Adv. Appl. Prob. **44,** 583–601 (2012)
Printed in Northern Ireland

TYPICAL DISTANCES IN ULTRASMALL RANDOM NETWORKS

STEFFEN DEREICH,* *Philipps-Universität Marburg*
CHRISTIAN MÖNCH ** AND
PETER MÖRTERS,** *** *University of Bath*

Abstract

We show that in preferential attachment models with power-law exponent $\tau \in (2, 3)$ the distance between randomly chosen vertices in the giant component is asymptotically equal to $(4 + o(1)) \log\log N/(-\log(\tau - 2))$, where N denotes the number of nodes. This is twice the value obtained for the configuration model with the same power-law exponent. The extra factor reveals the different structure of typical shortest paths in preferential attachment graphs.

Keywords: Scale-free network; preferential attachment; configuration model; power-law graph; conditionally Poissonian graph; graph distance; diameter

2010 Mathematics Subject Classification: Primary 05C80
Secondary 60C05; 90B15

1. Introduction

One of the central observations in the theory of scale-free random networks is that in the case of power-law exponents $\tau \in (2, 3)$ networks are *ultrasmall*, which means that the distance between two randomly chosen nodes in the giant component of a graph with N vertices is of asymptotic order $\log\log N$. The first analytical, but mathematically nonrigorous, evidence for this general phenomenon can be found, for example, in Cohen and Havlin [3] or Dorogovtsev *et al.* [10], and there are also some early papers with rigorous results for specific network models, in particular the work of Reittu and Norros [14] and the work of Chung and Lu [4].

In the present paper we refine this observation and identify graph distances including constant factors. Our main result is a universal technique for proving lower bounds for typical distances, which in a wide range of examples matches the upper bounds known from the recent literature. The result is presented in the form of two theorems, which reveal that ultrasmall networks can be divided into two different *universality classes*. For the class of ultrasmall preferential attachment models, the typical distances turn out to be twice as large as for fitness models. This difference corresponds to different structures of typical shortest paths in the network. We show that the two classes can be easily identified from the form of the attachment probability densities in the networks. We remark here that our work is focused on *typical distances* in networks, as results on diameters (see, e.g. [6]) tend to be model dependent and universality results are not to be expected.

Received 16 June 2011; revision received 31 October 2011.
* Postal address: Fachbereich Mathematik und Informatik, Philipps-Universität Marburg, Hans-Meerwein-Straße, 35032 Marburg, Germany.
** Postal address: Department of Mathematical Sciences, University of Bath, Claverton Down, Bath BA2 7AY, UK.
*** Email address: maspm@bath.ac.uk

At least informally, we have some structural insight into typical shortest paths in ultrasmall networks; see, for example, [13]. For the class we call *fitness models*, it turns out that typical vertices in the giant component can be connected with a few steps to a *core* of the network. Within this core there is a hierarchy of *layers* of nodes with increasing connectivity and at the top a small *inner core* of highly connected nodes with very small diameter. A typical shortest path inside the core runs from one layer to the next until the inner core is reached, and then climbing down again until a vertex in the lowest layer of the core is again connected to a typical vertex.

A high degree of a vertex is an indicator of its fitness and, thus, increases its connectivity to any other vertex. Hence, the layers can be identified by vertex degrees. Very roughly speaking, the jth layer consists of vertices with degree k_j, where

$$\log k_j \approx (\tau - 2)^{-j}$$

and there are about

$$\frac{\log \log N}{-\log(\tau - 2)}$$

layers. The graph distance of two randomly chosen vertices in the giant component is therefore

$$(2 + o(1)) \frac{\log \log N}{-\log(\tau - 2)}.$$

These asymptotics are rigorously confirmed for two variants of an inhomogeneous random graph model by Chung and Lu [4] and Norros and Reittu [12], and for the configuration model by van der Hofstad *et al.* [17]. See also van der Hofstad [16] for a summary of various results with detailed proofs. In general, upper bounds on the distances can be obtained by verifying the above strategy, while our Theorem 1 provides a flexible (i.e. model-independent) approach to the lower bound.

For the more complex class of ultrasmall *preferential attachment models*, existing results are far less complete. The work of Dommers *et al.* [6] strongly suggests that, for various ultrasmall preferential attachment models, the typical distance of two vertices in the giant component is bounded from above by

$$(4 + o(1)) \frac{\log \log N}{-\log(\tau - 2)}.$$

A corresponding lower bound, and, hence, confirmation of the exact factor 4, is stated as an interesting open problem by van der Hofstad and Hooghiemstra in [18, Section IV.B] and again in [6] (see the remark therein following Theorem 1.7 and Section 1.2). Our main result, Theorem 2, provides this bound and confirms, somewhat surprisingly, that the upper bound is sharp. Besides the models given in [6] we will also describe other examples of random network models in the same universality class, in which Theorem 2 applies.

Loosely speaking, the shortest paths in the class of preferential attachment models can be described as follows. Again, inside a core of highly connected vertices paths run from bottom to top and back through a hierarchy of layers defined as before. However, by construction of the preferential attachment models, a high degree of a vertex does not increase its connectivity to *all* vertices but only to those introduced *late* into the system (which are typically outside the core). Therefore, a path cannot directly connect one layer to another in one step, but it requires *two steps*. The paths run from one layer to a young vertex and from there back into the next higher layer. The distance between two typical vertices is therefore increased by a factor of two.

In the following section we formulate the precise results, consisting of two simple hypotheses on a random network leading to the two different lower bound results; see Theorems 1 and 2. We also provide a brief sketch of the proof technique and introduce the notation used in the proofs. In Section 3 we then discuss several examples of networks in the two universality classes. In all these examples upper bounds can either be found in the literature or derived by simple modifications of these proofs. Section 4 is devoted to the proofs of our main results.

2. Main results

A *(dynamic) network model* is a sequence of random graphs $(\mathcal{G}_N)_{N\in\mathbb{N}}$ with the set of vertices of $\mathcal{G}_N$ given by $[N] := \{1, 2, \dots, N\}$ and the set of unoriented edges of $\mathcal{G}_N$ given by a random symmetric subset of $[N] \times [N]$. Occasionally, we shall allow multiple edges between the same pair of vertices, but this has no bearing on the connectivity problems discussed here and is for convenience only. We write $v \leftrightarrow w$ if the vertices v and w are connected by an edge in the graph $\mathcal{G}_N$. The graph distance is given by

$$d_N(v, w) := \min\{n\colon \text{ there exists } v = v_0, v_1, \dots, v_n = w \in \mathcal{G}_N \text{ such that } v_{i-1} \leftrightarrow v_i \text{ for all } 1 \le i \le n\}.$$

The main aim of this paper is to provide techniques to find lower bounds on the *typical distance*, i.e. the asymptotic graph distance between two randomly chosen vertices in the graph $\mathcal{G}_N$.

We start with the easier of the two results, which is based on the following assumption.

Assumption FM(γ). *There exists κ such that, for all N and pairwise distinct vertices $v_0, \dots, v_\ell \in [N]$,*

$$\mathrm{P}\{v_0 \leftrightarrow v_1 \leftrightarrow v_2 \leftrightarrow \cdots \leftrightarrow v_\ell\} \le \prod_{k=1}^{\ell} \kappa v_{k-1}^{-\gamma} v_k^{-\gamma} N^{2\gamma-1}.$$

In random networks with power-law exponent τ, Assumption FM(γ) is typically satisfied for all $\gamma > (\tau-1)^{-1}$, and we expect these networks to be ultrasmall if $\frac{1}{2} < \gamma < 1$.

Theorem 1. *Let $(\mathcal{G}_N)_{N\in\mathbb{N}}$ be a dynamic network model that satisfies Assumption* FM(γ) *for some γ satisfying $\frac{1}{2} < \gamma < 1$. Then, for random vertices V and W chosen independently and uniformly from $[N]$, we have*

$$d_N(V, W) \ge 2\frac{\log\log N}{\log(\gamma/(1-\gamma))} + \mathcal{O}(1) \quad \textit{with high probability as } N \to \infty.$$

Five examples of network models in which Theorem 1 can be applied will be given in Section 3. Examples 3–5 in that section refer to the class of fitness models in which every vertex receives an a-priori fitness value which determines its likelihood to form future edges, including the configuration model, in which the fitness of a vertex equals its degree. In all these cases, if the degree distribution has power-law exponent $\tau \in (2, 3)$, Assumption FM(γ) is satisfied for all $\gamma > (\tau-1)^{-1}$, and Theorem 1 implies that

$$d_N(V, W) \ge (2 + o(1))\frac{\log\log N}{-\log(\tau-2)} \quad \text{with high probability as } N \to \infty.$$

For all examples of fitness models, matching upper bounds are known from the literature.

https://doi.org/10.1239/aap/1339878725 Published online by Cambridge University Press

While Theorem 1 applies to preferential attachment models, it does not give an optimal lower bound. Our main result is based on a strictly stronger assumption, which is tailored for the use in preferential attachment models and which gives an optimal lower bound for these models.

