A Geometric Preferential Attachment Model of Networks II

Abraham D. Flaxman[*], Alan M. Frieze[†], Juan Vera[‡]

September 20, 2007

Abstract

We study a random graph G_n that combines certain aspects of geometric random graphs and preferential attachment graphs. This model yields a graph with power-law degree distribution where the expansion property depends on a tunable parameter of the model.

The vertices of G_n are n sequentially generated points $x_1, x_2, \ldots, x_n$ chosen uniformly at random from the unit sphere in $\boldsymbol{R}^3$. After generating x_t, we randomly connect it to m points from those points in $x_1, x_2, \ldots, x_{t-1}$.

1 Introduction

During the last decade a large body of research has centered on understanding and modeling the structure of large-scale networks like the Internet and the World Wide Web. Several recent books provide a general introduction to this topic [38, 41]. One important feature identified in early experimental studies (including [4, 13, 23]) is that the vertex degree distribution of many real-world networks has a heavy-tailed property, which may follow a power-law (i.e., the proportion of vertices of degree at least k is proportional to $k^{-\alpha}$ for some constant α). This has driven the investigation of random graph distributions which generate heavy-tailed degree distributions, including the fixed degree sequence model, the copying model, and the preferential attachment model.

The preferential attachment model and its derivatives have been particularly popular for theoretical analysis. Preferential attachment was proposed as a model for real-world complex networks by Barabási and Albert [5]. The distribution was formalized by Bollobás and Riordan [10], and in [12] it was proved rigorously that **whp** a graph chosen according to this distribution has a power-law degree distribution with complementary cumulative distribution function (ccdf) $\mathbf{Pr}[\deg(v) \geq k] = \Theta(k^{-2})$. By changing the initial attactiveness or incorporating more random addition and deletion, the power of the ccdf power-law can be tuned to take any value in the interval $(1, \infty)$ [14, 18].

However, there are some significant differences between graphs generated by preferential attachment and those found in the real world. One major difference is found in

[*]Microsoft Research

[†]Carnegie Mellon University; supported in part by NSF grant CCF-200945.

[‡]Georgia Institute of Technology

their expansion properties. Mihail, Papadimitriou, and Saberi [36] showed that **whp** the preferential attachment model has conductance bounded below by a constant. On the other hand, Blandford, Blelloch and Kash [8] found that some WWW related graphs have smaller separators than the preferential attachment model predicts. This observation is consistient with observations due to Estrada [20], who found that half of the real-world networks he looked at were good expanders and the other half were not so good. The perturbed random graph framework provides one approach to understanding expansion in real-world networks [24], but it does not give a generative procedure. This paper investigates a generative procedure, based on a geometric modification of the preferential attachment model, which yields a graph that might or might not be a good expander, depending on a tunable parameter of the geometry. This is a strict generalization of the geometric preferential attachment graph developed in [25] which was designed specifically to avoid being a good expander.

The primary contribution of this paper is to provide a parameterised model that exhibits a sharp transition between low and high conductance. Choosing this parameter appropriately provides a unified approach to generating preferential attachment graphs with and without good expansion processes.

1.1 The random process

In [25] we studied a process which generates a sequence of graphs $G_t, t = 1, 2, \ldots, n$. The graph $G_t = (V_t, E_t)$ has t vertices and mt edges. Here V_t is a subset of t random points on S, the surface of the sphere in $\boldsymbol{R}^3$ of radius $\frac{1}{2\sqrt{\pi}}$ (so that $area(S) = 1$). After randomly choosing $x_{t+1} \in S$, it is connected, by preferential attachment (i.e. proportional to degree), to m vertices in V_t among those of distance at most r from x_{t+1}. We showed that this graph has a power law degree distribution, small seperators and a moderate diameter. In this paper we provide a "smoothed" version of this model, instead of choosing proportional to degree among those vertices within distance r of x_{t+1}, the m neighbors of x_t are chosen proportional to degree and some function of the distance to x_{t+1}.

Let $F : \mathbf{R}_+ \to \mathbf{R}_+$. Define

$$I = \int_S F(|u - u_0|)du = \frac{1}{2}\int_{x=0}^{\pi} F(x)\sin x dx$$

where u_0 is any point in S and $0 \leq |u - u_0| \leq \pi$ is the angular distance from u to u_0 along a great circle. Other parameters of the process are $m > 0$ the number of edges added in every step and $\alpha \geq 0$ a measure of the bias towards self loops.

