2.5 (Large community and small expander). In the base model G_n, let $r = n^{-1/2+\epsilon}$, where $\epsilon > 0$ and $m \geqslant K \ln n$, for some sufficiently large constant K. Then:

(i) If $R = n^{-1/2+\rho}$, for any $\rho > \epsilon$, then

$$|C_R(v)| = \Theta(n^{2\rho})$$
$$\Phi(C_R(v)) = O\left(\frac{1}{n^{\rho-\epsilon}}\right)$$

with high probability.

(ii) For all $R = o(r)$,

$$\Phi(C_R(v)) = \Omega(1)$$

with high probability.

Theorem 2.5 says that when $r = n^{-1/2+\epsilon}$, there exists some large community for every node, but it may not belong to any small community because the most natural candidate, that is, the small neighbourhood, is not a good community. Note that Theorem 2.5 may imply a new phenomenon in networks. It would be interesting to find some real-world networks in which there is a large fraction of nodes, each of which is contained in both a good but large community, and a small expander. Also note that the two generalised models exhibit the same phenomenon for large r.

The remaining sections of the paper are devoted to proving our main results, that is, Theorems 2.1–2.5. We introduce some basic tools for our proof in Section 3, and prove some basic properties of our network models; in Sections 4 and 5, we prove Theorems 2.1 and 2.2, respectively; in Section 6, we prove Theorems 2.3 and 2.4; and in Section 7, we prove Theorem 2.5. Finally, in Section 8, we discuss some further issues following on from the results in this paper.

3. Useful tools and basic facts

Before proving the main results, we will first give some basic facts, which will be useful in our proofs of the main results.

We will use the following form of the Chernoff bound – see, for example, Dubhashi and Panconesi (2009, Theorem 1.1).

Lemma 3.1. If $X_1, \ldots, X_t$ are independently distributed in $[0, 1]$ and $X = \sum_{i=1}^{t} X_i$, then for $0 < \zeta \leqslant 1$,

$$\Pr\left[|X - \mathrm{E}[X]| \geqslant \zeta \mathrm{E}[X]\right] \leqslant 2e^{-\frac{\zeta^2 \mathrm{E}[X]}{3}}. \tag{5}$$

We will use the following submartingale concentration inequality extensively in our proofs – see Chung and Lu (2006, Theorems 2.38 and 2.41).

Lemma 3.2. Suppose $\{X_0, \ldots, X_t\}$ is a sequence of random variables associated with a filter $\{\mathscr{F}_0, \ldots, \mathscr{F}_t\}$ and $\mathscr{G}$ is some event on the probability space. If for $1 \leqslant i \leqslant t$ we have

$$\mathrm{E}[X_i|\mathscr{F}_{i-1}, \mathscr{G}] \leqslant X_{i-1}$$
$$\mathrm{Var}[X_i|\mathscr{F}_{i-1}, \mathscr{G}] \leqslant \sigma_i^2$$
$$X_i - \mathrm{E}[X_i|\mathscr{F}_{i-1}, \mathscr{G}] \leqslant M$$

where σ_i^2, M are non-negative constants, then

$$\Pr[X_t \geqslant X_0 + \lambda] \leqslant e^{-\frac{\lambda^2}{2\sum_{i=1}^{t} \sigma_i^2 + M\lambda/3}} + \Pr[\neg\mathscr{G}]. \tag{6}$$

The supermartingale concentration inequality is similar, so we will not give it explicitly here.

In the following sections we will use constants c_0, c_1 and c_2, which may depend on ξ, to characterise some bounds. We state here the conditions that these three constants should satisfy:

$$(c_0 - c_1 - 1)(1 - 1/(\xi + 2)) < c_1 < 2(c_0 - c_1 - 1)(1 - 2/(2 + \xi)) \tag{7}$$

$$c_2 = c_1 \frac{\ln(\xi(1 + \xi/2) + 1)}{\ln((7 + 400/\xi)^2(\xi(1 + \xi/2) + 1))}. \tag{8}$$

https://doi.org/10.1017/S0960129511000570 Published online by Cambridge University Press

Note that for fixed ξ we can always choose c_0 to be large enough to guarantee that c_2 is also large, which will ensure that the bounds we obtain in the proof are good.

In the definition of our base model, a new vertex will create $2m$ self-loops if there is no existing vertex within a distance of at most r. This rule is made to guarantee that at each step the degree of the graph grows by $2m$, which facilitates further analysis. Moreover, in most interesting cases, when $r = r_0 = n^{-1/2}(\ln n)^{c_0}$, if t grows as large as

$$\tau = O\left(\frac{n}{(\ln n)^{2c_0-1}}\right),$$

then with high probability for any vertex that comes after time τ, there will be many existing nodes within a distance of at most r. Therefore, we will focus on the processes through which all the *later* comers will choose existing nodes as neighbours rather than creating $2m$ self-loops.

In analysing the degree sequence of our base model, it is convenient to compare the chosen probability given in Equation (2) with the traditional case (see, for example, Bollobás (2003)), in which at each step $t+1$ an existing vertex v with degree k is chosen with probability $k/2t$, where $2t$ is the total degree of all existing vertices. Thus, it is natural to consider using a good estimate of Equation (2) for further analysis. In particular, we would like to have some good bound on the normalised quantity of the denominator in Equation (2). Let $T_t(u)$ denote this quantity, namely,

$$T_t(u) = \sum_{v \in B_r(u) \cap V_t} (\deg_t(v) + \delta).$$

A closely related quantity is

$$Z_t(u) = \sum_{v \in B_r(u) \cap V_t} 1,$$

which is the number of vertices in $B_r(u)$ at time t. We will now prove several simple facts related to these two quantities.

Lemma 3.3. If $u \in S$ and $t > 0$, then the expectation of $T_t(u)$ is $A_r(2m+\delta)t$.

Proof. Note that

$$\begin{aligned} \mathrm{E}[T_t(u)] &= \mathrm{E}\left[\sum_{v \in B_r(u) \cap V_t} (\deg_t(v) + \delta)\right] \\ &= \mathrm{E}\left[\sum_{v \in V_t} (\deg_t(v) + \delta) 1_{v \in B_r(u)}\right] \\ &= \mathrm{E}\left[\sum_{v \in V_t} \deg_t(v) 1_{v \in B_r(u)}\right] + \delta A_r t. \end{aligned} \tag{9}$$

But from Flaxman *et al.* (2007a, Lemmas 1 and 2), we have

$$\mathrm{E}\left[\sum_{v \in V_t} \deg_t(v) 1_{v \in B_r(u)}\right] = 2A_r m t,$$

which completes the proof. □

https://doi.org/10.1017/S0960129511000570 Published online by Cambridge University Press

Let A_r denote the area of $B_r(v)$. Then $A_r = \text{area}(B_r(v)) \sim r^2/4$, for $r = o(1)$. Let

$$t_r = \frac{12(\ln n)^2 n^{c_1/c_0}}{r^{2(1-c_1/c_0)}}.$$

So $A_r t_r \sim 3(\ln n)^2(nr^2)^{c_1/c_0}$. We will consider $r \geqslant r_0 = n^{-1/2}(\ln n)^{c_0}$ and let

$$t_0 := t_{r_0} = \frac{12n}{(\ln n)^{2c_0-2c_1-2}}.$$

We first give an estimate of the quantity $Z_t(u)$.

Lemma 3.4. If $r \geqslant r_0$, then for any $t \geqslant t_r$, we have

$$|Z_t(u) - A_r t| \leqslant \frac{1}{(nr^2)^{c_1/2c_0}} A_r t,$$

with probability at least $1 - 2n^{-\ln n}$.

Proof. Noting that

$$Z_t(u) = \sum_{i=1}^{t} 1_{x_i \in B_r(u)}$$

and

$$\Pr[1_{x_i \in B_r(u)} = 1] = A_r,$$

we can obtain the result by simply applying the Chernoff bound. □

From the above lemma, we can give a rough bound on $T_t(u)$.

Lemma 3.5. If $r \geqslant r_0$, then for any $t \geqslant t_r$, we have

$$\left(1 - \frac{1}{(nr^2)^{c_1/2c_0}}\right)(1+\xi)mA_r t \leqslant T_t(u) \leqslant 4\left(1 + \frac{1}{(nr^2)^{c_1/2c_0}}\right)(2+\xi)mA_r \tag{10}$$

with probability at least $1 - 4n^{-\ln n}$.

Proof. The first inequality is obvious from the trivial relation that $T_t(u) \geqslant m(1+\xi)Z_t$ and the bound on Z_t given in Lemma 3.4.

To see the second inequality, note that the sum of the degrees of vertices in $B_r(u)$ is equal to the sum of the out-degrees of all vertices in $B_r(u)$, which is equal to mZ_t plus the sum of the in-degrees of vertices in $B_r(u)$, which is at most the sum of the out-degrees of

https://doi.org/10.1017/S0960129511000570 Published online by Cambridge University Press

all vertices in $B_{2r}(u)$. Therefore,

$$\begin{aligned} T_t(u) &\leqslant (m+\delta)Z_t + m \sum_{v \in V_t \cap B_{2r}(u)} 1 \\ &\leqslant (2m+\delta) \sum_{v \in V_t \cap B_{2r}(u)} 1 \\ &\leqslant (2m+\delta)A_{2r}t\left(1+\frac{1}{(nr^2)^{c_1/2c_0}}\right) \\ &= 4(2+\xi)mA_rt\left(1+\frac{1}{(nr^2)^{c_1/2c_0}}\right) \end{aligned}$$

with probability $1-2n^{-\ln n}$. □

4. Small-diameter property

It is obvious that the diameter in the base model is at least $\Omega(1/r) = \Omega(n^{1/2}(\ln n)^{-c_0})$ for all $r \geqslant r_0$ since any vertex can connect nodes within a distance of at most r, and the maximum distance of two vertices is $\Omega(1)$. However, with the addition of the ability to choose uniformly from the subset of previous vertices, the diameter can be reduced to $O(\ln n)$, with high probability. We will use the following classic result on the diameter and maximum degree of a uniform recursive tree.