Assumption PA(γ). *There exists κ such that, for all N and pairwise distinct vertices $v_0, \dots, v_\ell \in [N]$,*

$$\mathrm{P}\{v_0 \leftrightarrow v_1 \leftrightarrow v_2 \leftrightarrow \cdots \leftrightarrow v_\ell\} \le \prod_{k=1}^{\ell} \kappa (v_{k-1} \wedge v_k)^{-\gamma} (v_{k-1} \vee v_k)^{\gamma-1}.$$

In preferential attachment models with power-law exponent τ, Assumption PA(γ) is typically satisfied for all $\gamma > (\tau-1)^{-1}$. Hence, we again expect these networks to be ultrasmall if and only if $\frac{1}{2} < \gamma < 1$. Theorem 2 below, our main result, gives a lower bound on the typical distance under this assumption.

Theorem 2. *Let $(\mathcal{G}_N)_{N\in\mathbb{N}}$ be a dynamic network model that satisfies Assumption* PA(γ) *for some γ satisfying $\frac{1}{2} < \gamma < 1$. Then, for random vertices V and W chosen independently and uniformly from $[N]$, we have*

$$d_N(V, W) \ge 4 \frac{\log\log N}{\log(\gamma/(1-\gamma))} + \mathcal{O}(1)$$

with high probability.

Examples of network models in which Theorem 2 can be applied will be given as Examples 1 and 2 in Section 3. They comprise various preferential attachment models with power-law exponent $\tau \in (2, 3)$. In all these cases Assumption PA(γ) is satisfied for all $\gamma > (\tau-1)^{-1}$, and Theorem 2 implies that

$$d_N(V, W) \ge (4 + o(1)) \frac{\log\log N}{-\log(\tau - 2)} \quad \text{with high probability as } N \to \infty.$$

Matching upper bounds are known from the literature.

The proof of both theorems is based on a *constrained* or *truncated first-order method*, which we now briefly explain. We start with an explanation of the (unconstrained) first moment bound and its shortcomings. Let v and w be distinct vertices of $\mathcal{G}_N$. Then, for $\delta \in \mathbb{N}$,

$$\begin{aligned} \mathrm{P}\{d_N(v, w) \le 2\delta\} &= \mathrm{P}\Bigg(\bigcup_{k=1}^{2\delta} \bigcup_{(v_1,\dots,v_{k-1})} \{v \leftrightarrow v_1 \leftrightarrow \cdots \leftrightarrow v_{k-1} \leftrightarrow w\} \Bigg) \\ &\le \sum_{k=1}^{2\delta} \sum_{(v_1,\dots,v_{k-1})} \prod_{j=1}^{k} p(v_{j-1}, v_j), \end{aligned}$$

where $(v_0, \dots, v_k)$ is any collection of pairwise distinct vertices in $\mathcal{G}_N$ with $v_0 = v$, and $v_k = w$, and, for $m, n \in \mathbb{N}$,

$$p(m, n) := \begin{cases} \kappa (m \wedge n)^{-\gamma} (m \vee n)^{\gamma-1} & \text{if Assumption PA}(\gamma) \text{ holds}, \\ \kappa m^{-\gamma} n^{-\gamma} N^{2\gamma-1} & \text{if Assumption FM}(\gamma) \text{ holds}. \end{cases}$$

Note that we can assign to each path $(v_0, \dots, v_k)$ the weight

$$p(v_0, \dots, v_k) := \prod_{j=1}^{k} p(v_{j-1}, v_j),$$

https://doi.org/10.1239/aap/1339878725 Published online by Cambridge University Press

and the upper bound is just the sum over the weights of all paths from v to w of length no more than 2δ. The shortcoming of this bound is that the paths that contribute most to the total weight are those that connect v or w quickly to vertices with extremely small indices. Since these are typically not present in the network, such paths have to be removed in order to get a reasonable estimate.

To this end, we define a decreasing sequence $(\ell_k)_{k=0,\dots,\delta}$ of positive integers and consider a tuple of vertices $(v_0, \dots, v_n)$ as *admissible* if $v_k \wedge v_{n-k} \ge \ell_k$ for all $k \in \{0, \dots, \delta \wedge n\}$. We denote by $A_k^{(v)}$ the event that there exists a path $v = v_0 \leftrightarrow \cdots \leftrightarrow v_k$ in the network such that $v_0 \ge \ell_0, \dots, v_{k-1} \ge \ell_{k-1}, v_k < \ell_k$, i.e. a path that traverses the threshold after exactly k steps. For fixed distinct vertices $v, w \ge \ell_0$, the truncated first moment estimate is

$$\mathrm{P}\{d_N(v,w) \le 2\delta\} \le \sum_{k=1}^{\delta} \mathrm{P}(A_k^{(v)}) + \sum_{k=1}^{\delta} \mathrm{P}(A_k^{(w)}) + \sum_{n=1}^{2\delta} \sum_{\substack{(v_0,\dots,v_n)\\ \text{admissible}}} \mathrm{P}\{v_0 \leftrightarrow \cdots \leftrightarrow v_n\}, \tag{1}$$

where the admissible paths in the last sum start with $v_0 = v$ and end with $v_n = w$. By assumption,

$$\mathrm{P}\{v_0 \leftrightarrow \cdots \leftrightarrow v_n\} \le p(v_0, \dots, v_n),$$

so, for $v \ge \ell_0$ and $k = 1, \dots, \delta$,

$$\mathrm{P}(A_k^{(v)}) \le \sum_{v_1=\ell_1}^{N} \cdots \sum_{v_{k-1}=\ell_{k-1}}^{N} \sum_{v_k=1}^{\ell_k - 1} p(v, v_1, \dots, v_k). \tag{2}$$

Given $\varepsilon > 0$, we choose $\ell_0 = \lceil \varepsilon N \rceil$ and $(\ell_j)_{j=0,\dots,k}$ decreasing fast enough so that the first two summands on the right-hand side of (1) together are no larger than 2ε. For $k \in \{1, \dots, \delta\}$, set

$$\mu_k^{(v)}(u) := \mathbf{1}_{\{v \ge \ell_0\}} \sum_{v_1=\ell_1}^{N} \cdots \sum_{v_{k-1}=\ell_{k-1}}^{N} p(v, v_1, \dots, v_{k-1}, u),$$

and set $\mu_0^{(v)}(u) = \mathbf{1}_{\{v=u\}}$. To rephrase the truncated moment estimate in terms of μ, note that p is symmetric, so, for all $n \le 2\delta$ and $n^* := \lfloor n/2 \rfloor$,

$$\begin{aligned}\sum_{\substack{(v_0,\dots,v_n)\\ \text{admissible}}} \mathrm{P}\{v_0 \leftrightarrow \cdots \leftrightarrow v_n\} &\le \sum_{v_1=\ell_1}^{N} \cdots \sum_{v_{n^*}=\ell_{n^*}}^{N} \cdots \sum_{v_{n-1}=\ell_1}^{N} p(v, \dots, v_{n^*}) p(v_{n^*}, \dots, w) \\ &= \sum_{v_{n^*}=\ell_{n^*}}^{N} \mu_{n^*}^{(v)}(v_{n^*}) \mu_{n-n^*}^{(w)}(v_{n^*}).\end{aligned} \tag{3}$$

Using the recursive representation

$$\mu_{k+1}^{(v)}(n) = \sum_{m=\ell_k}^{N} \mu_k^{(v)}(m) p(m, n),$$

we establish upper bounds for $\mu_k^{(v)}(u)$, and use these to show that the rightmost term in (1) remains small if δ is chosen sufficiently small. Using the input from Assumptions PA(γ), respectively FM(γ), this will lead to the lower bounds for the typical distance in Theorem 2, respectively Theorem 1. Detailed proofs will be given in Section 4.

3. Examples

In this section we give five examples, corresponding to the best understood models of ultrasmall networks in the mathematical literature. Examples 1–2 are of preferential attachment type and will be discussed using our main result, Theorem 2, while Examples 3–5 are of fitness type and will be discussed using Theorem 1.