- **Time step** 0**:** To initialize the process, we start with G_0 being the Empty Graph.
- **Time step** $t+1$**:** We choose vertex x_{t+1} uniformly at random in S and add it to G_t. Let
$$T_t(x_{t+1}) = \sum_{v \in V_t} F(|x_{t+1} - v|) \deg_t(v).$$

We add m random edges (x_{t+1}, y_i), $i = 1, 2, \ldots, m$ incident with x_{t+1}. Here, each y_i is chosen independently from $V_{t+1} = V_t \cup \{x_{t+1}\}$ (parallel edges and loops are permitted), such that for each $i = 1, \ldots, m$, for all $v \in V_t$,

$$\mathbf{Pr}(y_i = v) = \frac{\deg_t(v) F(|x_{t+1} - v|)}{\max\left(T_t(x_{t+1}), \alpha m I t\right)}$$

and

$$\mathbf{Pr}(y_i = x_{t+1}) = 1 - \frac{T_t(x_{t+1})}{\max\left(T_t(x_{t+1}), \alpha m I t\right)}$$

(When $t = 0$ we have $\mathbf{Pr}(y_i = x_1) = 1$.)

For $z > 0$ we define

$$I_z = \frac{1}{2} \int_{x=0}^{z} F(x) \sin x \, dx \ \textit{and}\ J_z = I - I_z.$$

Where possible we will illustrate our theorems using the canonical functions:

$$\begin{aligned} F_0(u) &= 1_{|u| \le r}, & & r \ge n^{\epsilon - 1/2}. \\ F_1(u) &= \frac{1}{\max\{n^{-\delta}, u\}^{\beta}} & & \textit{where } \delta < 1/2. \\ F_2(u) &= e^{-\beta u} & & \beta = \beta(n) \ge 0. \end{aligned}$$

Notice that F_0 corresponds to the model presented in [25]. Also notice that without the $n^{-\delta}$ term in the definition of F_1 for $\beta \ge 2$ we would have $I = \infty$. One can justify its inclusion (for some value of δ) from the fact that **whp** the minimum distance between the points in V_n is greater than $1/n \ln n$.

Observe that

$$\begin{aligned} I_z(F_0) &= \frac{1}{2}(1 - \cos(\min\{z, r\})). \\ I_z(F_1) &= \begin{cases} \frac{\beta n^{\delta(\beta-2)}}{4(\beta-2)} + O(n^{(\beta-4)\delta} + z^{2-\beta}) & z \ge n^{-\delta},\ \beta > 2. \\ \Theta(z^{2-\beta}) + O(n^{(\beta-2)\delta}) & z \ge n^{-\delta},\ \beta < 2 \\ \ln(n^{\delta} z) + O(1) & z \ge n^{-\delta},\ \beta = 2 \end{cases} \\ I_z(F_2) &= \frac{1}{2(1+\beta^2)}(1 - e^{-\beta z}(\cos z + \beta \sin z)). \end{aligned}$$

Let $d_k(t)$ denote the number of vertices of degree k at time t and let $\overline{d}_k(t)$ denote the expectation of $d_k(t)$.

We will first prove the following result about the degree distribution and the existence of small separators:

Theorem 1

(a) *Suppose that $\alpha > 2$ and in addition that*

$$\int_{x=0}^{\pi} F(x)^2 \sin x dx = O(n^\theta I^2) \tag{1}$$

where $\theta < 1$ is a constant.

Then there exists a constant $\gamma_1 > 0$ such that for all $k = k(n) \geq m$,

$$\overline{d}_k(n) = e^{\varphi_k(m,\alpha)} \left(\frac{m}{k}\right)^{1+\alpha} n + O(n^{1-\gamma_1}) \tag{2}$$

where $\varphi_k(m, \alpha) = O(1)$ tends to a constant $\varphi_\infty(m, \alpha)$ as $k \to \infty$.