Lemma 4.1. With high probability, the diameter and maximum degree in a uniform recursive tree are $\Theta(\ln n)$.

Proof. This is a classic result – see, for example, Pittel (1994) and Devroye and Lu (1995) for a proof. □

We can now bound the diameter of the two generalised models as follows.

Proof of Theorem 2.1 (Small-diameter property). We consider the two models separately.

The hybrid model: In this case, no matter how quickly the local graph grows, the global graph is the same as the uniform recursive tree, which gives an upper bound $O(\ln n)$ on the diameter of the whole graph.

Self-loop model: In this case, the constructed tree in the flexible part is restricted to having degree at most δ, and thus may be different from a uniform recursive tree. However, by Lemma 4.1, the maximum degree of a uniform recursive tree is $L \ln n$, where L is the hidden constant in $\Theta(\ln n)$, from which we know that if $\delta \geqslant L \ln n$, then, with high probability, the constructed tree in the flexible part is the same as the uniform recursive tree. Therefore, the diameter of the self-loop model is again upper bounded by $O(\ln n)$. Finally, we note that $\delta = m\xi \geqslant L \ln n$ is equivalent to $m \geqslant L \ln n/\xi$, which completes the proof. □

5. The small-community phenomenon

In this section, we consider the community structure and will require that $r = r_0$. We start from the intuition that a group of people close to each other form a good community,

https://doi.org/10.1017/S0960129511000570 Published online by Cambridge University Press

which can be thought of as a geographical community (Kurucz and Benczúr 2010). In particular, for a node v, we define the R-neighbourhood $C_R(v)$ of v to be the set of vertices within a distance of at most R from v in G_n, that is, $C_R(v) = B_R(v) \cap V_n$. Letting $R_0 = n^{-1/2}(\ln n)^{2c_0}$, we will show that $C_{R_0}(v)$ is a good community. In this section, we will assume that $m \geqslant K_2(\xi)\ln n$, where $K_2(\xi)$ is some large constant depending on ξ.

Note that given v, the probability that a node generated uniformly at random from S will land in $B_{R_0}(v)$ is

$$A_{R_0} \sim R_0^2/4 = \frac{(\ln n)^{4c_0}}{4n}.$$

Using the Chernoff bound, it is easy to show that with high probability the number of nodes in C_{R_0} is $\Theta((\ln n)^{4c_0})$, which means that the size of such an R-neighbourhood is small. Now we consider the connectivity of the subgraph induced by $C_{R_0}(v)$.

Lemma 5.1. In the base model, if $r = r_0 = n^{-1/2}(\ln n)^{c_0}$, then for any $v \in V_n$, the R_0-neighbourhood $C_{R_0}(v)$ induces a connected subgraph in G_n with high probability.

Proof. We will first show that for every v, we have $C_{r/2}(v)$ induces a connected subgraph in G_n with high probability. The lemma then follows from the fact that any two vertices u, u' in $C_{R_0}(v)$ can be connected by a set of paths between vertices $u = v_1, v_2, \cdots, v_k = u'$ such that each vertex pair (v_i, v_{i+1}) is within a distance of $r/2$.

Now we consider the connectivity of $C_{r/2}(v)$.

Let $A_r T = 12 \ln n$. Then

$$T = \frac{12n}{(\ln n)^{2c_0-1}}.$$

Let H_0 be the subgraph induced by nodes within a distance of at most $r/2$ from v at time T. Now let $x_{t_1}, \cdots, x_{t_k}$ be the nodes that land in $B_{r/2}(v)$ after time T, and H_s be the corresponding subgraph when vertex x_{t_s} is added in $B_{r/2}(v)$. Since every vertex x_j will land in $B_{r/2}(v)$ with probability $A_{r/2}$, we know that with high probability, for $t \geqslant T$, the number of nodes in $B_{r/2}(v)$ will be in the range $[\kappa_1 A_{r/2} t, \kappa_2 A_{r/2} t]$ for some constants κ_1, κ_2. In particular, we have that

$$|H_0| \leqslant \kappa_2 A_{r/2} T = 3\kappa_2 \ln n$$

and

$$\kappa_1 A_{r/2} t_s \leqslant |H_s| \leqslant \kappa_2 A_{r/2} t_s.$$

Now let X_s be the number of connected components of H_s, and Y_s be the number of connected components of H_s connected to $x_{t_{s+1}}$. Then we have

$$X_{s+1} = X_s - Y_s + 1, X_0 \leqslant 3\kappa_2 \ln n.$$

We show that if $s \leqslant 6\kappa_2 \ln n$, then X_s decreases by at least 1 for every $s \geqslant 1$ with probability at least 7/10, from which we know that the probability that $H_{6\kappa_2 \ln n}$ is not connected is bounded by $O(n^{-3})$. The Lemma then follows from the fact that each of the later vertices $x_{t_{s+1}}$ such that $s \geqslant 6\kappa_2 \ln n$ will connect the H_s with probability at least $1 - O(n^{-10})$.

https://doi.org/10.1017/S0960129511000570 Published online by Cambridge University Press

Let $\mathscr{E}$ denote the event that for any $u \in V_n$ and for each $t \geqslant T$, we have

$$T_t(u) \leqslant 32(2+\xi)mA_{r/2}t.$$

Then, as in the proof of Lemma 3.5, the probability that $\mathscr{E}$ holds is $1 - O(n^{-4})$. Now, conditioned on $\mathscr{E}$, for each $1 \leqslant s \leqslant 6\kappa_2 \ln n$, since x_{t_s} is in $B_{r/2}(v)$, we have that $|x_{t_s} - u| \leqslant r$ for every vertex $u \in H_{s-1}$, so x_{t_s} will connect u with probability at least

$$\frac{m+\delta}{T_{t_s-1}(x_{t_s})} \geqslant \frac{1}{32A_{r/2}t_s}.$$

Therefore, the probability that x_{t_s} will not connect any vertex in $B_{r/2}(v)$ is

$$\Pr[Y_s = 0] \leqslant \left(1 - \frac{|H_s|}{32A_{r/2}t_s}\right)^m \leqslant n^{-10},$$

where the second inequality follows from the fact that $m \geqslant K_2(\xi)\ln n$.

Now we consider the case where H_s has at least two connected components, namely, $X_s \geqslant 2$. The probability that $x_{t_{s+1}}$ will connect at most one component is

$$\Pr[Y_s = 1 | X_s \geqslant 2] \leqslant 2\left(1 - \frac{1}{32A_{r/2}t_s}\right)^m \leqslant 1/10,$$

where we have used the fact that $32A_{r/2}t_s \geqslant 96\ln n$ and $m \geqslant K_2(\xi)\ln n$.

Therefore, X_s decreases by at least 1 for every $1 \leqslant s \leqslant 6\kappa_2 \ln n$ with probability at least 7/10, which completes our proof. □

We will now show that the conductance of $C_{R_0}(v)$ in each model is small.

Lemma 5.2. In both the hybrid and self-loop models, for any $v \in V_n$, we have

$$\Phi(C_{R_0}(v)) = O\left(\frac{1}{|C_{R_0}(v)|^{1/4c_0}}\right) \tag{11}$$

with high probability.

Proof. We first consider the hybrid model. For convenience, we abbreviate $C_{R_0}(v)$ as C. Let $e(C, \bar{C})$ denote the set of edges connecting C and its complement. Let $e_1(C, \bar{C})$ and $e_2(C, \bar{C})$ denote edges in $e(C, \bar{C})$ that are local and long, respectively. Then we have

$$e(C, \bar{C}) = e_1(C, \bar{C}) \cup e_2(C, \bar{C}).$$

Local edges connecting C and $\bar{C}$ must lie between the two spherical segments separated by the boundary of $C_{R_0}(v)$. More specifically, if $e = (u, w) \in e_1(C, \bar{C})$, then one of u, w lies on the strip $str_1 = B_{R_0+r}(v) \backslash B_{R_0}(v)$, and the other point lies on the strip $str_2 = B_{R_0-r}(v) \backslash B_{R_0}(v)$. With high probability, the total number of vertices in str_1 is at most $n(2rR_0 + r^2)$, and the total number of vertices in str_2 is at most $n(2rR_0 - r^2)$. Hence, the number of local edges lying between the two strips is at most $4mnrR_0$, namely, $|e_1(C, \bar{C})| \leqslant 4mnrR_0$.

Now we consider the long edges that connect C to the rest of the graph. We will show that the number of these edges is relatively small compared with the number of local edges within C. More precisely, we have the following lemma.

Lemma 5.3. In the hybrid model, let Y_t denote the sum of the long-degrees of vertices in $B_{R_0}(v) \cap V_t$. Then $Y_n \leqslant cA_{R_0}n$ for some constant c, with high probability.