Example 1. (*Preferential attachment with fixed outdegree.*) This class of models was studied in the work of Hooghiemstra, van der Hofstad, and coauthors. We base our discussion on [6], where three qualitatively similar models were considered; see also [16] for a survey. We focus on the first model studied in [6], which is the most convenient to define, as the two variants can be treated with the same method. The model depends on two parameters, an integer $\mathfrak{m} \geq 1$ and a real $\delta > -\mathfrak{m}$. Roughly speaking, in every step a new vertex is added to the network and connected to $\mathfrak{m}$ existing vertices with a probability proportional to their degree plus δ. Note that in the case $\mathfrak{m} = 1$ the network has the metric structure of a tree, making this a degenerate case of less interest. The case famously studied by Bollobás and Riordan [2] corresponds to $\delta = 0$ and $\mathfrak{m} \geq 2$, and leads to a network with $\tau = 3$ and typical distance $\log N/\log\log N$, so that it lies outside the class of ultrasmall networks.

We first generate a dynamic network model $(\mathcal{G}_N)$ for the case $\mathfrak{m} = 1$. By $Z[n, N]$, $n \leq N$, we denote the degree of vertex n in $\mathcal{G}_N$ (with the convention that self-loops add two towards the degree of the vertex to which they are attached).

- $\mathcal{G}_1$ consists of a single vertex, labelled 1, with one self-loop.
- In each further step, given $\mathcal{G}_N$, we insert one new vertex, labelled $N + 1$, and one new edge into the network such that the new edge connects the new vertex to vertex $m \in [N]$ with probability
$$\mathrm{P}\{m \leftrightarrow N+1 \mid \mathcal{G}_N\} = \frac{Z[m, N] + \delta}{N(2+\delta) + 1 + \delta},$$
or to itself with probability
$$\frac{1+\delta}{N(2+\delta) + 1 + \delta}.$$

To generalise the model to arbitrary values of $\mathfrak{m}$, we take the graph $\mathcal{G}'_{\mathfrak{m}N}$ constructed using parameters $\mathfrak{m}' = 1$ and $\delta' = \delta/\mathfrak{m}$, and merge vertices $\mathfrak{m}(k-1)+1, \dots, \mathfrak{m}k$ in the graph $\mathcal{G}'_{\mathfrak{m}N}$ into a single vertex denoted k, keeping all edges. We obtain asymptotic degree distributions which are power laws with exponent $\tau = 3 + \delta/\mathfrak{m}$, so we expect the model to be in the ultrasmall range if and only if $-\mathfrak{m} < \delta < 0$.

Proposition 1. *For independent, uniformly chosen vertices V and W in the giant component of the preferential attachment model with parameters $\mathfrak{m} \geq 2$ and $-\mathfrak{m} < \delta < 0$, we have*

$$d_N(V, W) = (4 + o(1)) \frac{\log\log N}{-\log(1+\delta/\mathfrak{m})} \quad \textit{with high probability.}$$

Remark 1. The upper bound can be proved by an adaption of the argument in [6]; see below. This paper leaves the problem of finding a lower bound open. We resolve this problem by verifying Assumption PA(γ) for $\gamma = (2 + \delta/\mathfrak{m})^{-1}$ and applying Theorem 2.

Proof of Proposition 1. For the lower bound, we look at $\mathfrak{m} = 1$ first. In this case we have, for $1 \le m < n \le N$,

$$\mathrm{P}\{m \leftrightarrow n\} = \frac{\mathrm{E}\, Z[m, n-1] + \delta}{n(2+\delta) - 1}. \tag{4}$$

It is easy to see that

$$\mathrm{E}[Z[m,n] + \delta \mid Z[m, n-1]] = (Z[m, n-1] + \delta)\frac{n(2+\delta)}{n(2+\delta)-1},$$

and, hence,

$$\mathrm{E}[Z[m,n] + \delta] = (1+\delta)\frac{\Gamma(n+1)\Gamma(m - 1/(2+\delta))}{\Gamma(n + (1+\delta)/(2+\delta))\Gamma(m)}.$$

In particular, there exist constants $0 < c < C$ such that

$$c\left(\frac{n}{m}\right)^{1/(2+\delta)} \le \mathrm{E}\, Z[m,n] \le C\left(\frac{n}{m}\right)^{1/(2+\delta)} \quad \text{for all } 1 \le m < n.$$

Combining this with (4) yields, for $\gamma = 1/(2+\delta)$ and a suitable $\kappa_1 > 0$,

$$\mathrm{P}\{m \leftrightarrow n\} \le \frac{C(n/m)^{\gamma} + \delta}{n(2+\delta) - 1} \le \kappa_1 n^{\gamma - 1} m^{-\gamma} \quad \text{for all } 1 \le m < n. \tag{5}$$

To verify Assumption PA(γ), following [6, Lemma 2.1] we find that, for distinct vertices $v_0, \dots, v_l$, all events of the form $\{v_{j-1} \leftrightarrow v_j \leftrightarrow v_{j+1}\}$ with $j \in \{1, \dots, l-1\}$ and $v_j < v_{j-1}, v_{j+1}$, and all events $\{v_{j-1} \leftrightarrow v_j\}$ which are not part of these, are nonpositively correlated, in the sense that the probability of all of them occurring is smaller than the product of the probabilities. Recalling also (5), it remains to show that, for $m < v, w$,

$$\mathrm{P}\{v \leftrightarrow m \leftrightarrow w\} \le \kappa_2 v^{\gamma-1} w^{\gamma-1} m^{-2\gamma} \tag{6}$$

for some finite constant $\kappa_2 > 0$. To this end, we let $\{(Z_n^{(k,m)})_{n \ge m} : k, m \in \mathbb{N}\}$ denote the collection of right-continuous Markov jump processes starting at $Z_{m-}^{(k,m)} = k$, jumping instantly at time m and subsequently at integer time steps following the rule

$$\mathrm{P}\{Z_n^{(k,m)} = Z_{n-}^{(k,m)} + 1 \mid Z_{n-}^{(k,m)}\} = \frac{Z_{n-}^{(k,m)} + \delta}{n(2+\delta) - \delta} = 1 - \mathrm{P}\{Z_n^{(k,m)} = Z_{n-}^{(k,m)} \mid Z_{n-}^{(k,m)}\}.$$

Note that $(Z[m,n])_{n \ge m} = (Z_n^{(1,m)})_{n \ge m}$ in law and that, for $m < n$, the event $\{m \leftrightarrow n\}$ corresponds to $\{\Delta Z_n^{(k,m)} = 1\}$, where we write $\Delta Z_n^{(k,m)} := Z_n^{(k,m)} - Z_{n-}^{(k,m)}$. Note also that $Z_n^{(k_0,m)}$ is stochastically dominated by $Z_n^{(k,m)}$ for $k \ge k_0$. Hence, for $m < n_1 < n_2$,

$$\begin{aligned}
\mathrm{E}[Z_{n_2}^{(2,m)} \mid \Delta Z_{n_1}^{(2,m)} = 1] &= \sum_{j=2}^{m-n_2+2} \sum_{k=2}^{m-n_1+1} j\, \mathrm{P}\{Z_{n_2}^{(2,m)} = j \mid Z_{n_1-}^{(2,m)} = k,\ \Delta Z_{n_1}^{(2,m)} = 1\} \\
&\qquad \times \mathrm{P}\{Z_{n_1-}^{(2,m)} = k \mid \Delta Z_{n_1}^{(2,m)} = 1\} \\
&\le \sum_{j=2}^{m-n_2+2} \sum_{k=2}^{m-n_1+1} \frac{j\, \mathrm{P}\{Z_{n_2}^{(k+1,n_1)} = j\}(k+\delta)\, \mathrm{P}\{Z_{n_1-}^{(2,m)} = k\}}{(n_1(2+\delta) + 1 + \delta)\, \mathrm{P}\{\Delta Z_{n_1}^{(2,m)} = 1\}} \\
&= \sum_{k=2}^{m-n_1+1} \frac{(k+\delta)\, \mathrm{P}\{Z_{n_1-}^{(2,m)} = k\}\, \mathrm{E}\, Z_{n_2}^{(k+1,n_1)}}{(n_1(2+\delta) + 1 + \delta)\, \mathrm{P}\{\Delta Z_{n_1}^{(2,m)} = 1\}}.
\end{aligned}$$

As in the derivation of (5), the expectation in the last line can be bounded from above by $c_0(k+1)n_2^{\gamma}n_1^{-\gamma}$ for some $c_0 > 0$. Similarly, we obtain $\mathrm{P}\{\Delta Z_{n_1}^{(2,m)} = 1\} \geq c_1 n_1^{\gamma-1} m^{-\gamma}$ and

$$\mathrm{E}[(Z_{n_1-}^{(2,m)})^2] \leq c_2 m^{-2/(2+\delta)} n_1^{2/(2+\delta)}$$

for further constants $c_1, c_2 > 0$. Summarising, we obtain

$$\mathrm{E}[Z_{n_2}^{(2,m)} \mid \Delta Z_{n_1}^{(2,m)} = 1] \leq c_3 n_2^{\gamma} n_1^{-2\gamma} m^{\gamma} \sum_{k=2}^{m-n_1+1} k^2\, \mathrm{P}\{Z_{n_1-}^{(2,m)} = k\} \leq c_4 n_2^{\gamma} m^{-\gamma}$$

for some $c_3, c_4 > 0$, and this establishes (6). Finally, passing from $\mathfrak{m} = 1$ to general $\mathfrak{m}$ can be achieved by a simple union bound.