Furthermore, for n sufficiently large, the random variable $d_k(n)$ satisfies the following concentration inequality:

$$\mathbf{Pr}(|d_k(n) - \overline{d}_k(n)| \geq I^2 n^{\max\{1/2, 2/\alpha\}+\delta}) \leq e^{-n^\delta}. \tag{3}$$

(b) *Suppose that $\alpha > 0$ and $m_0 \leq m$ where m_0 is a sufficiently large constant and $\varphi, \eta = o(1)$ are such that $\eta n \to \infty$ and $J_\eta \leq \varphi I$. Then* **whp***, V_n can be partitioned into $T, \bar{T}$ such that $|T|, |\bar{T}| \sim n/2$, and there are $\tilde{O}((\eta + \varphi)mn)$ edges between T and $\bar{T}$.*

Remark 1 *Note that the exponent in (a) does not depend on the particular function F. F manifests itself only through the error terms.*

For Part (a) of the above theorem:

$F = F_0$**:** $\theta = 1 - 2\epsilon$.

$F = F_1$, $\beta > 2$**:** $\theta = 2\delta$.

$F = F_1$, $\beta < 2$**:** $\theta = 0$.

$F = F_1$, $\beta = 2$**:** $\theta = 2\delta$.

$F = F_2$**:** $\theta = 0$.

For Part (b) of the above theorem:

$F = F_0$**:** $\eta = r$, $\varphi = 0$.

$F = F_1$, $\beta > 2$**:** $\eta = n^{-\delta/2}$, $\varphi = O(n^{-(\beta-2)\delta/2})$.

$F = F_1$, $\beta = 2$**:** $\eta = \frac{\ln \ln n}{\ln n}$, $\varphi = O(\eta)$.

We now consider the connectivity and diameter of G_n. For this we will place some more restrictions on F.

Define the parameter $\rho(\mu)$ by

$$I_\rho = \mu I. \tag{4}$$

As we will see in Theorem 3, $F = F_1, \beta < 2$ does not fit the hypotheses of part (b) of this theorem.

We will say that F is *smooth* (for some value of μ) if

(S1) F is monotone non-increasing.

(S2) $\rho^2 n \geq L \ln n$ for some sufficiently large constant L.

(S3) $\rho^2 F(2\rho) \geq c_3 I$ for some c_3 which is bounded below.

Theorem 2 *Suppose that $\alpha > 2$ and F is smooth for some constant $\mu > 0$ and $m \geq K \ln n$ for K sufficiently large. Then* **whp**

(a) *G_n is connected.*

(b) *G_n has diameter $O(\ln n/\rho)$.*

For the above theorem:

$F = F_0$**:** $I \sim r^2/4$ and so we can take $\mu \sim 1/4$, $\rho = r/2$, $c_3 \sim 1$.

$F = F_1$, $\beta > 2$**:** $I \sim \frac{n^{\delta(\beta-2)}}{2(\beta-2)}$ and so we can take $\mu \sim 1/4$, $\rho = n^{-\delta}/2$, $c_3 \sim (\beta - 2)/2$.

$F = F_1$, $\beta < 2$**:** $I = \Theta(1)$ and we can take $\rho = 1$, $\mu = \Omega(1)$, $c_3 = \Omega(1)$.

$F = F_2$**:** $I = \Theta(1)$ and we can take $\rho = 1$, $\mu = \Omega(1)$, $c_3 = \Omega(1)$.

We have a problem fitting the case of F_1 with $\beta = 2$ into the theorem.

We now consider conditons under which G_n is an expander.

Let F be *tame* if there exist absolute constants C_1, C_2 such that

(T1) $F(x) \geq C_1$ for $0 < x \leq \pi$.

(T2) $I \leq C_2$.

We note that F_1 with $\beta < 2$ is tame since $F_1(x) \geq \pi^{-\beta}$ for $0 \leq \pi$ and

$$I = \frac{1}{2} \int_{x=0}^{\pi} \sin x x^{-\beta} dx \leq \frac{\pi^{2-\beta}}{2(2-\beta)}.$$

The *conductance* Φ of G_n is defined by

$$\Phi = \min_{\deg_n(K) \leq mn} \Phi(K) = \min_{\deg_n(K) \leq mn} \frac{|E(K : \bar{K})|}{\deg_n(K)}.$$

Theorem 3 *If $\alpha > 2$ and F is tame and $m \geq K \ln n$ for K sufficiently large then* **whp**

(a) *G_n has conductance bounded from below by a constant.*

(b) *G_n is connected.*

(c) *G_n has diameter $O(\log_m n)$.*

Mihail et al [32] have some empirical results on the conductance of G_n in the case where $F = F_1$. They observe poor conductance when $\beta < 2$ and good conductance when $\beta > 2$. This fits nicely with the results of Theorems 2 and 3.