Proof. By definition, we have the following recurrence for Y_t.

$$\mathrm{E}[Y_{t+1}|Y_t] = Y_t + A_{R_0} + \frac{|B_{R_0}(v) \cap V_t|}{t}. \tag{12}$$

Let $A_{R_0}T = 12 \ln n$, so

$$T = \frac{12n}{(\ln n)^{4c_0-1}}.$$

Let $\mathscr{F}$ denote the event that for all $t \geqslant T$, the relation $|B_{R_0}(v) \cap V_t| \in [\kappa_1 A_{R_0} t, \kappa_2 A_{R_0} t]$ holds for some constants κ_1 and κ_2 and that the maximum long-degree of vertices $x_1, \cdots, x_T$ is $L \ln n$. By Lemma 4.1 and the Chernoff bound, we know that $\Pr[\mathscr{F}] \geqslant 1 - O(n^{-3})$.

Now we know that for $t \geqslant T$,

$$\mathrm{E}[Y_{t+1}|Y_t, \mathscr{F}] \leqslant Y_t + A_{R_0} + \frac{\kappa_2 A_{R_0} t}{t},$$

from which we have

$$\mathrm{E}[Y_{t+1}|Y_t, \mathscr{F}] - (1+\kappa_2)A_{R_0}(t+1) \leqslant Y_t - (1+\kappa_2)A_{R_0}t. \tag{13}$$

Conditioned on $\mathscr{F}$, we know that the number of vertices in $B_{R_0}(v) \cap V_t$ is

$$\kappa_2 A_{R_0} T \leqslant 12\kappa_2 \ln n,$$

and every vertex in this set has degree at most $L \ln n$, from which we know that $Y_T \leqslant 12\kappa_2 L(\ln n)^2$. We now define

$$X_\tau = \begin{cases} Y_\tau - (1+\kappa_2)A_{R_0}\tau & \text{for } \tau \geqslant T+1 \\ 12\kappa_2 L(\ln n)^2 & \text{for } \tau = T. \end{cases}$$

By inequality (13), $X_T, \cdots, X_t$ forms a submartingale with error $O(n^{-3})$. We also have that for $\tau > T$,

$$X_\tau - \mathrm{E}[X_\tau | X_{\tau-1}] \leqslant 1,$$

and

$$\begin{aligned} \mathrm{Var}[X_\tau|X_{\tau-1}] &= \mathrm{Var}[Y_\tau|X_{\tau-1}] \\ &\leqslant \mathrm{E}[(Y_\tau - Y_{\tau-1})^2|X_{\tau-1}] \\ &\leqslant (1+\kappa_2)A_{R_0}. \end{aligned}$$

We now apply the submartingale concentration inequality, as in Lemma 3.2, to get

$$\begin{aligned} \Pr[X_t \geqslant X_T + \lambda] &\leqslant e^{-\frac{\lambda^2}{2(\sum_{\tau=T+1}^{t}(1+C_2)A_{R_0} + \lambda/3)}} + O(n^{-3}) \\ &\leqslant e^{-\frac{\lambda^2}{2t(1+C_2)A_{R_0} + 2\lambda/3}} + O(n^{-3}). \end{aligned}$$

Let $\lambda = c'\sqrt{\ln n A_{R_0} t}$ for some constant c'. Then

$$\Pr\left[X_t \geqslant X_T + c'\sqrt{\ln n A_{R_0} t}\right] \leqslant O(n^{-3}).$$

https://doi.org/10.1017/S0960129511000570 Published online by Cambridge University Press

Finally, using $X_t = Y_t - (1+\kappa_2)A_{R_0}t$, we have

$$\Pr\left[Y_t \geqslant (1+\kappa_2)A_{R_0}t + c'\sqrt{\ln n A_{R_0} t} + 12\kappa_2 L(\ln n)^2\right] \leqslant O(n^{-3}).$$

In particular, $Y_n \leqslant cA_{R_0}n$ with high probability for some constant c, which completes the proof of Lemma 5.3. □

Continuing the proof of Lemma 5.2, we know by Lemma 5.3 that $|e_2(C,\bar{C})| \leqslant cA_{R_0}n$, so the total number of edges between C and $\bar{C}$ is

$$|e(C,\bar{C})| = O(mrR_0n + R_0^2 n). \tag{14}$$

The volume of C is at least $m|C| \sim mR_0^2 n$, which means that

$$\Phi(C) = O\left(\frac{m4rR_0n + R_0^2 n}{mR_0^2 n}\right) = O((\ln n)^{-1}) = O\left(\frac{1}{|C|^{1/(4c_0)}}\right). \tag{15}$$

Finally, we briefly discuss the proof for the self-loop model. Let $\delta \geqslant K_1(\xi)\ln n$. Then, with high probability, the constructed tree in the flexible part of the model is a uniform recursive tree in the same way as in the proof of Theorem 2.1. Therefore, the edges that connect an R-neighbourhood and its complement can also be bounded by the same argument as in the case of the hybrid model, which then gives the same result as Equation (15).

This completes the proof of Lemma 5.2. □

We can now show that the two models have the small-community phenomenon.

Proof of Theorem 2.2 (small-community phenomenon). For each $v \in V_n$, the R_0-neighbourhood $C_{R_0}(v)$ is of size $\Theta((\ln n)^{4c_0})$. By Lemmas 7 and 8, we know that $C_{R_0}(v)$ is an (α,β,γ)-community of v, where α is the hidden constant in the term

$$O\left(\frac{1}{|C_{R_0}(v)|^{1/4c_0}}\right)$$

in Equation (15), $\beta = 1/4c_0$ and $\gamma = 4c_0$. This completes the proof of Theorem 2.2. □

Note that the proof of Theorem 2.2 also implies that the base model G_t has the small-community phenomenon. In fact, in that case, we do not need to consider the effect of the edges generated in the uniform recursive tree, which simplifies the analysis. We can easily show that the R_0-neighbourhood $C_{R_0}(v)$ has small size, induces a connected subgraph and has conductance

$$\begin{aligned}\Phi(C_{R_0}(v)) &= O\left(\frac{m4rR_0n}{mR_0^2n}\right)\\ &= O((\ln n)^{-c_0})\\ &= O\left(\frac{1}{|C_{R_0}(v)|^{1/4}}\right)\\ &\leqslant \frac{\alpha'}{|C_{R_0}(v)|^{1/4}},\end{aligned}$$

that is, every node in the base model is contained in an $(\alpha', 1/4, 4c_0)$-community.

6. The power law degree distribution

In this section, we prove Theorems 2.3 and 2.4. In particular, in Section 6.1, we prove Theorem 2.3 by assuming a concentration inequality of the degree sequence, which we prove in Section 6.2 by developing an *alternating concentration method*, and, finally, we prove Theorem 2.4 in Section 6.3.

6.1. *The degree sequence on the base model*

To prove Theorem 2.3, we analyse a recurrence on $\mathrm{E}[d_k(t)]$ in the usual way. Recall that

$$T_t(u) = \sum_{v \in B_r(u) \cap V_t} (\deg_t(v) + \delta).$$

As mentioned earlier, we will first give a good estimate of $T_t(u)$ and show that $T_t(u)$ concentrates around its expected value. Building on this, we can derive the degree sequence from the recurrence on $\mathrm{E}[d_k(t)]$.

Recall that

$$t_r = \frac{12(\ln n)^2 n^{c_1/c_0}}{r^{2(1-c_1/c_0)}}$$

for any $r \geqslant r_0$. We have the following concentration inequality for $T_t(u)$.

Lemma 6.1 (Alternating Concentration Theorem). If $r \geqslant r_0$, then for all $t \geqslant t_r$, we have

$$\Pr\left[|T_t(u) - (2+\xi)mA_r t| \geqslant \frac{1}{(nr^2)^{c_2/2c_0}} mA_r t\right] = O(n^{-2}), \tag{16}$$

where c_1 and c_2 are constants satisfying the conditions in Equations (7) and (8).

Lemma 6.1 is one of the key technical contributions made in this paper, and is interesting in its own right. To prove it, we will need to develop an *alternating concentration method*, through which we will alternately and iteratively apply the submartingale and supermartingale inequalities to prove the desired concentration result.

The role of Lemma 6.1 is to give a good estimate of

$$\mathrm{E}\left[\frac{1_{|x_{t+1}-v|\leqslant r}}{T_t(x_{t+1})} | G_t\right]$$

to analyse the recurrence of $\mathrm{E}[d_k(t)]$. In this section, we will use Lemma 6.1 to prove Theorem 2.3; the full proof of Lemma 6.1 is given in Section 6.2.

Proof of Theorem 2.3. We define

$$D_k(t) := \{v \in V(G_t) | \deg_{G_t}(v) = k\}.$$

Then $d_k(t) = |D_k(t)|$.

https://doi.org/10.1017/S0960129511000570 Published online by Cambridge University Press

The recurrence for the expectation of $d_k(t)$ can be written as follows.

$$\begin{aligned} \mathrm{E}[d_k(t+1)|G_t] = d_k(t) & \\ & + \sum_{v \in D_{k-1}(t)} \left(m\mathrm{E}\left[\frac{(k-1+\delta)1_{|x_{t+1}-v|\leqslant r}}{T_t(x_{t+1})} |G_t \right] \right) \\ & - \sum_{v \in D_k(t)} \left(m\mathrm{E}\left[\frac{(k+\delta)1_{|x_{t+1}-v|\leqslant r}}{T_t(x_{t+1})} |G_t \right] \right) \\ & + O(m\mathrm{E}[\eta_k(G_t, x_{t+1})|G_t]), \end{aligned} \tag{17}$$

where $\eta_k(G_t, x_{t+1})$ denotes the probability that a parallel edge from the new vertex x_{t+1} to a vertex of degree no more than k is created, and is at most

$$\binom{m}{2} \sum_{i=m}^{k} \sum_{v \in D_i(t)} (i+\delta)^2 \left(\frac{1_{|v-x_{t+1}|\leqslant r}}{T_t(x_{t+1})} \right)^2 .$$