For the upper bound, we work in the graph $\mathcal{G}_{2N}$ with $\mathfrak{m} \geq 2$ and $\delta \in (-\mathfrak{m}, 0)$. Using the terminology of [6], we define the *core* of $\mathcal{G}_{2N}$ to be

$$\mathrm{core}_N = \{m \in [N] \colon Z[m, N] \geq (\log N)^{\sigma}\},$$

where $\sigma > -\mathfrak{m}/\delta$. Dommers *et al.* [6, Theorem 3.1] stated that the diameter of the core in $\mathcal{G}_{2N}$ is bounded by $(4 + o(1)) \log\log N |\log(1 + \delta/\mathfrak{m})|^{-1}$; thus, all we need to show is that, for fixed $\varepsilon > 0$, a uniformly chosen vertex $V \in [\lfloor (2-\varepsilon)N \rfloor]$ can be connected to the core using no more then $o(\log\log N)$ edges in $\mathcal{G}_{2N}$. This is done in two steps. For the first step, we explore the neighbourhood of V in $\mathcal{G}_M$ for $M = \lfloor (2-\varepsilon)N \rfloor$ until we find a vertex w with degree $Z[w, N] \geq u_0$, where u_0 will be determined below. Denote by S_k, $k \geq 0$, the set of all vertices in $\mathcal{G}_M$ that can be reached from V using exactly k different edges from $\mathcal{G}_M$. If we fix $u \in \mathbb{N}$ and set $T_u^{(V)} = \min\{k \colon S_k \cap \{n \colon Z[n, N] \geq u\} \neq \varnothing\}$, then it is straightforward to verify, similarly to the proof of [6, Theorem 3.6], that we can find a large constant $C_{u,\varepsilon} > 0$ such that

$$\mathrm{P}\{T_u^{(V)} > C_{u,\varepsilon}\} < \varepsilon, \tag{7}$$

if N is sufficiently large. The second step is to show that any vertex w satisfying $Z[w, N] \geq u_0$ for sufficiently large u_0 can be joined to the core by using $\mathcal{O}(\log\log\log N)$ edges. To this end, we apply [6, Lemma A.1], as in the proof of [6, Proposition 3.3], to obtain, for any vertex $j \in \mathcal{M} = \{\lceil (2-\varepsilon)N \rceil, \dots, 2N\}$ and any vertex $a \in [N]$ with $Z[a, N] \geq u_a$,

$$\mathrm{P}\{j \leftrightarrow a,\ j \leftrightarrow b \text{ for some } b \text{ with } Z[b, N] \geq u_b \mid \mathcal{G}_M\} \geq \frac{c u_a u_b^{-(1+\delta/\mathfrak{m})}}{N}$$

for a positive constant c with probability exceeding $1 - o(N^{-1})$, and as long as both u_a and u_b do not exceed a small power of N. Hence, with the same probability,

$$\begin{aligned} &\mathrm{P}\{\text{there exists } j \in \mathcal{M} \colon j \leftrightarrow a,\ j \leftrightarrow b \text{ for some } b \text{ with} Z[b, N] \geq u_b \mid \mathcal{G}_M\} \\ &\quad\leq \left(1 - \frac{c u_a u_b^{-(1+\delta/\mathfrak{m})}}{N}\right)^{\#\mathcal{M}} \\ &\quad\leq \exp(-2c\varepsilon u_a u_b^{-(1+\delta/\mathfrak{m})}). \end{aligned} \tag{8}$$

Starting from the initial vertex w with $Z[w, N] \geq u_0$ and defining, for $k \geq 1$,

$$u_{k+1} = \left(\frac{\varepsilon c u_k}{\log(k+1) - (\log\varepsilon)/2}\right)^{1/(1+\delta/\mathfrak{m})}, \tag{9}$$

https://doi.org/10.1239/aap/1339878725 Published online by Cambridge University Press

it is straightforward to check that, for $u_a = u_k$ and $u_b = u_{k+1}$, the right-hand side of (8) equals $\varepsilon(k+1)^{-2}$. Summing over these error bounds, we therefore find that (9) defines an increasing sequence $(u_k)_{k=0}^K$ of lower bounds on degrees at time N, for which we know that, with probability at least $1-\pi^2\varepsilon/6$, there is a path of length $2K$ which alternates between high degree vertices and vertices from $\mathcal{M}$ and connects w to a vertex of degree u_K. The recursive definition (9) implies that

$$\log u_K \geq \left(\frac{1}{1+\delta/\mathfrak{m}}\right)^K (\log u_0 - C_\varepsilon)$$

for some large $C_\varepsilon > 0$; thus, if $u_0 = \exp(2C_\varepsilon)$, we can connect w to a vertex belonging to core_N by choosing $K \geq D_{\sigma,\varepsilon} \log\log\log N$, where $D_{\sigma,\varepsilon} > 0$ depends only on σ and ε. Fixing $u = \exp(2C_\varepsilon)$ in (7) and starting the above construction in $u_0 = u$, we obtain, for a uniformly chosen vertex $V \in \mathcal{G}_{2N}$,

$$\mathrm{P}\{d_{2N}(V, \text{core}_N) > 2D_{\sigma,\varepsilon} \log\log\log N + C_{e^{2C_\varepsilon},\varepsilon}\} \leq \left(2+\frac{\pi^2}{6}\right)\varepsilon,$$

if N is sufficiently large, showing that the diameter of core_N is the dominating contribution to typical distances in $\mathcal{G}_{2N}$.

A different class of preferential attachment models was introduced in [7] and further studied in [9]. Here a new vertex is connected to any existing vertex independently with a probability depending (possibly nonlinearly) on its degree. In this model the number of edges created in every step is asymptotically Poisson distributed.

Example 2. (*Preferential attachment with variable outdegree.*) This model was studied in the work of Dereich, Mörters, and coauthors; see [8] for a survey. The model depends on a concave function $f\colon \mathbb{N} \cup \{0\} \to (0, \infty)$, which is called the *attachment rule*. Roughly speaking, in every step a new vertex is added to the network and oriented edges from the new vertex to existing vertices are introduced independently with a probability proportional to the current degree of the existing vertex.

More precisely, to generate a dynamic network model $(\mathcal{G}_N)$, we assume that f satisfies $f(0) \leq 1$ and $f(1) - f(0) < 1$. An important parameter derived from f is the limit

$$\gamma := \lim_{n\to\infty} \frac{f(n)}{n},$$

which always exists with $0 \leq \gamma < 1$, by concavity. By $Z[n, N]$, $n \leq N$, we denote the number of *younger* vertices to which vertex n is connected in $\mathcal{G}_N$.

- $\mathcal{G}_1$ consists of a single vertex, labelled 1, and no edges.
- In the $(N+1)$th step, given $\mathcal{G}_N$, we insert one new vertex, labelled $N+1$, and independently for any $m \in [N]$, we introduce an edge from $N+1$ to m with probability

$$\frac{f(Z[m, N])}{N}.$$

By [7, Theorem 1.1(b)], the conditional distribution given $\mathcal{G}_N$ of the number of edges created in the $(N+1)$th step converges to a Poisson distribution and the empirical distribution of the

https://doi.org/10.1239/aap/1339878725 Published online by Cambridge University Press

degrees converges to a power law with exponent $\tau = 1 + 1/\gamma$, or, more precisely, to a random probability vector (μ_k) satisfying

$$\lim_{k\to\infty} \frac{\log \mu_k}{\log k} = 1 + \frac{1}{\gamma}.$$

We therefore expect the network to be ultrasmall if and only if $\gamma > \frac{1}{2}$.

Proposition 2. *For independent, uniformly chosen vertices V and W in the giant component of the preferential attachment model with attachment rule f and derived parameter $\gamma > \frac{1}{2}$, we have*

$$d_N(V, W) = (4 + o(1))\frac{\log \log N}{\log(\gamma/(1-\gamma))} \quad \textit{with high probability.}$$

Remark 2. The upper bound can be proved by adapting the argument of [6]; see the forthcoming thesis [11] for details. For the lower bound, we verify Assumption PA$(\gamma+\varepsilon)$ for any $\varepsilon > 0$ and apply Theorem 2.