The role of α: This parameter was introduced in [25] as a means of overcoming a difficult technical problem. When $\alpha > 2$ it facilitates a proof of Lemma 2. On the positive side, it does give a parameter that effects the power law. On the negative side, when $\alpha > 2$, there will **whp** be isolated vertices, unless we make m grow at least as fast as $\ln n$. It is for us, an interesting open question, as to how to prove our results with $\alpha = 0$.

2 Outline of the paper

We prove a likely power law for the degree sequence in Section 3. We follow a standard practise and prove a recurrence for the expected number of vertices of degree k at time step t. Unfortunately, this involves the estimation of the expectation of the reciprocal of a random variable and to handle this, we show that this random variable is concentrated. This is quite technical and is done in Section 3.2. In Section 4 we show that under the assumptions of Theorem 1(b) there are small separators. This is relatively easy, since any give great circle can **whp** be used to define a small separator.

Section 5 proves connectivity when m grows logarithmically with n. The idea is to show that **whp** the sub-graph $G_n(B)$ induced by a ball B of radius ρ, centered in $u \in S$, is connected. and has small diameter. We then show that the union of the $G_n(B)$'s for $u = x_1, x_2, \ldots, x_n$ is connected and has small diameter.

Section 6 deals with the case of tame functions.

3 Proving a power law

3.1 Establishing a recurrence for $\overline{d}_k(t)$: the expected number of vertices of degree k at time t

Our approach to proving Theorem 1(a) is to find a recurrence for $\overline{d}_k(t)$. For $k \in \mathbf{N}$ define $D_k(t) = \{v \in V_t : \deg_t(v) = k\}$. Thus $d_k(t) = |D_k(t)|$. Also, define $d_{m-1}(t) = 0$ and $\overline{d}_{m-1}(t) = 0$ for all integers t with $t > 0$. Let $\eta_k(G_t, x_{t+1})$ denote the (conditional) probability that a parallel edge from x_{t+1} to a vertex of degree no more than k is created at time $t + 1$. Then,

$$\eta_k(G_t, x_{t+1}) = O\left(\min\left\{\sum_{i=m}^{k} \sum_{v \in D_i(t)} \frac{F(|x_{t+1} - v|)^2\, i^2}{\max\{\alpha m I t, T_t(x_{t+1})\}^2}, 1\right\}\right). \tag{5}$$

Then for $k \geq m$,

$$\begin{aligned} \mathbf{E}\left[d_k(t+1) \mid G_t, x_{t+1}\right] &= d_k(t) \\ &+ m \sum_{v \in D_{k-1}(t)} \frac{(k-1)F(|x_{t+1} - v|)}{\max\{\alpha m I t, T_t(x_{t+1})\}} - m \sum_{v \in D_k(t)} \frac{kF(|x_{t+1} - v|)}{\max\{\alpha m I t, T_t(x_{t+1})\}} \\ &+ \mathbf{Pr}\left[deg_{t+1}(x_{t+1} = k) \mid G_t, x_{t+1}\right] + O(m\eta_k(G_t, x_{t+1})). \end{aligned} \tag{6}$$

Let $\mathcal{A}_t$ be the event

$$\{|T_t(x_{t+1}) - 2mIt| \leq C_1 I m t^{\gamma} \ln n\}$$

where

$$\max\{2/\alpha, \theta\} < \gamma < 1$$

and C_1 is some sufficiently large constant.