Now for $t \geqslant t_r$, we use $\mathscr{A}_t$ to denote the event

$$|T_t(u) - (2+\xi)mA_r t| \leqslant \frac{1}{(nr^2)^{c_2/2c_0}} mA_r t.$$

By Lemma 6.1, we have

$$\Pr[\mathscr{A}_t] = 1 - O(n^{-2}).$$

Therefore, for $t \geqslant t_r$,

$$\begin{aligned} & \mathrm{E}\left[\sum_{v \in D_k(t)} \frac{(k+\delta)1_{|x_{t+1}-v|\leqslant r}}{T_t(x_{t+1})} \right] \\ & = \mathrm{E}\left[\sum_{v \in D_k(t)} \frac{(k+\delta)1_{|x_{t+1}-v|\leqslant r}}{(2+\xi)mA_r t} \left(1 + O\left(\frac{1}{(nr^2)^{c_2/2c_0}}\right)\right) |\mathscr{A}_t \right] \Pr[\mathscr{A}_t] + O(n^{-2}) \\ & = \frac{(k+\delta)}{(2+\xi)mt} \left(1 + O\left(\frac{1}{(nr^2)^{c_2/2c_0}}\right)\right) \mathrm{E}[d_k(t)|\mathscr{A}] \Pr[\mathscr{A}] + O(n^{-2}) \\ & = \frac{(k+\delta)}{(2+\xi)mt} \left(1 + O\left(\frac{1}{(nr^2)^{c_2/2c_0}}\right)\right) (\mathrm{E}[d_k(t)] - \mathrm{E}[d_k(t)|\neg\mathscr{A}] \Pr[\neg\mathscr{A}]) + O(n^{-2}) \\ & = \frac{(k+\delta)\mathrm{E}[d_k(t)]}{(2+\xi)mt} + O\left(\frac{1}{(nr^2)^{c_2/2c_0}}\right). \end{aligned}$$

Similarly, we have

$$\mathrm{E}\left[\sum_{v \in D_{k-1}(t)} \frac{(k-1+\delta)1_{|x_{t+1}-v|\leqslant r}}{T_t(x_{t+1})} \right] = \frac{(k-1+\delta)\mathrm{E}[d_{k-1}(t)]}{(2+\xi)mt} + O\left(\frac{1}{(nr^2)^{c_2/2c_0}}\right).$$

The error term can be bounded as follows.

$$\begin{aligned}
&\mathrm{E}[\eta_k(G_t, x_{t+1})] \\
&\quad \leqslant \binom{m}{2} \mathrm{E}\left[\sum_{i=m}^{k} \sum_{v \in D_i(t)} (k+\delta)^2 \left(\frac{1_{|v-x_{t+1}| \leqslant r}}{T_t(x_{t+1})}\right)^2\right] \\
&\quad \leqslant \binom{m}{2} \mathrm{E}\left[\sum_{i=m}^{k} \sum_{v \in D_i(t)} (k+\delta)^2 \frac{1}{m^2 A_r t^2}\left(1 + O\left(\frac{1}{(nr^2)^{c_2/2c_0}}\right)\right)\right] + O(n^{-2}) \\
&\quad \leqslant O\left(\frac{(k+\delta)^2}{A_r t}\right) + O(n^{-2}).
\end{aligned}$$

If

$$k + \delta \leqslant k_0(t) = (nr^2)^{c_1/2c_0 - c_2/4c_0},$$

then

$$\mathrm{E}[\eta_k(G_t, x_{t+1})] = O\left(\frac{1}{(\ln n)^2 (nr^2)^{c_2/2c_0}}\right)$$

and

$$\mathrm{E}[m\eta_k(G_t, x_{t+1})] = O\left(\frac{1}{(nr^2)^{c_2/2c_0}}\right),$$

given the fact that $m = O(\ln^2 n)$.

Let $\bar{d}_k(t) := \mathrm{E}[d_k(t)]$. Now the recurrence can be simplified as

$$\begin{aligned}
\bar{d}_k(t+1) = \bar{d}_k(t) - \frac{(k+\delta)\bar{d}_k(t)}{(2+\xi)t} + \frac{(k-1+\delta)\bar{d}_{k-1}(t)}{(2+\xi)t} \\
+ 1_{k=m} + O\left(\frac{1}{(nr^2)^{c_2/2c_0}}\right).
\end{aligned} \tag{18}$$

We now define a new recurrence related to (18). For $j < m$, let $f_j = 0$, and for $j \geqslant m$, let

$$f_k = \frac{k-1+\delta}{2+\xi} f_{k-1} - \frac{k+\delta}{2+\xi} f_k + 1_{d=m}, \tag{19}$$

which has solution

$$f_m = \frac{2+\xi}{2+\xi+m+\delta},$$

and for $k \geqslant m+1$,

$$\begin{aligned}
f_k &= \prod_{j=m+1}^{k} \frac{j-1+\delta}{2+\xi+j+\delta} f_m \\
&= \frac{\Gamma(k+\delta)\Gamma(m+4+\xi+\delta)}{\Gamma(3+\xi+k+\delta)\Gamma(m+1+\delta)} \frac{2+\xi}{2+\xi+m+\delta} \\
&= \frac{\phi_k(m,\delta)}{k^{3+\xi}},
\end{aligned}$$

where $\phi_k(m, \delta)$ tends to a limit $\phi_\infty(m, \delta)$ that only depends on m and δ as $k \to \infty$.

https://doi.org/10.1017/S0960129511000570 Published online by Cambridge University Press

We will now show that

$$|\bar{d}_k(t) - f_k t| \leqslant M\left(t_r + \frac{n + Lt}{(nr^2)^{c_2/2c_0}}\right), \tag{20}$$

where M is some large constant and L is the hidden constant in the term

$$O\left(\frac{1}{(nr^2)^{c_2/2c_0}}\right)$$

in Equation (18).

We now prove (20) by induction:

— $t \leqslant t_r$:
In this case the relation holds trivially.

— $t \geqslant t_r$ and $k \geqslant k_0(t)$:
In this case the inequality follows from the fact that $\bar{d}_k(t) \leqslant 2mt/k$.

— $t \geqslant t_r$ and $k \leqslant k_0(t)$:
We have

$$\begin{aligned}
&|\bar{d}(k+1) - f_k(t+1)| \\
&\quad = \left|\bar{d}_k(t) - f_k(t+1) - \frac{(k+\delta)\bar{d}_k(t)}{(2+\xi)t} + \frac{(k-1+\delta)\bar{d}_{k-1}(t)}{(2+\xi)t} \right. \\
&\qquad \left. + O\left(\frac{1}{(nr^2)^{c_2/2c_0}}\right)\right| \\
&\quad = \left|\bar{d}_k(t) - f_k t - \left(\frac{k-1+\delta}{2+\xi} f_{k-1} - \frac{k+\delta}{2+\xi} f_k\right) \right. \\
&\qquad - \frac{(k+\delta)\bar{d}_k(t)}{(2+\xi)t} + \frac{(k-1+\delta)\bar{d}_{k-1}(t)}{(2+\xi)t} \\
&\qquad \left. + O\left(\frac{1}{(nr^2)^{c_2/2c_0}}\right)\right| \\
&\quad \leqslant \left(1 - \frac{k+\delta}{(2+\xi)t}\right)|\bar{d}_k(t) - f_k t| \\
&\qquad + \frac{k-1+\delta}{(2+\xi)t}|d_{k-1}(t) - f_{k-1}t| \\
&\qquad + O\left(\frac{1}{(nr^2)^{c_2/2c_0}}\right) \\
&\quad \leqslant M\left(t_r + \frac{n+Lt}{(nr^2)^{c_2/2c_0}}\right) + L\frac{1}{(nr^2)^{c_2/2c_0}} \\
&\quad \leqslant M\left(t_r + \frac{n+L(t+1)}{(nr^2)^{c_2/2c_0}}\right),
\end{aligned}$$

which completes the induction and thus the proof of Theorem 2.3. □

6.2. *Estimation of $T_t(u)$ – alternating concentration analysis*

In this section, we prove the Alternating Concentration Theorem (Lemma 6.1).

As mentioned earlier, Flaxman *et al.* (2007a; 2007b) and van den Esker (2008) introduced a new parameter $\alpha > 2$ to facilitate the analysis, and then used the traditional coupling technique to bound $T_t(u)$. We do not need to use the additional parameter α here, but will still get a nice bound. Our idea is to develop a refined method based on the recurrence directly implied in the definition of $T_t(u)$. By using this recurrence, we can start from the weak bound given in Lemma 3.5, and iteratively improve both the upper and lower bounds of $T_t(u)$. This improvement can be done using the submartingale and supermartingale concentration inequalities as in the proof of Lemma 5.3. This allows us to show that the accumulated error in the whole process is small, which guarantees the desired bound.

We will first show that a lower bound can be achieved from a rough lower bound on $T_t(u)$.

Lemma 6.2. Fix $r \geqslant r_0$. If for any $t \geqslant t_r$,

$$\Pr\left[T_t(u) \leqslant \left(b_l - \frac{r_l}{(nr^2)^{c_1/2c_0}}\right)(2+\xi)mA_r t\right] \leqslant \epsilon_l \tag{21}$$

for some $b_l \in [1/2, 1)$ and $r_l = o((nr^2)^{c_1/2c_0})$, then for any $t \geqslant t_r$,

$$\Pr\left[T_t(u) \geqslant \left(b_u + \frac{r_u}{(nr^2)^{c_1/2c_0}}\right)(2+\xi)mA_r t\right] \leqslant \epsilon_u, \tag{22}$$

where

$$\begin{aligned} b_u &= \frac{\xi+1}{2+\xi-\frac{1}{b_l}} \in (1,\infty) \\ r_u &= 7 + 40 r_l/\xi \\ \epsilon_u &= n\epsilon_l + 5n^{-\ln n+1} \end{aligned}$$

and c_1 is some constant satisfying the condition given in Equation (7).