Proof of Proposition 2. We first note that, for $v < w \in [N]$,

$$\mathrm{P}\{v \leftrightarrow w\} = \frac{\mathrm{E}\, f(Z[v, w-1])}{w-1}.$$

To estimate the expectation, we note that, by concavity, given $\varepsilon > 0$, there exists k such that, for all $n \ge k$, we have $f(n) \le f(k) + (\gamma+\varepsilon)(n-k)$. An easy calculation (see [9, Lemma 2.7]) shows that

$$\mathrm{E}\, f(Z[v, w-1]) \le C_1 w^{\gamma+\varepsilon} v^{-\gamma-\varepsilon} \quad \text{for a suitable constant } C_1 > 0. \tag{10}$$

We now use (10) to verify Assumption PA$(\gamma+\varepsilon)$. For $v < w \in [N]$, all events $\{v \leftrightarrow w\}$ with different values of v are independent. Hence, $\mathrm{P}\{v_0 \leftrightarrow \cdots \leftrightarrow v_n\}$ can be decomposed into factors of the form $\mathrm{P}\{v_{j-1} \leftrightarrow v_j \leftrightarrow v_{j+1}\}$ with $v_j < v_{j-1}, v_{j+1}$ and factors of the form $\mathrm{P}\{v_{j-1} \leftrightarrow v_j\}$ for the remaining edges. It remains to estimate factors of the former form. We may assume that $v < u < w$ and obtain

$$\mathrm{P}\{u \leftrightarrow v \leftrightarrow w\} = \frac{\mathrm{E}[f(Z[v, u-1]) f(Z[v, w-1])]}{(u-1)(w-1)}.$$

Arguing as in the derivation of (10) we obtain, for a suitable constant $C_2 > 0$,

$$\mathrm{E}[f(Z[v, w-1]) \mid Z[v, u-1] = k] \le C_2 f(k) w^{\gamma+\varepsilon} u^{-\gamma-\varepsilon}.$$

Hence,

$$\mathrm{E}[f(Z[v, u-1]) f(Z[v, w-1])] \le C_2\, \mathrm{E}[f(Z[v, u-1])^2] w^{\gamma+\varepsilon} u^{-\gamma-\varepsilon},$$

and, using a similar argument as above, we obtain $C_3 > 0$ such that

$$\mathrm{E}[f(Z[v, u-1])^2] \le C_3 u^{2(\gamma+\varepsilon)} v^{-2(\gamma+\varepsilon)}.$$

Summarising, we obtain a constant $C_4 > 0$ such that

$$\mathrm{P}\{u \leftrightarrow v \leftrightarrow w\} \le C_4 u^{\gamma-1+\varepsilon} v^{-2(\gamma+\varepsilon)} w^{\gamma-1+\varepsilon},$$

as required to complete the proof.

We now give three examples of random networks in the universality class of fitness models. The first two belong to the wide class of inhomogeneous random graphs, whose essential feature is the independence between different edges.

Example 3. (*Expected degree random graph.*) This model was studied in the work of Chung and Lu; see [4] or [5] for a survey. In its general form the model depends on a triangular scheme $w_1^{(N)}, \dots, w_N^{(N)}$ of positive weights, where the weight $w_i^{(N)}$ plays the role of the expected degree of vertex i in $\mathcal{G}_N$. The model is defined by the following two requirements:

- for every pair (i, j) with $1 \le i \ne j \le N$, the events $\{i \leftrightarrow j\}$ are independent,
- for every pair (i, j) with $1 \le i \ne j \le N$, we have

$$\mathrm{P}\{i \leftrightarrow j\} = \frac{w_i^{(N)} w_j^{(N)}}{\ell_N} \wedge 1, \quad \text{where} \quad \ell_N := \sum_{i=1}^{N} w_i^{(N)}.$$

Proposition 3. *For independent, uniformly chosen vertices V and W in the expected degree random graph with weights satisfying*

$$c\left(\frac{N}{i}\right)^{\gamma} \le w_i^{(N)} \le C\left(\frac{N}{i}\right)^{\gamma} \quad \textit{for all } 1 \le i \le N,$$

for some $\gamma > \frac{1}{2}$ and constants $0 < c \le C$, we have

$$d_N(V, W) = (2 + o(1)) \frac{\log\log N}{\log(\gamma/(1-\gamma))} \quad \textit{with high probability.}$$

Proof. The upper bound is sketched in [4]. For the lower bound, we have to check Assumption FM(γ). Note that, using the upper bound on the weights,

$$\mathrm{P}\{i \leftrightarrow j\} \le \frac{w_i^{(N)} w_j^{(N)}}{\ell_N} \le C^2 \frac{N^{2\gamma}}{\ell_N} (ij)^{-\gamma}.$$

From the lower bound on the weights we obtain $\ell_N \ge cN$ for some $c > 0$, and, hence, $\mathrm{P}\{i \leftrightarrow j\} \le \kappa N^{2\gamma-1} i^{-\gamma} j^{-\gamma}$ for a suitable κ. Using the independence assumption, we see that Assumption FM(γ) holds, and the lower bound follows from Theorem 1.

Example 4. (*Conditionally Poissonian random graph.*) This model was studied in the work of Norros and Reittu; see [12]. It is based on drawing an independent, identically distributed sequence $\Lambda_1, \Lambda_2, \dots$ of positive capacities. Conditional on this sequence, the dynamical network model is constructed as follows.

- $\mathcal{G}_1$ consists of a single vertex, labelled 1, and no edges.
- In the $(N + 1)$th step, given $\mathcal{G}_N$, we insert one new vertex, labelled $N + 1$, and independently for any $m \in [N]$, we introduce a random number of edges between $N + 1$ and m according to a Poisson distribution with parameter

$$\frac{\Lambda_i \Lambda_{N+1}}{L_{N+1}} \quad \text{for } L_n := \sum_{k=1}^{n} \Lambda_k.$$

https://doi.org/10.1239/aap/1339878725 Published online by Cambridge University Press

- We further remove each edge in $\mathcal{G}_N$ independently with probability $1 - L_N/L_{N+1}$, and, thus, obtain $\mathcal{G}_{N+1}$.

Recall that having possibly several edges between two vertices has no relevance for the typical distances in the giant component. In order to be in the ultrasmall regime, we require the law of the capacities to be power laws with exponent $2 < \tau < 3$.

Proposition 4. *Assume that the capacities in the conditionally Poissonian random graph satisfy*

$$\mathrm{P}\{\Lambda_1 > x\} = x^{1-\tau}(c + o(1)) \quad \textit{for all sufficiently large } x,$$

where $2 < \tau < 3$ and $c > 0$ is constant. For independent, uniformly chosen vertices V and W in the giant component, we have

$$d_N(V, W) = (2 + o(1)) \frac{\log\log N}{-\log(\tau - 2)} \quad \textit{with high probability.}$$

Remark 3. The upper bound was proved in [12, Theorem 4.2], where it was also shown that a giant component exists. For the lower bound, we verify Assumption FM(γ) for $\gamma = 1/(\tau - 1)$ and apply Theorem 1.

Proof of Proposition 4. We check that Assumption FM(γ) holds with high probability, conditionally given the capacities. For fixed N, we put the capacities in decreasing order,

$$\Lambda_N^{(1)} > \Lambda_N^{(2)} > \cdots > \Lambda_N^{(N)},$$

and relabel the vertices so that the jth vertex has weight $\Lambda_N^{(j)}$. We recall from [12, Proposition 2.1] that the number of edges between vertices i and j in $\mathcal{G}_N$ is Poisson distributed with parameter $\Lambda_N^{(i)}\Lambda_N^{(j)}/L_N$. As the edges are conditionally independent, we only have to verify that, given $\varepsilon > 0$, there exists $\kappa > 0$ such that

$$1 - \exp\left(-\frac{\Lambda_N^{(i)}\Lambda_N^{(j)}}{L_N}\right) \le \kappa N^{2\gamma-1} i^{-\gamma} j^{-\gamma} \quad \text{for all } 1 \le i < j \le N,$$

with probability greater than or equal to $1 - 2\varepsilon$. By the law of large numbers, L_N is of order N, so it suffices to establish that $\Lambda_N^{(i)} \le \kappa(N/i)^\gamma$ for all $1 \le i \le N$. To this end, we denote by $S_N^{(i)}$ the number of potential values exceeding $\kappa(N/i)^\gamma$. The random variable $S_N^{(i)}$ is binomially distributed with parameters N and $p := \mathrm{P}\{\Lambda_1 > \kappa(N/i)^\gamma\} \le c(\kappa)i/N$, where $c(\varkappa) \downarrow 0$ for $\varkappa \uparrow \infty$. By Bernstein's inequality, see, e.g. [1, Equation (8)],

$$\mathrm{P}\{S_N^{(i)} > 2i\} \le \exp\left(\frac{-i^2/2}{\mathrm{var}(S_N^{(i)}) + i/3}\right) \le \mathrm{e}^{-3i/8} \quad \text{if } c(\kappa) < 1.$$

Hence, we may choose M large enough so that $\sum_{i=M}^{\infty} \exp(-\frac{3}{8}i) < \varepsilon$, ensuring that, with probability exceeding $1 - \varepsilon$, we have $\Lambda_N^{(2i)} \le \kappa(N/i)^\gamma$ for all $i \ge M$. It remains to give bounds on $\Lambda_N^{(1)}, \ldots, \Lambda_N^{(2M)}$. By a standard Poisson approximation result, see, e.g. [15, Proposition 3.21], we note that, for any $1 \le i \le 2M$, $S_N^{(i)}$ converges weakly to a Poisson distribution with parameter $\lambda := \lim_{N\to\infty} N\,\mathrm{P}\{\Lambda_1 > \kappa(N/i)^\gamma\} \le 2c(\kappa)M$, and, hence, by choosing κ large, we can ensure that, for large N, we have $\sum_{i=1}^{2M} \mathrm{P}\{S_N^{(i)} > i\} \le \varepsilon$, which completes the proof.