Note that if

$$t \geq t_0 = (\ln n)^{2/(1-\gamma)} \tag{7}$$

then

$$\mathcal{A}_t \text{ implies } T_t(x_{t+1}) \leq \alpha m I t.$$

Then, for $t \geq t_0$,

$$\begin{aligned}
&\mathbf{E}\left[\sum_{v\in D_k(t)} \frac{kF(|x_{t+1}-v|)}{\max\{\alpha m I t, T_t(x_{t+1})\}}\right] \\
&\quad = \mathbf{E}\left[\sum_{v\in D_k(t)} \frac{kF(|x_{t+1}-v|)}{\max\{\alpha m I t, T_t(x_{t+1})\}} \;\Bigg|\; \mathcal{A}_t\right] \mathbf{Pr}\left[\mathcal{A}_t\right] \\
&\qquad\qquad + \mathbf{E}\left[\sum_{v\in D_k(t)} \frac{kF(|x_{t+1}-v|)}{\max\{\alpha m I t, T_t(x_{t+1})\}} \;\Bigg|\; \neg\mathcal{A}_t\right] \mathbf{Pr}\left[\neg\mathcal{A}_t\right] \\
&\quad = \frac{k}{\alpha m t}\mathbf{E}\left[d_k(t)|\mathcal{A}_t\right]\mathbf{Pr}\left[\mathcal{A}_t\right] + O\left(1\right)\mathbf{Pr}\left[\neg\mathcal{A}_t\right] \\
&\quad = \frac{k\bar{d}_k(t)}{\alpha m t} - \frac{k}{\alpha m t}\mathbf{E}\left[d_k(t)|\neg\mathcal{A}_t\right]\mathbf{Pr}\left[\neg\mathcal{A}_t\right] + O\left(1\right)\mathbf{Pr}\left[\neg\mathcal{A}_t\right] \\
&\quad = \frac{k\bar{d}_k(t)}{\alpha m t} + O\left(k\right)\mathbf{Pr}\left[\neg\mathcal{A}_t\right]
\end{aligned}$$

In Lemma 2 below we prove that

$$\mathbf{Pr}\left[\neg\mathcal{A}_t\right] = O\left(n^{-2}\right). \tag{8}$$

Thus, if $t \geq t_0$ then

$$\mathbf{E}\left[\sum_{v\in D_k(t)} \frac{kF(|x_{t+1}-v|)}{\max\{\alpha m I t, T_t(x_{t+1})\}}\right] = \frac{k\bar{d}_k(t)}{\alpha m t} + O\left(k/n^2\right). \tag{9}$$

In a similar way

$$\mathbf{E}\left[\sum_{v\in D_{k-1}(t)} \frac{(k-1)F(|x_{t+1}-v|)}{\max\{\alpha m I t, T_t(x_{t+1})\}}\right] = \frac{(k-1)\bar{d}_{k-1}(t)}{\alpha m t} + O\left(k/n^2\right). \tag{10}$$

On the other hand, given G_t, x_{t+1}, if

$$p = 1 - \frac{T_t(x_{t+1})}{\max\left(T_t(x_{t+1}), \alpha m I t\right)}$$

then

$$\mathbf{Pr}\left[\deg_{t+1}(x_{t+1}=k) \mid G_t, x_{t+1}\right] = \mathbf{Pr}\left[\mathrm{Bi}(m,p) = k-m\right]$$

So, if $t \geq t_0$,

$$\begin{aligned}
\mathbf{Pr}\left[\deg_{t+1}(x_{t+1}=k)\right] &= \binom{m}{k-m}\mathbf{E}\left[p^{k-m}(1-p)^{2m-k} \;\Big|\; \mathcal{A}_t\right]\mathbf{Pr}\left[\mathcal{A}_t\right] + O(\mathbf{Pr}\left[\neg\mathcal{A}_t\right]) \\
&= \binom{m}{k-m}\left(1-\frac{2}{\alpha}\right)^{k-m}\left(\frac{2}{\alpha}\right)^{2m-k}(1+O(mt^{\gamma-1}\ln n))\mathbf{Pr}\left[\mathcal{A}_t\right] + O(n^{-2}) \\
&= \binom{m}{k-m}\left(1-\frac{2}{\alpha}\right)^{k-m}\left(\frac{2}{\alpha}\right)^{2m-k} + O(mt^{\gamma-1}\ln n).
\end{aligned}$$

Now note that from equations (5) and (8) that if

$$t \geq t_1 = n^{(\gamma+\theta)/2\gamma}$$

and

$$k \leq k_0(t) = n^{(\gamma-\theta)/4}$$

then, from (1), we see that

$$\mathbf{E}(m\eta_k(G_t, x_{t+1})) = O\left(\frac{k^2 n^\theta}{mt}\right) = O(t^{\gamma-1}). \tag{11}$$