Proof of Lemma 6.2. We will mainly use the following recurrence.

$$\begin{aligned} \mathrm{E}[T_{t+1}(u)|G_t] = {} & T_t(u) + m(1+\xi)\mathrm{E}[1_{|x_{t+1}-u|\leqslant r}|G_t] \\ & + \sum_{v\in V_t} m \Pr\left[y_i^{t+1} = v|G_t\right] 1_{|u-v|\leqslant r}, \end{aligned}$$

where

$$\Pr[y_i^{t+1} = v|G_t] = \mathrm{E}\left[\frac{(\deg_t(v)+\delta)1_{|x_{t+1}-v|\leqslant r}}{T_t(x_{t+1})}|G_t\right].$$

Let $\mathcal{G}$ denote the event that for all $t \geqslant t_r$, the following inequalities hold:

$$T_t(u) \geqslant \left(b_l - \frac{r_l}{(nr^2)^{c_1/2c_0}}\right)(2+\xi)mA_r t$$

$$\left(1 - \frac{1}{(nr^2)^{c_1/2c_0}}\right)(1+\xi)mA_r t \leqslant T_t(u) \leqslant 4(2+\xi)mA_r\left(1 + \frac{1}{(nr^2)^{c_1/2c_0}}\right).$$

https://doi.org/10.1017/S0960129511000570 Published online by Cambridge University Press

Then by Lemma 3.5 and the bound given in (21), $\Pr[\neg\mathscr{G}] \leqslant n\epsilon_l + 4n^{-\ln n+1}$. Conditioned on $\mathscr{G}$, for $t \geqslant t_r$, we have

$$\begin{aligned}
\Pr[y_i^{t+1} = v | G_t, \mathscr{G}] &\leqslant \mathrm{E}\left[\frac{(\deg_t(v) + \delta)1_{v \in B_r(x_{t+1})}}{\left(b_l - \frac{r_l}{(nr^2)^{c_1/2c_0}}\right)(2+\xi)mA_r t} | G_t, \mathscr{G}\right] \\
&= \frac{\deg_t(v) + \delta}{\left(b_l - \frac{r_l}{(nr^2)^{c_1/2c_0}}\right)(2+\xi)mt} \\
&\leqslant \frac{\deg_t(v) + \delta}{b_l(2+\xi)mt}\left(1 + \frac{4r_l}{(nr^2)^{c_1/2c_0}}\right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathrm{E}[T_{t+1}(u)|G_t, \mathscr{G}] &\leqslant T_t(u) + m(1+\xi)A_r + \frac{1}{b_l(2+\xi)t}\left(1 + \frac{4r_l}{(nr^2)^{c_1/2c_0}}\right)T_t(u) \\
&\leqslant \left(1 + \frac{1}{b_l(2+\xi)t}\right)T_t(u) + \left(\xi + 1 + \frac{40r_l}{(nr^2)^{c_1/2c_0}}\right)mA_r,
\end{aligned}$$

where the second inequality uses the rough upper bound on $T_t(u)$ given in Lemma 3.5.

Let

$$b_u = \frac{\xi + 1}{2 + \xi - \frac{1}{b_l}},$$

and $s = 40r_l/\xi$. Then

$$\begin{aligned}
&\mathrm{E}[T_{t+1}(u)|G_t, \mathscr{G}] - \left(b_u + \frac{s}{(nr^2)^{c_1/2c_0}}\right)(2+\xi)mA_r(t+1) \\
&\quad\leqslant \left(1 + \frac{1}{b_l(2+\xi)t}\right)\left(T_t(u) - \left(b_u + \frac{s}{(nr^2)^{c_1/2c_0}}\right)(2+\xi)mA_r t\right) \\
&\qquad + \left(\frac{b_u}{b_l} + \xi + 1 - b_u(2+\xi) + (2s + 40r_l - s(2+\xi))\frac{1}{(nr^2)^{c_1/2c_0}}\right)mA_r \\
&\quad\leqslant \left(1 + \frac{1}{b_l(2+\xi)t}\right)\left(T_t(u) - \left(b_u + \frac{s}{(nr^2)^{c_1/2c_0}}\right)(2+\xi)mA_r t\right). \qquad (23)
\end{aligned}$$

Now define

$$X_i = \begin{cases} \dfrac{T_i(u) - \left(b_u + \frac{s}{(nr^2)^{c_1/2c_0}}\right)(2+\xi)mA_r i}{\prod_{j=t_r}^{i-1}\left(1 + \frac{1}{b_l(2+\xi)j}\right)} & \text{for } i > t_r \\[2ex] T_{t_r}(u) - \left(b_u + \frac{s}{(nr^2)^{c_1/2c_0}}\right)(2+\xi)mA_r t_r & \text{for } i = t_r. \end{cases}$$

https://doi.org/10.1017/S0960129511000570 Published online by Cambridge University Press

From inequality (23), we know that $\mathrm{E}[X_i|G_{i-1},\mathscr{G}] \leqslant X_{i-1}$ for $t_r < i \leqslant t$. Let

$$\Delta_i = \prod_{j=t_r}^{i}\left(1+\frac{1}{b_l(2+\xi)j}\right) \sim \left(\frac{i}{t_r}\right)^{1/b_l(2+\xi)}.$$

We have

$$X_i - \mathrm{E}[X_i|G_{i-1},\mathscr{G}] = \frac{T_i(u)-\mathrm{E}[T_i(u)|G_{i-1},\mathscr{G}]}{\Delta_{i-1}} \leqslant (2+\xi)m,$$

and

$$\begin{aligned}\mathrm{Var}[X_i|G_{i-1},\mathscr{G}] &= \frac{\mathrm{Var}[T_i(u)|G_{i-1},\mathscr{G}]}{\Delta_{i-1}^2}\\ &\leqslant \frac{\mathrm{E}[(T_i(u)-T_{i-1}(u))^2|G_{i-1},\mathscr{G}]}{\Delta_{i-1}^2}\\ &\leqslant (2+\xi)m\frac{\frac{T_{i-1}(u)}{b_l(2+\xi)(i-1)}+\left(\xi+1+\frac{40r_l}{(nr^2)^{c_1/2c_0}}\right)mA_r}{\Delta_{i-1}^2}\\ &\leqslant \frac{(\xi+3)^2m^2A_r}{\Delta_{i-1}^2}.\end{aligned}$$

Therefore, the sequence $X_{t_r},\ldots,X_t$ satisfies the conditions in Lemma 3.2 with

$$\Pr[\neg\mathscr{G}] \leqslant n\epsilon + 4n^{-\ln n+1}$$

and

$$\begin{aligned}\sum_{i=t_r+1}^{t}\mathrm{Var}[X_i|G_{i-1},\mathscr{G}] &\leqslant \sum_{i=t_r+1}^{t}\frac{(\xi+3)^2m^2A_r}{\Delta_{i-1}^2}\\ &\leqslant \sum_{i=t_r+1}^{t}\frac{(\xi+3)^2m^2A_rt_r^{2/b_l(2+\xi)}}{i^{2/b_l(2+\xi)}} \qquad (24)\\ &\leqslant \left(\frac{mA_rt_r}{\ln n}\right)^2.\end{aligned}$$

The last inequality can be seen using the fact that $A_rt_r \sim 3(\ln n)^2(nr^2)^{c_1/c_0}$ and the assumption that $(c_0-c_1-1)(1-1/(\xi+2)) < c_1$. Specifically:

— If $2/b_l(\xi+2)=1$, then

$$\begin{aligned}\sum_{i=t_r+1}^{t}\frac{(\xi+3)^2m^2A_rt_r^{2/b_l(2+\xi)}}{i^{2/b_l(2+\xi)}} &\leqslant O\left(m^2A_rt_r\ln(t/t_r)\right)\\ &= O\left(\frac{m^2A_r^2t_r^2}{A_rt_r/\ln\ln n}\right)\\ &\leqslant \left(\frac{mA_rt_r}{\ln n}\right)^2.\end{aligned}$$

https://doi.org/10.1017/S0960129511000570 Published online by Cambridge University Press

— If $2/b_l(\xi+2) > 1$, then

$$\sum_{i=t_r+1}^{t} \frac{(\xi+3)^2 m^2 A_r t_r^{2/b_l(2+\xi)}}{i^{2/b_l(2+\xi)}} \leqslant O(m^2 A_r t_r) = O\left(\frac{m^2 A_r^2 t_r^2}{A_r t_r}\right) \leqslant \left(\frac{m A_r t_r}{\ln n}\right)^2.$$

— If $2/b_l(\xi+2) < 1$, then

$$\begin{aligned}\sum_{i=t_r+1}^{t} \frac{(\xi+3)^2 m^2 A_r t_r^{2/b_l(2+\xi)}}{i^{2/b_l(2+\xi)}} &\leqslant O\left(m^2 A_r t \left(\frac{t_r}{t}\right)^{2/b_l(\xi+2)}\right)\\ &= O\left(\frac{m^2 A_r^2 t_r^2}{A_r t_r (\frac{t_r}{t})^{1-2/b_l(\xi+2)}}\right)\\ &\leqslant \frac{m^2 A_r^2 t_r^2}{3(\ln n)^{2+2\left(1-\frac{2}{b_l(\xi+2)}\right)} (nr^2)^{c_1/c_0+(c_1/c_0-1)\left(1-\frac{2}{b_l(\xi+2)}\right)}}\\ &\leqslant \frac{m^2 A_r^2 t_r^2}{3(\ln n)^{2+2\left(1-\frac{2}{b_l(\xi+2)}\right)} (nr^2)^{c_1/c_0+(c_1/c_0-1)\left(1-\frac{2}{(\xi+2)}\right)}}\\ &\leqslant \left(\frac{m A_r t_r}{\ln n}\right)^2.\end{aligned}$$