A model which also falls in the universality class of fitness models is the random network with fixed degree sequence, or configuration model. This model is well studied and very detailed results on average distances in the case of power laws with exponent $\tau \in (2, 3)$ have been obtained, in particular by van der Hofstad *et al.* [17].

Example 5. (*Random networks with fixed degree sequence.*) The idea behind this class of models is to enforce a particular power-law exponent by fixing the degree sequence of the network in a first step. We therefore choose a sequence $D_1, D_2, \ldots$ of independent and identically distributed random variables with values in the nonnegative integers. For given N, we assume that

$$L_N := \sum_{j=1}^{N} D_j$$

is even, which may be achieved by replacing D_N by $D_N - 1$ if necessary. Thus, given $D_1, \ldots, D_N$, we construct the network $\mathcal{G}_N$ as follows.

- To any vertex $m \in [N]$ we attach D_m half-edges or stubs.
- The L_N stubs are given an (arbitrary) order.
- We start by pairing the first stub with another (uniformly) randomly chosen stub, and continue pairing the lowest numbered unpaired stub with a remaining randomly chosen stub until all stubs are matched.
- Any pair of stubs connect to form an edge.

Obviously, the resulting network can have self-loops and double edges, but this has no relevance for the typical distances in the giant component. In order to be in the ultrasmall regime, we require the law of the degrees to be a power law with exponent $2 < \tau < 3$.

Proposition 5. *Assume that there exists* $c > 0$ *such that*

$$\mathrm{P}\{D_1 > x\} = x^{1-\tau}(c + o(1)) \quad \textit{for all sufficiently large } x.$$

For independent, uniformly chosen vertices V *and* W *in the giant component, we have*

$$d_N(V, W) = (2 + o(1)) \frac{\log\log N}{-\log(\tau - 2)} \quad \textit{with high probability}.$$

Remark 4. This and much more is proved in [17, Theorem 1.2]. For an alternative approach to the lower bound, we now verify Assumption FM(γ) for any $\gamma < 1/(\tau - 1)$ and paths of length up to $\ell = \mathcal{O}(\log\log N)$, which is clearly sufficient to apply Theorem 1.

Proof of Proposition 5. We observe that, given $D_1, \ldots, D_N$, for pairwise disjoint vertices $v_1, \ldots, v_{\ell+1}$,

$$\mathrm{P}\{v_\ell \leftrightarrow v_{\ell+1} \mid v_1 \leftrightarrow v_2 \leftrightarrow \cdots \leftrightarrow v_{\ell-1} \leftrightarrow v_\ell\} \le \frac{D_{v_\ell} D_{v_{\ell+1}}}{L_N - 2\sum_{k=1}^{\ell} D_{v_k}},$$

where the denominator is a rough lower bound on the number of stubs unaffected by the conditioning event. In particular, $\mathrm{P}\{i \leftrightarrow j\} \le D_i D_j / (L_N - 2D_i)$. Using the law of large numbers, we can easily see that there exists a $c > 0$ such that

$$L_N - 2\sum_{k=1}^{\ell} D_{v_k} \ge cN \quad \text{with high probability}$$

for any choice of $v_1, \ldots, v_\ell$ if $\ell = \mathcal{O}(\log\log N)$. Therefore, to verify Assumption FM(γ), we only need to find appropriate bounds on the degrees of given vertices, which can be achieved (using the same relabelling) by a similar argument as in Example 4.

4. Proofs

4.1. Proof of Theorem 2

In this section we assume the validity of Assumption PA(γ) for a $\gamma \in (\frac{1}{2}, 1)$ with a fixed constant κ. Given a vector $(q(1), \dots, q(n))$, we use the notation

$$q[m] := \sum_{i=1}^{m} q(i) \quad \text{for all } 1 \le m \le n.$$

We adopt the notation of the discussion at the end of Section 2. In particular, recall the definition of $\mu_k^{(v)}$ and the key estimates (1), (2), and (3), which combine to give

$$\mathrm{P}\{d_N(v, w) \le 2\delta\} \le \sum_{k=1}^{\delta} \mu_k^{(v)}[\ell_k - 1] + \sum_{k=1}^{\delta} \mu_k^{(w)}[\ell_k - 1] + \sum_{n=1}^{2\delta} \sum_{u=\ell_{n^*}}^{N} \mu_{n^*}^{(v)}(u)\mu_{n-n^*}^{(w)}(u). \quad (11)$$

The remaining task of the proof is to choose $\delta \in \mathbb{N}$ and $2 \le \ell_\delta \le \dots \le \ell_0 \le N$ which allow the required estimates for the right-hand side. To do so, we will make use of the recursive representation

$$\mu_{k+1}^{(v)}(n) = \sum_{m=\ell_k}^{N} \mu_k^{(v)}(m) p(m, n) \quad \text{for } k \in \{0, \dots, \delta - 1\} \text{ and } n \in [N],$$

where $\mu_0^{(v)}(n) = \mathbf{1}_{\{v=n\}}$ and

$$p(m, n) = \kappa(m \wedge n)^{-\gamma}(m \vee n)^{\gamma-1}.$$

Denote by $\bar{\mu}_k^{(v)}(m) = \mathbf{1}_{\{m \ge \ell_k\}} \mu_k^{(v)}(m)$ the truncated version of $\mu_k^{(v)}$, and conceive $\mu_k^{(v)}$ and $\bar{\mu}_k^{(v)}$ as row vectors. Then

$$\mu_{k+1}^{(v)} = \bar{\mu}_k^{(v)} \boldsymbol{P}_N, \quad (12)$$

where $\boldsymbol{P}_N = (p(m, n))_{m,n=1,\dots,N}$. Our aim is to provide a majorant of the form

$$\mu_k^{(v)}(m) \le \alpha_k m^{-\gamma} + \mathbf{1}_{\{m > \ell_{k-1}\}} \beta_k m^{\gamma-1}$$

for suitably chosen parameters $\alpha_k, \beta_k \ge 0$. Key to this choice is the following lemma.

Lemma 1. *Suppose that* $2 \le \ell \le N$, $\alpha, \beta \ge 0$, *and* $q \colon [N] \to [0, \infty)$ *satisfies*

$$q(m) \le \mathbf{1}_{\{m \ge \ell\}}(\alpha m^{-\gamma} + \beta m^{\gamma-1}) \quad \textit{for all } m \in [N].$$

Then there exists a constant $c > 1$ *(depending only on* γ *and* κ*) such that*

$$q\boldsymbol{P}_N(m) \le c\left(\alpha \log\left(\frac{N}{\ell}\right) + \beta N^{2\gamma-1}\right)m^{-\gamma} + \mathbf{1}_{\{m>\ell\}} c\left(\alpha \ell^{1-2\gamma} + \beta \log\left(\frac{N}{\ell}\right)\right)m^{\gamma-1}$$

for all $m \in [N]$.