Taking expectations on both sides of (6) and using (9,10,11), we see that if $t \geq t_0$ and $k \leq k_0(t)$ then

$$\overline{d}_k(t+1) = \overline{d}_k(t) + \frac{k-1}{\alpha t}\overline{d}_{k-1}(t) - \frac{k}{\alpha t}\overline{d}_k(t) \\ + \binom{m}{k-m}\left(1-\frac{2}{\alpha}\right)^{k-m}\left(\frac{2}{\alpha}\right)^{2m-k} + O\left(mt^{\gamma-1}\ln n\right) \tag{12}$$

We consider the recurrence given by $f_{m-1} = 0$ and for $k \geq m$,

$$f_k = \frac{k-1}{\alpha}f_{k-1} - \frac{k}{\alpha}f_k + \binom{m}{k-m}\left(1-\frac{2}{\alpha}\right)^{k-m}\left(\frac{2}{\alpha}\right)^{2m-k}, \tag{13}$$

which, for $k > 2m$, has solution

$$\begin{aligned} f_k &= f_{2m}\prod_{i=2m+1}^{k}\frac{i-1}{i+\alpha} \\ &= f_{2m}e^{\varphi_k(m,\alpha)}\left(\frac{m}{k}\right)^{\alpha+1}. \end{aligned} \tag{14}$$

Here $\varphi_k(m,\alpha) = O(1)$ tends to a limit $\varphi_\infty(m,\alpha)$ depending only on m, α as $k \to \infty$. Furthermore, $\lim_{m\to\infty}\varphi_\infty(\alpha, m) = 0$. We also have

$$f_{m+i} = f_{2m}\prod_{j=i+1}^{m}\left(1+\frac{\alpha+1}{m+j-1}\right) \leq e^{2\alpha+3}f_{2m}.$$

It follows that (14) is also valid for $m \leq k \leq 2m$ with $\varphi_k(m,\alpha) = O(1)$.

We finish the proof of (2) by showing that there exists a constant $M > 0$ such that

$$|\overline{d}_k(t) - f_k t| \leq M(t_1 + mt^\gamma \ln n) \tag{15}$$

for all $0 \leq t \leq n$ and $m \leq k \leq k_0(t)$.

We have that (15) is trivially true for $t < t_1$, and for $t \geq t_1$ and $k > k_0(t)$ it follows from $\overline{d}_k(t) \leq 2mt/k$.

Now, let $\Theta_k(t) = \overline{d}_k(t) - f_k t$. Then for $t \geq t_1$ and $m \leq k \leq k_0(t)$,

$$\Theta_k(t+1) = \frac{k-1}{\alpha t}\Theta_{k-1}(t) - \frac{k}{\alpha t}\Theta_k(t) + O(mt^{\gamma-1}\ln n). \tag{16}$$

Let L denote the hidden constant in $O(mt^{\gamma-1}\ln n)$ of (16). Our inductive hypothesis $\mathcal{H}_t$ is that

$$|\Theta_k(t)| \leq M(t_1 + mt^\gamma \ln n)$$

for every $m \leq k \leq k_0(t)$ and M sufficiently large. Assume that $t \geq t_1$. Then $k \ll t$ in the current range of interest, and so from (16),

$$\begin{aligned} |\Theta_k(t+1)| &\leq M(t_1 + mt^\gamma \ln n) + Lmt^{\gamma-1}\ln n \\ &\leq M(t_1 + m(t+1)^\gamma \ln n). \end{aligned}$$

This verifies $\mathcal{H}_{t+1}$ and completes the proof by induction.

3.2 Concentration of $T_t(u)$

Now we turn our attention to prove that $T_t(u)$ is concentrated around its mean.

Lemma 1 *Let $u \in S$ and $t > 0$ then* $\mathbf{E}[T_t(u)] = 2Imt$

Proof

$$\mathbf{E}[T_t(u)] = \mathbf{E}\left[\sum_{v\in V_t} \deg_t(v)F(|u-v|)\right] = I\sum_{v\in V_t}\deg_t(v) = 2Imt$$

□

Lemma 2 *If $t > 0$ and u is chosen randomly from S then*

$$\mathbf{Pr}\left[|T_t(u) - 2Imt| \geq mI(t^{2/\alpha} + t^{1/2}\ln t)\ln n\right] = O\left(n^{-2}\right).$$