If we let $\lambda = 2mA_r t_r$, then using the submartingale concentration inequality, we have

$$\begin{aligned}\Pr[X_t \geqslant X_{t_r} + \lambda] &\leqslant e^{-\frac{\lambda^2}{2\sum_{j=t_r+1}^{t} \mathrm{Var}[X_i|G_{i-1},\mathscr{G}]+2(2+\xi)m\lambda/3}} + \Pr[\neg\mathscr{G}]\\ &\leqslant n\epsilon_l + 5n^{-\ln n+1}.\end{aligned}$$

On the other hand, we have $X_{t_r} \leqslant 5mA_r t_r$ conditioned on $\mathscr{G}$. Thus,

$$\begin{aligned}\Delta_{t-1}(X_{t_r}+\lambda) &\leqslant 7\left(\frac{t}{t_r}\right)^{1/b_l(2+\xi)} mA_r t_r\\ &= 7\left(\frac{t_r}{t}\right)^{1-1/b_l(2+\xi)} mA_r t\\ &\leqslant \frac{7}{(nr^2)^{c_1/2c_0}} mA_r t,\end{aligned}$$

where the last inequality follows from the assumption that

$$(2c_0 - 2c_1 - 2)(1 - 2/(2+\xi)) > c_1.$$

https://doi.org/10.1017/S0960129511000570 Published online by Cambridge University Press

Therefore,

$$\Pr\left[T_t(u) \geqslant \left(b_u + \frac{s+7}{(nr^2)^{c_1/2c_0}}\right)(2+\xi)mA_r t\right]$$

$$\leqslant \Pr\left[\frac{T_t(u) - \left(b_u + \frac{s}{(nr^2)^{c_1/2c_0}}\right)(2+\xi)mA_r t}{\Delta_{t-1}} \geqslant X_{t_r} + \lambda\right]$$

$$\leqslant \Pr[X_t \geqslant X_{t_r} + \lambda]$$

$$\leqslant n\epsilon_l + 5n^{-\ln n + 1}.$$

We then complete the proof by letting

$$r_u = 7 + 40 r_l/\xi$$

$$\epsilon_u = n\epsilon_l + 5n^{-\ln n+1}.$$

Similarly, we can obtain an upper bound on $T_t(u)$ from a rough upper bound. □

Lemma 6.3. Fix $r \geqslant r_0$. If for any $t \geqslant t_r$,

$$\Pr\left[T_t(u) \geqslant \left(b_u + \frac{r_u}{(nr^2)^{c_1/2c_0}}\right)(2+\xi)mA_r t\right] \leqslant \epsilon_u, \tag{25}$$

for some $b_u \in (1,4)$ and $r_u = o\left((nr^2)^{c_1/2c_0}\right)$, then for any $t \geqslant t_0$,

$$\Pr\left[T_t(u) \leqslant \left(b_l - \frac{r_l}{(nr^2)^{c_1/2c_0}}\right)(2+\xi)mA_r t\right] \leqslant \epsilon_l, \tag{26}$$

where

$$b_l = \frac{\xi+1}{2+\xi-\frac{1}{b_u}} \in (1/2, 1)$$

$$r_l = 7 + 40 r_u/\xi$$

$$\epsilon_l = n\epsilon_u + 5n^{-\ln n+1},$$

and c_1 is some constant satisfying the condition given in Equation (7).

Proof. The proof is similar to the proof of Lemma 6.2. Note that in this case we use the supermartingale concentration inequality and let $\mathscr{G}'$ denote the good event defined in a similar way to $\mathscr{G}$ in the above proof, which then leads to the following recurrence:

$$\mathrm{E}[T_{t+1}(u)|G_t,\mathscr{G}'] \geqslant T_t(u) + m(1+\xi)A_r + \frac{1}{b_u(2+\xi)t}\left(1 - \frac{r_u}{(nr^2)^{c_1/2c_0}}\right)T_t(u)$$

$$\geqslant \left(1 + \frac{1}{b_u(2+\xi)t}\right)T_t(u) + \left(\xi + 1 - \frac{5r_u}{(nr^2)^{c_1/2c_0}}\right)mA_r.$$

Let

$$b_l = \frac{\xi+1}{2+\xi-\frac{1}{b_u}}$$

$$s' = 40 r_u/\xi.$$

https://doi.org/10.1017/S0960129511000570 Published online by Cambridge University Press

Then

$$\begin{aligned}
\mathrm{E}[T_{t+1}(u)|G_t,\mathscr{G}'] &- \left(b_l - \frac{s'}{(nr^2)^{c_1/2c_0}}\right)(2+\xi)mA_r(t+1) \\
&\geqslant \left(1+\frac{1}{b_u(2+\xi)t}\right)\left(T_t(u) - \left(b_l - \frac{s'}{(nr^2)^{c_1/2c_0}}\right)(2+\xi)mA_r t\right) \\
&\quad + \left(\frac{b_l}{b_u} + \xi + 1 - b_l(2+\xi) + (-s' - 5r_u + s'(2+\xi))\frac{1}{(nr^2)^{c_1/2c_0}}\right) mA_r \\
&\geqslant \left(1+\frac{1}{b_u(2+\xi)t}\right)\left(T_t(u) - \left(b_l - \frac{s'}{(nr^2)^{c_1/2c_0}}\right)(2+\xi)mA_r t\right). \qquad (27)
\end{aligned}$$

We then define the corresponding supermartingale $X'_{t_r},\ldots,X'_t$ using the above inequality. In this case, we will also use the conditions given by Equation (7) on the constants c_0 and c_1. Then, by setting

$$\lambda' = \frac{29}{4} mA_r t,$$

where λ' corresponds to the parameter λ in Lemma 6.2, and using

$$X'_{t_r} \geqslant \frac{1}{4} mA_r t_r,$$

we get

$$\begin{aligned}
\Pr&\left[T_t(u) \leqslant \left(b_l - \frac{s'+7}{(nr^2)^{c_1/2c_0}}\right)(2+\xi)mA_r t\right] \\
&\leqslant \Pr\left[\frac{T_t(u) - \left(b_l - \frac{s'}{(nr^2)^{c_1/2c_0}}\right)(2+\xi)mA_r t}{\Delta_{t-1}} \leqslant X'_{t_r} - \lambda'\right] \\
&\leqslant \Pr[X'_t \leqslant X'_{t_r} - \lambda'] \\
&\leqslant n\epsilon_u + 5n^{-\ln n + 1}.
\end{aligned}$$

We then complete the proof by letting

$$\begin{aligned}
r_l &= 7 + 40 r_u/\xi \\
\epsilon_l &= n\epsilon_u + 5n^{-\ln n+1}.
\end{aligned}$$

□

We are now ready to prove Lemma 6.1. Intuitively, we will apply the above two lemmas iteratively and show that if we start with a rough lower bound l_1, then, by Lemma 6.2, we can get an upper bound u, from which we can again get a new lower bound l_2 by Lemma 6.3. We prove that $l_2 > l_1$, which means that we get a better lower bound at each iteration. The same holds for the upper bound.

Proof of Lemma 6.1. If

$$\frac{\xi+1}{2+\xi-\frac{2+\xi}{1+\xi}} > 4,$$

we start our iterative process from the rough upper bound in Lemma 3.5; otherwise, we start the process from the rough lower bound.

https://doi.org/10.1017/S0960129511000570 Published online by Cambridge University Press

Assume we start from the rough lower bound; the case of starting from the rough upper bound is similar. By Lemma 3.5, we know that for all $t \geqslant t_r$,

$$T_t(u) \geqslant \left(1 - \frac{1}{(nr^2)^{c_1/2c_0}}\right)(1+\xi)mA_r t$$

with probability at least $1 - 4n^{-\ln n}$. We define the start point of our iterative process by letting

$$b_l^{(1)} = \frac{1+\xi}{2+\xi} \in (1/2, 1)$$
$$r_l^{(1)} = 1$$
$$\epsilon_l^{(1)} = 5n^{-\ln n}.$$

For $i \geqslant 1$, we assume that we have

$$T_t(u) \geqslant \left(b_l^{(i)} - \frac{r_l^{(i)}}{(nr^2)^{c_1/2c_0}}\right)(2+\xi)mA_r t$$

with error probability $\epsilon_l^{(i)}$ for any $t \geqslant t_r$. We now substitute the corresponding parameters in Lemma 6.2 to give an upper bound

$$T_t(u) \leqslant \left(b_u^{(i)} + \frac{r_u^{(i)}}{(nr^2)^{c_1/2c_0}}\right)(2+\xi)mA_r t$$

for all $t \geqslant t_r$ with error probability $\epsilon_u^{(i)}$, where

$$b_u^{(i)} = \frac{1+\xi}{2+\xi-\frac{1}{b_l^{(i)}}} \in (1,4]$$
$$r_u^{(i)} = (7+40/\xi)r_l^{(i)} \geqslant 7 + 40r_l^{(i)}/\xi$$
$$\epsilon_u^{(i)} = n\epsilon_l^{(i)} + 5n^{-\ln n+1}.$$