Proof. We have

$$\begin{aligned} q\boldsymbol{P}_N(m) &= \mathbf{1}_{\{m>\ell\}} \sum_{k=\ell}^{m-1} q(k)p(k, m) + \sum_{k=m\vee\ell}^{N} q(k)p(k, m) \\ &\le \mathbf{1}_{\{m>\ell\}} \sum_{k=\ell}^{m-1} \kappa(\alpha k^{-\gamma} + \beta k^{\gamma-1})k^{-\gamma}m^{\gamma-1} + \sum_{k=m\vee\ell}^{N} \kappa(\alpha k^{-\gamma} + \beta k^{\gamma-1})k^{\gamma-1}m^{-\gamma} \end{aligned}$$

https://doi.org/10.1239/aap/1339878725 Published online by Cambridge University Press

$$\leq \kappa\Big(\alpha \sum_{k=m\vee\ell}^{N} k^{-1} + \beta \sum_{k=m\vee\ell}^{N} k^{2\gamma-2}\Big) m^{-\gamma}$$
$$+ \mathbf{1}_{\{m>\ell\}}\, \kappa\Big(\alpha \sum_{k=\ell}^{m-1} k^{-2\gamma} + \beta \sum_{k=\ell}^{m-1} k^{-1}\Big) m^{\gamma-1}$$
$$\leq \kappa\Big(\alpha \log\Big(\frac{m}{\ell-1}\Big) + \frac{\beta}{2\gamma-1} N^{2\gamma-1}\Big) m^{-\gamma}$$
$$+ \mathbf{1}_{\{m>\ell\}}\, \kappa\Big(\frac{\alpha}{1-2\gamma}(\ell-1)^{1-2\gamma} + \beta \log\Big(\frac{m}{\ell-1}\Big)\Big) m^{\gamma-1}.$$

This immediately implies the assertion since $\ell \geq 2$ by assumption.

We apply Lemma 1 iteratively. Fix $\varepsilon > 0$ small and start with

$$\ell_0 = \lceil \varepsilon N \rceil, \qquad \alpha_1 = \kappa(\varepsilon N)^{\gamma-1}, \qquad \text{and} \qquad \beta_1 = \kappa(\varepsilon N)^{-\gamma}.$$

Fix $v \geq \ell_0$. Then, for all $m \in [N]$,

$$\mu_1^{(v)}(m) = p(v, m) \leq \kappa \ell_0^{\gamma-1} m^{-\gamma} + \mathbf{1}_{\{m>\ell_0\}}\, \kappa \ell_0^{-\gamma} m^{\gamma-1}$$
$$\leq \alpha_1 m^{-\gamma} + \mathbf{1}_{\{m>\ell_0\}}\, \beta_1 m^{\gamma-1}.$$

Now suppose that, for some $k \in \mathbb{N}$, we have chosen α_k, β_k, and an integer ℓ_{k-1} such that

$$\mu_k^{(v)}(m) \leq \alpha_k m^{-\gamma} + \beta_k m^{\gamma-1} \quad \text{for all } m \in [N].$$

We choose ℓ_k to be an integer satisfying

$$\frac{6\varepsilon}{\pi^2 k^2} \geq \frac{1}{1-\gamma} \alpha_k \ell_k^{1-\gamma}, \tag{13}$$

and assume that $\ell_k \geq 2$. Pick α_{k+1} and β_{k+1} such that

$$\alpha_{k+1} \geq c\Big(\alpha_k \log\Big(\frac{N}{\ell_k}\Big) + \beta_k N^{2\gamma-1}\Big), \qquad \beta_{k+1} \geq c\Big(\alpha_k \ell_k^{1-2\gamma} + \beta_k \log\Big(\frac{N}{\ell_k}\Big)\Big). \tag{14}$$

By the induction hypothesis we can apply Lemma 1 with $\ell = \ell_k$ and $q(m) = \bar{\mu}_k^{(v)}(m)$. Then, using (12),

$$\mu_{k+1}^{(v)}(m) \leq \alpha_{k+1} m^{-\gamma} + \mathbf{1}_{\{m>\ell_k\}}\, \beta_{k+1} m^{\gamma-1} \quad \text{for all } m \in [N], \tag{15}$$

showing that the induction can be carried forward up to the point where $\ell_k < 2$, say in step K. Summing (15) over $m \leq \ell_k - 1$ and using (13), we obtain

$$\mu_k^{(v)}[\ell_k - 1] \leq \frac{1}{1-\gamma} \alpha_k \ell_k^{1-\gamma} \leq \frac{6\varepsilon}{\pi^2 k^2}.$$

Hence, the first two summands on the right-hand side of (11) are together smaller than 2ε. It remains to choose $\delta = \delta(N)$ as large as possible while ensuring that $\delta < K$ and

$$\lim_{N\to\infty} \sum_{n=1}^{2\delta} \sum_{u=\ell_{n^*}}^{N} \mu_{n^*}^{(v)}(u) \mu_{n-n^*}^{(w)}(u) = 0.$$

https://doi.org/10.1239/aap/1339878725 Published online by Cambridge University Press

To this end, assume that ℓ_k is the largest integer satisfying (13), and that the parameters α_k and β_k are defined via the equalities in (14). To establish lower bounds for the decay of ℓ_k, we investigate the growth of $\eta_k := N/\ell_k > 0$.

Going backwards through the definitions yields, for an integer $k \geq 0$ with $k+1 < K$ and if the right-hand side is smaller or equal to $(N/3)^{1-\gamma}$,

$$\left(\eta_{k+2}^{-1} + \frac{1}{N}\right)^{\gamma-1} \leq \frac{c^2(k+2)^2}{k^2}\eta_k^{\gamma} + 2c^2\frac{(k+2)^2}{(k+1)^2}\eta_{k+1}^{1-\gamma}\log\eta_{k+1}.$$

In particular, it follows that $K > k+2$ in this case.

It is easy to check inductively that, for any solution of this system, there exist constants $b, B > 0$ (not depending on N) such that

$$\eta_k \leq b\exp\left(B\left(\sqrt{\frac{\gamma}{1-\gamma}}\right)^k\right) \tag{16}$$

for $k < K$ and, moreover, the right-hand side exceeds $(N/3)^{1-\gamma}$ before step K. We now use (15) to estimate

$$\begin{aligned}\sum_{n=1}^{2\delta}\sum_{u=\ell_k}^{N}\mu_{n^*}^{(v)}(u)\mu_{n-n^*}^{(w)}(u) &\leq 2\sum_{k=1}^{\delta}\sum_{u=\ell_k}^{N}(\alpha_k u^{-\gamma} + \beta_k u^{\gamma-1})^2\\ &\leq \frac{4}{2\gamma-1}\sum_{k=1}^{\delta}(\alpha_k^2\ell_k^{1-2\gamma} + \beta_k^2 N^{2\gamma-1})\\ &\leq \frac{4}{2\gamma-1}\delta(\alpha_\delta^2\ell_\delta^{1-2\gamma} + \beta_\delta^2 N^{2\gamma-1}).\end{aligned}$$

Using (13) and (16), the first summand in the bracket can be estimated as

$$\alpha_\delta^2\ell_\delta^{1-2\gamma} \leq \left(\delta^{-2}\frac{6\varepsilon}{\pi^2}(1-\gamma)\right)^2\ell_\delta^{-1} \leq \left(\frac{6\varepsilon}{b\pi^2}(1-\gamma)\right)^2\frac{1}{N\delta^4}\exp\left(B\left(\frac{\gamma}{1-\gamma}\right)^{\delta/2}\right).$$

Using equality in (14), we obtain $\beta_\delta \leq c(\alpha_\delta\ell_\delta^{1-2\gamma} + \alpha_\delta N^{1-2\gamma}\log(N/\ell_\delta))$. Noting that the second summand on the right-hand side is bounded by a multiple of the first, we find a constant $C_1 > 0$ such that $\beta_\delta^2 N^{2\gamma-1} \leq C_1\alpha_\delta^2\ell_\delta^{1-2\gamma}$, and, thus, for a suitable constant $C_2 > 0$,

$$\sum_{n=1}^{2\delta}\sum_{u=\ell_k}^{N}\mu_{n^*}^{(v)}(u)\mu_{n-n^*}^{(w)}(u) \leq C_2\frac{1}{N\delta^3}\exp\left(B\left(\frac{\gamma}{1-\gamma}\right)^{\delta/2}\right).$$

Hence, for a suitable constant $C > 0$, choosing

$$\delta \leq \frac{\log\log N}{\log\sqrt{\gamma/(1-\gamma)}} - C,$$

we find that the term we consider goes to 0 as $\mathcal{O}((\log\log N)^{-3})$. Note from (16) that this choice also ensures that $\ell_\delta \geq 2$. We have thus shown that

$$\mathrm{P}\{d_N(v,w) \geq 2\delta\} \leq 2\varepsilon + \mathcal{O}((\log\log N)^{-3}),$$

whenever $v, w \geq \ell_0 = \lceil\varepsilon N\rceil$, which implies the statement of Theorem 2.

4.2. Proof of Theorem 1

In this section we assume the validity of Assumption FM(γ) for some $\gamma \in (\frac{1}{2}, 1)$ with a fixed constant $\kappa \geq 1$. Recall again the notation and framework from the introductory section. We use the same approach as in the proof of Theorem 2, but now we have to consider the *symmetric* matrix $\boldsymbol{P}_N := (p(m,n))_{m,n\in[N]}$ given by

$$p(m,n) := \kappa m^{-\gamma} n^{-\gamma} N^{2\gamma-1} \quad \text{for } m,n \in [N]. \tag{17}$$

We obtain the following lemma, which is the analogue of Lemma 1.