Proof We use Azuma-Hoeffding inequality (see for example [2]). One may be a little concerned here that our probability space is not discrete. Although it is not really necessary, one could replace S by 2^{2^n} randomly chosen points X and sample uniformly from these. Then **whp** the change in distribution would be negligible. With this reassurance, fix τ, with $1 \leq \tau < t$. Fix G_τ and let $G_t = G_t(G_\tau, x_{\tau+1}, y_1, \ldots, y_m)$ and $\hat{G}_t = G_t(G_\tau, \hat{x}_{\tau+1}, \hat{y}_1, \ldots, \hat{y}_m)$, where $x_{\tau+1}, \hat{x}_{\tau+1} \in S$ and $y_1, \ldots, y_m, \hat{y}_1, \ldots, \hat{y}_m \in V_\tau$. We couple the construction of G_t and $\hat{G}_t$, starting at time step $\tau+1$ with the graph G_τ and $\hat{G}_\tau$ respectively. Then, for every step $\sigma > \tau + 1$ we choose the same point $x_\sigma \in S$ in both and for every $i = 1, \ldots, m$ we choose $u_i, \hat{u}_i \in V_\sigma$ such that each marginal is the correct marginal and such that the probability of choosing the same vertex is maximized.

Notice that we have

$$\mathbf{Pr}[u_i = v = \hat{u}_i] = \min\left(\frac{\deg_{G_{\sigma-1}}(v)F(|v-x_\sigma|)}{\max(T_{\sigma-1}(x_\sigma), \alpha mI(\sigma-1))}, \frac{\deg_{\hat{G}_{\sigma-1}}(v)F(|v-x_\sigma|)}{\max\left(\hat{T}_{\sigma-1}(x_\sigma), \alpha mI(\sigma-1)\right)}\right)$$

for every $v \in V_{\sigma-1}$. Also,

$$\mathbf{Pr}[u_i = x_\sigma = \hat{u}_i] = 1 - \max\left(\frac{T_{\sigma-1}(x_\sigma)}{\max(T_{\sigma-1}(x_\sigma), \alpha mI(\sigma-1))}, \frac{\hat{T}_{\sigma-1}(x_\sigma)}{\max\left(\hat{T}_{\sigma-1}(x_\sigma), \alpha mI(\sigma-1)\right)}\right)$$

Now, for $u \in S$ let

$$\Delta_\sigma(u) := \Delta_{\sigma,\tau}(u) = \sum_{\rho=\tau}^{\sigma} \sum_{i=1}^{m} |F(|u - u_i^\rho|) - F(|u - \hat{u}_i^\rho|)|.$$

Lemma 3 *Let $t \geq 1$ and let u be a random point in S. Then for some constant $C > 0$,*

$$\mathbf{E}\left[\Delta_t(u)\right] \leq CmI \left(\frac{t}{\tau}\right)^{2/\alpha}.$$

Proof We begin with

$$\mathbf{E}\left[|F(|w - u_i^\rho|) - F(|w - \hat{u}_i^\rho|)|\big|u_i^j, \hat{u}_i^j : i = 1, \ldots, m,\ j = 1, \ldots, \sigma\right] \leq 2I 1_{u_i^\rho \neq \hat{u}_i^\rho}.$$

Therefore if we define for every $\tau < \sigma \leq t$

$$\Delta_\sigma = \sum_{\rho=\tau}^{\sigma} \sum_{i=1}^{m} 1_{u_i^\sigma \neq \hat{u}_i^\sigma},$$

we have

$$\mathbf{E}\left[\Delta_\sigma(u)\right] \leq 2I\mathbf{E}\left[\Delta_\sigma\right].$$

Fix $\tau < \sigma \leq t$. We have then

$$\Delta_\sigma = \Delta_{\sigma-1} + \sum_{i=1}^{m} 1_{u_i^\sigma \neq \hat{u}_i^\sigma}. \tag{17}$$