We again substitute the corresponding parameters in Lemma 6.3 to give an improved lower bound

$$T_t(u) \geqslant \left(b_l^{(i+1)} - \frac{r_l^{(i+1)}}{(nr^2)^{c_1/2c_0}}\right)(2+\xi)mA_r t$$

for all $t \geqslant t_r$ with error $\epsilon_l^{(i+1)}$, where

$$b_l^{(i+1)} = \frac{1+\xi}{2+\xi-\frac{1}{b_u^{(i)}}} \in (1/2,1)$$
$$r_l^{(i+1)} = (7+40/\xi)r_u^{(i)} \geqslant 7 + 40r_u^{(i)}/\xi$$
$$\epsilon_l^{(i+1)} = n\epsilon_u^{(i)} + 5n^{-\ln n+1}.$$

https://doi.org/10.1017/S0960129511000570 Published online by Cambridge University Press

Let $C(\xi) = 7 + 40/\xi$. Then

$$r_l^{(i+1)} = C(\xi)^2 r_l^{(i)}$$
$$\epsilon_l^{(i+1)} \leqslant n^2 \epsilon_l^{(i)} + 10n^{-\ln n + 2}.$$

We now show that for every i, we have $b_l^{(i+1)}$ is strictly greater than $b_l^{(i)}$, that is, the process gives a better lower bound after every pair of consecutive steps. Then, by the fact that $b_l^{(i)} < 1$, we have that $\{b_l^{(i)}\}_{i \geqslant 1}$ converges to 1. We can then show similarly that the procedure gives a better upper bound; in other words, $\{b_u^{(i)}\}_{i \geqslant 1}$ is a decreasing sequence that converges to 1. In the following, we will actually prove the stronger result that after each iteration, the distance between $b_l^{(i)}$ and 1 decreases by a multiple factor, which guarantees that the $\{b_l^{(i)}\}_{i \geqslant 1}$ converges *quickly* to 1.

We calculate the distance between $b_l^{(i+1)}$ and 1, which gives

$$\begin{aligned} 1 - b_l^{(i+1)} &= 1 - \frac{1+\xi}{2+\xi-\frac{1}{b_u^{(i)}}} \\ &= 1 - \frac{1+\xi}{2+\xi-\frac{1}{\frac{1+\xi}{2+\xi-\frac{1}{b_l^{(i)}}}}} \\ &= \frac{1-b_l^{(i)}}{\xi(2+\xi)b_l^{(i)}+1} \\ &\leqslant \frac{1-b_l^{(i)}}{\xi(1+\xi/2)+1}. \end{aligned}$$

Therefore, the sequence $\{1 - b_l^{(i)}\}_{i \geqslant 1}$ decreases by a multiple factor of at least

$$\frac{1}{\xi(1+\xi/2)+1}$$

at each step. On the other hand, since

$$T_t(u) \geqslant \left[1 - \left(1 - b_l^{(i)}\right) - \frac{r_l^{(i)}}{(\ln n)^{c_1}}\right](2+\xi)mA_r t,$$

the best bound is determined by the maximum of

$$\frac{r_l^{(i)}}{(\ln n)^{c_1}} \quad \text{and} \quad 1 - b_l^{(i)},$$

which is at most

$$\frac{1/2}{(\xi(1+\xi/2)+1)^i}.$$

We terminate the iteration at the step

$$k_0 = \left\lceil \frac{(c_1/2c_0)\ln(nr^2)}{\ln(C(\xi)^2(\xi(1+\xi/2)+1))} \right\rceil \leqslant \frac{\ln n}{4},$$

https://doi.org/10.1017/S0960129511000570 Published online by Cambridge University Press

in which case

$$\frac{1/2}{(\xi(1+\xi/2)+1)^{k_0}} \leqslant \frac{r_l^{(k_0)}}{(nr^2)^{c_1/2c_0}} = \frac{C(\xi)^{2k_0}}{(nr^2)^{c_1/2c_0}},$$

and

$$\begin{aligned}\Pr\left[T_t(u) \leqslant \left(1 - \frac{1}{(nr^2)^{c_2/2c_0}}\right)(2+\xi)mA_r t\right] &\leqslant \Pr\left[T_t(u) \leqslant \left(1 - \frac{2C(\xi)^{2k_0}}{(nr^2)^{c_1/2c_0}}\right)(2+\xi)mA_r t\right] \\ &\leqslant \epsilon_l^{(k_0)} \\ &\leqslant 2n^{2k_0 - \ln n + 2} \\ &\leqslant n^{-\ln n/2+2},\end{aligned}$$

where we have used the assumption that

$$c_2 = c_1 \frac{\ln(\xi(1+\xi/2)+1)}{\ln(C(\xi)^2(\xi(1+\xi/2)+1))}.$$

The upper bound can be obtained similarly by noting that the sequence $\{b_u^{(i)} - 1\}_{i\geqslant 1}$ decreases by a multiple factor of at least

$$\frac{1}{\xi(2+\xi)+1} \leqslant \frac{1}{\xi(1+\xi/2)+1}$$

at each step. Hence, we have

$$\Pr\left[|T_t(u) - (2+\xi)mA_r t| \geqslant \frac{1}{(nr^2)^{c_2/2c_0}} mA_r t\right] \leqslant n^{-2}, \tag{28}$$

which completes the proof. □

6.3. *Power law distribution of the generalised models*

In this section, we prove Theorem 2.4, using both the result of Theorem 2.3 and its proof.

Proof of Theorem 2.4 (Power law degree distribution). Since the local-degree sequences in the hybrid model are exactly the same as the degree sequences in the base model, by Theorem 2.3, the local graph of G_n^{H} has the power law degree distribution.

For the self-loop model, the degree of a node v can be expressed as $\deg_t(v) + \delta$, where $\deg_t(v)$ is the number of non-flexible edges incident to v at time t. We can now write the recurrence as follows:

$$\begin{aligned} \mathrm{E}[d_{k+\delta}(t+1)|G_t] = d_{k+\delta}(t) \\ + \sum_{v\in D_{k-1+\delta}(t)} \left(m\mathrm{E}\left[\frac{(k-1+\delta)1_{|x_{t+1}-v|\leqslant r}}{T_t(x_{t+1})} |G_t \right] \right) \\ - \sum_{v\in D_{k+\delta}(t)} \left(m\mathrm{E}\left[\frac{(k+\delta)1_{|x_{t+1}-v|\leqslant r}}{T_t(x_{t+1})} |G_t \right] \right) \\ + O\left(m\mathrm{E}[\eta_k(G_t, x_{t+1})|G_t] \right), \end{aligned} \tag{29}$$

Solving the recurrence, we can also arrive at Equation (19), which means the solution has the form

$$\frac{\phi_k'(m,\delta)}{(k+\delta)^{3+\xi}},$$

where $\phi_k'(m,\delta)$ tends to a limit $\phi_\infty'(m,\delta)$ that depends only on m and δ as $k \to \infty$. This completes the proof that the degree sequence of the self-loop model follows a power law distribution. □

7. Large community and small expander

In this section, we will prove Theorem 2.5.

Before proving the result, we will give a brief discussion on the choice of r. In the previous sections, we considered the case when $r = n^{-1/2}(\ln n)^{c_0}$ for some sufficiently large constant c_0. Both the base model and the two generalised models have the small-community phenomenon and the power law degree distribution. Now we consider other choices of r and show that if r is too small or too large, there is some strong evidence suggesting that the model does not have the power law degree distribution in the first case or the small-community phenomenon in the second.

When r is as small as $r = n^{-1/2-\epsilon}$ for any $\epsilon > 0$, each node connects only a very small fraction of its neighbours, and the whole graph is almost surely disconnected (Penrose 2003). Furthermore, there are many isolated vertices in the base model for this range of r, which indicates that the base model is very unlikely to have the power law degree distribution.

When r is as large as $r = n^{-1/2+\epsilon}$ for any $\epsilon > 0$, we have shown that the models have the power law degree distribution. However, the small-community phenomenon does not seem to exist in this situation. In particular, there is an interesting division of the structure of the R-neighbourhood when R varies. Specifically, we showed in Li and Peng (2011) that for this range of r, if $R = n^{-1/2+\rho}$ for any $\rho > \epsilon$, then with high probability we have $C_R(v)$, for any v, is an (α,β)-community for some constants α, β of size $\Theta(n^{2\rho})$, which indicates that every node belongs to some large community. Here we show that with high probability, for all $R = o(r)$, and for any $v \in V_n$, the conductance $\Phi(C_R(v))$ of $C_R(v)$ is larger than some constant, which indicates that the R-neighbourhood is not a good community.

We will now give the proof of Theorem 2.5.

https://doi.org/10.1017/S0960129511000570 Published online by Cambridge University Press

Proof of Theorem 2.5 (Large community and small expander). The proof of the first part of the theorem was given in Li and Peng (2011), so we will just prove the second part here.

For some fixed $R = o(r)$, we let $C = C_R(v)$ and $C' = C_{r-R}(v)$ for convenience. Then for any vertex $u \in C$ and $u' \in C'$, the distance between u and u' is at most r. The areas of $B_R(v)$ and $B_{r-R}(v)$ are given by

$$\text{area}(B_R(v)) \sim R^2/4$$
$$\text{area}(B_{r-R}(v)) \sim (r-R)^2/4 \sim r^2/4,$$

respectively, which means that a uniformly generated point will land in $B_R(v)$ and $B_{r-R}(v)$ with probabilities $R^2/4$ and $r^2/4$, respectively.