Lemma 2. *Suppose that* $2 \leq \ell \leq N$ *and* $q\colon [N] \to [0,\infty)$ *satisfies*

$$q(m) \leq \mathbf{1}_{\{m\geq \ell\}}\, m^{\gamma-1}\ell^{-\gamma} \quad \textit{for all } m \in [N].$$

Then, for all $m \in [N]$*,*

$$q\boldsymbol{P}_N(m) \leq \kappa m^{-\gamma} N^{\gamma-1}\left(\frac{N}{\ell}\right)^{\gamma} \log\left(\frac{N}{\ell-1}\right).$$

Proof. By (17) and the assumption on q,

$$q\boldsymbol{P}_N(m) = \sum_{i=1}^{N} q(i)p(i,m) \leq \kappa m^{-\gamma}\ell^{-\gamma}N^{2\gamma-1}\sum_{i=\ell}^{N}\frac{1}{i} \leq \kappa m^{-\gamma}\ell^{-\gamma}N^{2\gamma-1}\log\left(\frac{N}{\ell-1}\right),$$

which implies the statement of the lemma.

For fixed $\varepsilon > 0$, we first construct inductively a strictly decreasing sequence of integers (ℓ_k) by letting $\ell_0 = \lceil \varepsilon N \rceil$ and defining ℓ_{k+1} as the largest integer such that

$$\frac{\kappa}{1-\gamma}\left(\frac{\ell_{k+1}}{N}\right)^{1-\gamma} \leq \frac{6\varepsilon}{\pi^2(k+1)^2}\left(\log\left(\frac{N}{\ell_k - 1}\right)\right)^{-1}\left(\frac{\ell_k}{N}\right)^{\gamma}. \tag{18}$$

We stop once we find $\ell_k \leq 1$, say in step K. Recall the definition and recursive formula for $\mu_k^{(v)}$, and let $\bar{\mu}_k^{(v)}(m) := \mathbf{1}_{\{m\geq \ell_k\}}\mu_k^{(v)}(m)$. Then $\mu_{k+1}^{(v)}(m) = \bar{\mu}_k^{(v)}\boldsymbol{P}_N(m)$. We now show for $k = 1, \ldots, K-1$ that

$$\mu_k^{(v)}(m) \leq \kappa m^{-\gamma}N^{\gamma-1}\left(\frac{N}{\ell_{k-1}}\right)^{\gamma}\log\left(\frac{N}{\ell_{k-1}-1}\right) \leq m^{-\gamma}\ell_k^{\gamma-1} \quad \text{for all } m \in [N]. \tag{19}$$

Indeed, for $k = 1$, the statement follows from (17) and (18). We then continue by induction using Lemma 2. Considering the truncated first moment estimate (1) for $\delta < K$ and our choice of (ℓ_k), we obtain, from (19),

$$\mathrm{P}(A_k^{(v)}) \leq \mu_k^{(v)}[\ell_k - 1] \leq \frac{\kappa}{1-\gamma}\left(\frac{\ell_k}{N}\right)^{1-\gamma}\left(\frac{N}{\ell_{k-1}}\right)^{\gamma}\log\left(\frac{N}{\ell_{k-1}-1}\right).$$

Hence, (18) entails that $\sum_{k=1}^{\delta}\mathrm{P}(A_k^{(v)}) \leq \varepsilon$. The last step is to choose $\delta = \delta(N)$ as large as possible while ensuring that $\delta < K$ and

$$\sum_{n=1}^{2\delta}\sum_{u=\ell_{n^*}}^{N}\mu_{n^*}^{(v)}(u)\mu_{n-n^*}^{(w)}(u) \tag{20}$$

https://doi.org/10.1239/aap/1339878725 Published online by Cambridge University Press

goes to 0 as $N \to \infty$. By (19), this term can be bounded by a constant multiple of $N^{2\gamma-2} \sum_{k=1}^{\delta} \ell_k^{1-2\gamma}$. To verify (20), we have to bound the growth of the values $\eta_k := N/\ell_k$. The choice made in (18) implies that, for $k < K$ and if the right-hand side is smaller than $(N/3)^{1-\gamma}$,

$$\left(\eta_{k+1}^{-1} + \frac{1}{N}\right)^{\gamma-1} < \frac{\pi^2 \kappa}{1-\gamma} \frac{(k+1)^2}{6\varepsilon} \eta_k^{\gamma} \log(2\eta_k) \quad \text{for } k \geq 0.$$

In particular, we have $k + 1 < K$ in this case.

From this, it is straightforward to verify inductively the existence of constants $b, B > 0$, which only depend on ε, κ, and γ, such that

$$\eta_k \leq b \exp\left(B\left(\frac{\gamma}{1-\gamma}\right)^k\right) \quad \text{for } k < K,$$

and, moreover, the right-hand side exceeds $(N/3)^{1-\gamma}$ before step K. Hence, we may choose a suitable constant $C > 0$ such that, for

$$\delta \leq \frac{\log \log N}{\log(\gamma/(1-\gamma))} - C,$$

we have $\ell_\delta \geq 2$. To complete the proof, we note that

$$N^{2\gamma-2} \sum_{k=1}^{\delta} \ell_k^{1-2\gamma} \leq \frac{1}{N} \sum_{k=1}^{\delta} \eta_k^{2\gamma-1} \leq \delta b N^{B(\gamma/(1-\gamma))^{-C}-1},$$

which implies convergence in (20) when C is chosen large enough.

Acknowledgements

We would like to acknowledge fruitful discussions with Remco van der Hofstad at the 2011 BMS Summer School clarifying the upper bound in Proposition 1, and with Ilkka Norros at the Isaac-Newton Institute, Cambridge, during the programme on Stochastic Processes in Communication Sciences. The third author would like to acknowledge the support of the EPSRC through the award of an Advanced Research Fellowship.

References

[1] Bennett, G. (1962). Probability inequalities for the sum of independent random variables. *J. Amer. Statist. Assoc.* **57,** 33–45.

[2] Bollobás, B. and Riordan, O. (2004). The diameter of a scale-free random graph. *Combinatorica* **24,** 5–34.

[3] Cohen, R. and Havlin, S. (2003). Scale-free networks are ultrasmall. *Phys. Rev. Lett.* **90,** 058701, 4 pp.

[4] Chung, F. and Lu, L. (2003). The average distance in a random graph with given expected degrees. *Internet Math.* **1,** 91–113.

[5] Chung, F. and Lu, L. (2006). *Complex Graphs and Networks* (CBMS Regional Conf. Math. **107**). American Mathematical Society, Providence RI.

[6] Dommers, S., van der Hofstad, R. and Hooghiemstra, G. (2010). Diameters in preferential attachment models. *J. Statist. Phys.* **139,** 72–107.

[7] Dereich, S. and Mörters, P. (2009). Random networks with sublinear preferential attachment: degree evolutions. *Electron. J. Prob.* **14,** 1222–1267.

[8] Dereich, S. and Mörters, P. (2011). Random networks with concave preferential attachment rule. *Jahresber. Deutschen Math. Ver.* **113,** 21–40.

[9] Dereich, S. and Mörters, P. (2012). Random networks with sublinear preferential attachment: the giant component. To appear in *Ann. Prob.*.

[10] Dorogovtsev, S. N., Mendes, J. F. F. and Samukhin, A. N. (2003). Metric strucure of random networks. *Nuclear Phys. B* **653,** 307–338.
[11] Mönch, C. (2012). Distances in preferential attachment networks. Doctoral Thesis, University of Bath.
[12] Norros, I. and Reittu, H. (2006). On a conditionally Poissonian graph process. *Adv. Appl. Prob.* **38,** 59–75.
[13] Norros, I. and Reittu, H. (2008). Network models with a 'soft hierarchy': a random graph construction with loglog scalability. *IEEE Network* **22,** 40–46.
[14] Reittu, H. and Norros, I. (2002). On the effect of very large nodes in internet graphs. In *GLOBECOM '02*, Vol. III (Proc. Global Telecommun. Conf., Taipei, 2002), pp. 2624–2628.
[15] Resnick, S. I. (2008). *Extreme Values, Regular Variation and Point Processes*. Springer, New York.
[16] Van der Hofstad, R. (2010). Random graphs and complex networks. Lecture Notes, Eindhoven University of Technology.
[17] Van der Hofstad, R., Hooghiemstra, G. and Znamenski, D. (2007). Distances in random graphs with finite mean and infinite variance degrees. *Electron. J. Prob.* **12,** 703–766.
[18] Van der Hofstad, R. and Hooghiemstra, G. (2008). Universality for distances in power-law random graphs. *J. Math. Phys.* **49,** 125209, 14 pp.