We will show that there are many edges between $C' \backslash C$ and C. To be more specific, let C_1 (or C_1') be the vertices in C (or C') that were created at or before time $n/2$, and C_2 (or C_2') be the set of vertices in C (or C') that were created after time $n/2$. We will show that the sum of the number of edges $e(C_1, C_2')$ between C_1 and C_2', and the number of edges $e(C_2, C_1')$ between C_2 and C_1' is large.

Let $\mathscr{E}$ denote the event that for any $u \in V_n$ and for each $t \geqslant t_0$,

$$T_t(u) \leqslant 8(2+\xi)mA_r t.$$

Then, by Lemma 3.5, the probability that $\mathscr{E}$ holds is $1 - O(n^{-\ln n})$. Now, conditioned on $\mathscr{E}$, for any vertex $x_j \in C_2'$, the probability that the ith contact of x_j lies in C_1 is at least

$$\frac{(m+\delta)|C_1|}{T_{j-1}(x_j)} \geqslant \frac{(1+\xi)|C_1|}{4(2+\xi)A_r n} \geqslant \frac{|C_1|}{8A_r n}.$$

Hence, $|e(C_1, C_2')|$ dominates

$$Bi\left(m|C_2'|, \frac{|C_1|}{8A_r n}\right),$$

where $Bi(N, p)$ denotes the binomial distribution with parameters N and p.

Similarly, for any vertex $x_j \in C_2$, the probability that the ith contact of x_j lies in C_1' is thus at least

$$\frac{(m+\delta)|C_1'|}{T_{j-1}(x_j)} \geqslant \frac{(1+\xi)|C_1'|}{4(2+\xi)A_r n} \geqslant \frac{|C_1'|}{8A_r n}.$$

Hence, $|e(C_2, C_1')|$ dominates

$$Bi\left(m|C_2|, \frac{|C_1'|}{8A_r n}\right).$$

In total, the expected number of edges between C and $C' \backslash C$ is

$$\mathrm{E}[|e(C, C' \backslash C)|] \geqslant \frac{m|C_2'||C_1|}{8A_r n} + \frac{m|C_2||C_1'|}{8A_r n},$$

which is at least $m|C|/16$ conditioned on the event $\mathscr{A}$ that C_1' and C_2' are both of size at least $A_r n/4$. Therefore, by Hoeffdings inequality and the fact that $\Pr[\neg\mathscr{A}] = O(n^{-3})$, we have

$$|e(C, \bar{C})| \geqslant |e(C, C' \backslash C)| \geqslant m|C|/32$$

https://doi.org/10.1017/S0960129511000570 Published online by Cambridge University Press

with probability at least

$$1 - e^{-m|C|/32}.$$

On the other hand, $|C| = o(A_r n)$ with high probability. Therefore,

$$\Pr[\exists R = o(r), \exists v, |e(C_R(v), \bar{C}_R(v))| \leqslant m|C_R(v)|/32] \leqslant \sum_{k=1}^{o(A_r n)} \binom{n}{k} e^{-mk/32} = o(1),$$

where the last inequality follows from the assumption that $m \geqslant K \ln n$ for some large constant K.

We now note that

$$\mathrm{vol}(C_R(v)) \leqslant m|C_R(v)| + |e(C_R(v), \bar{C}_R(v))|,$$

so we have

$$\Phi(C_R(v)) \geqslant \frac{m|C_R(v)|/32}{m|C_R(v)| + m|C_R(v)|/32} = \Omega(1) \tag{30}$$

with high probability, which completes the proof of Theorem 2.5. □

Finally, note that the above proof can be adapted to the two generalised models G_n^{H} and G_n^{S}. Since the number of long edges is relatively small compared with the number of local edges, the effect of long edges does not change the community structure too much. Specifically, to show that for $R = o(r)$, with $C_R(v)$ an expander in G_n^{H} and G_n^{S}, we just need to use the fact that

$$\mathrm{vol}(C_R(v)) \leqslant (m+1)|C_R(v)| + |e(C_R(v), \bar{C}_R(v))|$$

and

$$|e(C_R(v), \bar{C}_R(v))| \geqslant m|C|/32,$$

which follows in exactly the same way as in the above.

8. Conclusion

We have investigated the small-community phenomenon in networks and given two models that unify the three properties typical of large-scale networks: the power law degree distribution; the small-community phenomenon; and the small-diameter property. The proposed network models provide insights into how real networks evolve, and may have potential applications in, for example, wireless *ad hoc* models and sensor networks.

We have shown that the choice of parameters is subtle if one wants all three properties to coexist. We discussed the fundamental conflicts between them, that is, the fact that the power law degree distribution generated by the preferential attachment scheme and the small-diameter property always lead to an expander-like graph, while the small-community phenomenon corresponds naturally to an anti-expander in some sense, which means that the conductance of many subsets of small size is of order $o(1)$. Other reasons for such conflicts are worth further investigation.

Finally, our proof technique for the power law degree distribution is interesting in its own right, and partially solves the open problems in Flaxman *et al.* (2007a). It would be interesting to find other applications of this method, in particular, in the analysis of randomised algorithms and network modelling.

References

Albert, R., Jeong, H. and Barabási, A.-L. (1999) The diameter of the World Wide Web. *Nature* **401** 130–131.

Allen, C. (2004) Life with alacrity: The dunbar number as a limit to group sizes. Available at http://www.lifewithalacrity.com/2004/03/the_dunbar_numb.html.

Avin, C. (2008) Distance graphs: from random geometric graphs to Bernoulli graphs and between. In: *Proceedings of the fifth international workshop on Foundations of mobile computing* 71–78.

Bollobás, B. (2001) *Random Graphs* (second edition), Cambridge University Press.

Bollobás, B. (2003) Mathematical results on scale-free random graphs. In: *In Handbook of Graphs and Networks*, Wiley-VCH 1–34.

Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A. and Wiener, J. (2000) Graph structure in the web. *Computer Networks* **33** (1–6) 309–320.

Chung, F. and Lu, L. (2004) The small world phenomenon in hybrid power law graphs. In: Ben-Naim, E. *et al.* (eds.) *Complex Networks*, Springer-Verlag 91–106.

Chung, F. and Lu, L. (2006) *Complex Graphs and Networks (CBMS Regional Conference Series in Mathematics)*, American Mathematical Society.

Devroye, L. and Lu, J. (1995) The strong convergence of maximal degrees in uniform random recursive trees and dags. *Random Structures and Algorithms* **7** 1–14.

Dubhashi, D. and Panconesi, A. (2009) *Concentration of Measure for the Analysis of Randomized Algorithms* (first edition), Cambridge University Press.

Flaxman, A. D., Frieze, A. M. and Vera, J. (2007a) A geometric preferential attachment model of networks. *Internet Mathematics* **3** (2) 187–205.

Flaxman, A. D., Frieze, A. M. and Vera, J. (2007b) A geometric preferential attachment model of networks II. *Internet Mathematics* **4** (1) 87–111.

Fraigniaud, P. and Giakkoupis, G. (2009) The effect of power-law degrees on the navigability of small worlds [extended abstract]. In: *Proceedings of the 28th ACM symposium on Principles of distributed computing: PODC '09*, ACM 240–249.

Groh, G. and Rappel, V. (2009) Towards demarcation and modeling of small sub-communities/groups in p2p social networks. In: *CSE '09: Proceedings of the 2009 International Conference on Computational Science and Engineering*, IEEE Computer Society 304–311.

Horton, D. and Wohl, R. R. (1956) Mass communication and para-social interaction: Observations on intimacy at a distance. *Psychiatry* **19** (3) 215–229.

Jordan, J. (2010) Degree sequences of geometric preferential attachment graphs. *Advances in Applied Probability* **42** 319–330.

Kannan, R., Vempala, S. and Vetta, A. (2004) On clusterings: Good, bad and spectral. *Journal of the ACM* **51** (3) 497–515.

Kleinberg, J. (2000) The small-world phenomenon: an algorithmic perspective. In: *Proceedings of the 32nd ACM Symposium on the Theory of Computing*.

Kleinberg, J. M., Kumar, R., Raghavan, P., Rajagopalan, S. and Tomkins, A. S. (1999) The Web as a graph: measurements, models and methods. In: Proceedings of the 5th Annual International Computing and Combinatorics Conference (COCOON). *Springer-Verlag Lecture Notes in Computer Science* **1627** 1–18.

Kurucz, M. and Benczúr, A. A. (2010) Geographically organized small communities and the hardness of clustering social networks. *Annals of Information Systems – Data Mining for Social Network Analyis.*

Lang, K. (2005) Fixing two weaknesses of the spectral method. In: NIPS '05. *Advances in Neural Information Processing Systems* **18**.

Leskovec, J., Lang, K. J., Dasgupta, A. and Mahoney, M. W. (2008) Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. (Informal publication: CoRR abs/0810.1355.)

Leskovec, J., Lang, K. J. and Mahoney, M. (2010) Empirical comparison of algorithms for network community detection. In: *Proceedings of the 19th international conference on World wide web: WWW '10*, ACM 631–640.

Li, A. and Peng, P. (2011) Communities structures in classical network models. *Internet Mathematics* **7** (2) 81–106.

Penrose, M. (2003) *Random Geometric Graphs*, Oxford University Press.

Pittel, B. (1994) Note on the heights of random recursive trees and random *m*-ary search trees. *Random Structures and Algorithms* **5** 337–347.

Ravasz, E. and Barabási, A.-L. (2003) Hierarchical organization in complex networks. *Physical Review E* **67** 026112.

Smythe, R. and Mahmoud, H. (1995) A survey of recursive trees. *Theory of Probability and Mathematical Statistics* **51** 1–27.

van den Esker, H. (2008) A geometric preferential attachment model with fitness. (Available at http://arxiv.org/abs/0801.1612.)