Now fix $1 \leq i \leq m$. Taking expectations with respect to our coupling,

$$\begin{aligned}
&\mathbf{E}\left[1_{u_i^\sigma \neq \hat{u}_i^\sigma} | G_{\sigma-1}, \hat{G}_{\sigma-1}, x_\sigma\right] = 1 - \mathbf{Pr}\left[u_i^\sigma = \hat{u}_i^\sigma | G_{\sigma-1}, \hat{G}_{\sigma-1}, x_\sigma\right] \\
&= \max\left(\frac{T_{\sigma-1}(x_\sigma)}{\max\left(T_{\sigma-1}(x_\sigma), \alpha mI(\sigma-1)\right)}, \frac{\hat{T}_{\sigma-1}(x_\sigma)}{\max\left(\hat{T}_{\sigma-1}(x_\sigma), \alpha mI(\sigma-1)\right)}\right) \\
&\quad - \sum_{v \in V_{\sigma-1}} \min\left(\frac{\deg_{G_{\sigma-1}}(v) F(|v - x_\sigma|)}{\max\left(T_{\sigma-1}(x_\sigma), \alpha mI(\sigma-1)\right)}, \frac{\deg_{\hat{G}_{\sigma-1}}(v) F(|v - x_\sigma|)}{\max\left(\hat{T}_{\sigma-1}(x_\sigma), \alpha mI(\sigma-1)\right)}\right) \\
&\leq \frac{\max\left(T_{\sigma-1}(x_\sigma), \hat{T}_{\sigma-1}(x_\sigma)\right) - \sum_{v \in V_{\sigma-1}} \min\left(\deg_{G_{\sigma-1}}(v), \deg_{\hat{G}_{\sigma-1}}(v)\right) F(|v - x_\sigma|)}{\max\left(T_{\sigma-1}(x_\sigma), \hat{T}_{\sigma-1}(x_\sigma), \alpha mI(\sigma-1)\right)} \qquad (18) \\
&\leq \frac{\sum_{v \in V_{\sigma-1}} |\deg_{G_{\sigma-1}}(v) - \deg_{\hat{G}_{\sigma-1}}(v)| F(|v - x_\sigma|)}{\max\left(T_{\sigma-1}(x_\sigma), \hat{T}_{\sigma-1}(x_\sigma), \alpha mI(\sigma-1)\right)} \qquad (19) \\
&\leq \frac{\sum_{v \in V_{\sigma-1}} |\deg_{G_{\sigma-1}}(v) - \deg_{\hat{G}_{\sigma-1}}(v)| F(|v - x_\sigma|)}{\alpha mI(\sigma-1)}
\end{aligned}$$

Inequality (18), follows from

$$\max\left(\frac{a}{\max(a,c)},\frac{b}{\max(b,c)}\right)=\frac{\max(a,b)}{\max(a,b,c)}$$

and

$$\min\left(\frac{a}{b},\frac{c}{d}\right)\geq\frac{\min(a,c)}{\max(b,d)}.$$

Inequality (19) is a consequence of $\max\{\sum_i a_i,\sum_i b_i\}-\sum_i \min\{a_i,b_i\}\leq\sum_i |a_i-b_i|$.

Therefore

$$\mathbf{E}\left[\Delta_\sigma \,\middle|\, G_{\sigma-1},\hat{G}_{\sigma-1}\right]\leq\Delta_{\sigma-1}+\frac{\sum_{v\in V_{\sigma-1}}|\deg_{G_{\sigma-1}}(v)-\deg_{\hat{G}_{\sigma-1}}(v)|}{\alpha(\sigma-1)}. \tag{20}$$

But, for each $v\in V_{\sigma-1}$ we have

$$|\deg_{G_{\sigma-1}}(v)-\deg_{\hat{G}_{\sigma-1}}(v)|\leq\sum_{j=\tau}^{\sigma-1}\sum_{i=1}^{m}(1_{u_i^j=v,\,\hat{u}_i^j\neq v}+1_{u_i^j\neq v,\,\hat{u}_i^j=v})$$

and thus

$$\sum_{v\in V_{\sigma-1}}|\deg_{G_{\sigma-1}}(v)-\deg_{\hat{G}_{\sigma-1}}(v)|\leq\sum_{j=\tau}^{\sigma-1}\sum_{i=1}^{m}\sum_{v\in V_{\sigma-1}}\left(1_{u_i^j=v,\,\hat{u}_i^j\neq v}+1_{u_i^j\neq v,\,\hat{u}_i^j=v}\right)\leq 2\Delta_{\sigma-1}.$$

Going back to (20) we have

$$\mathbf{E}[\Delta_\sigma]\leq\mathbf{E}[\Delta_{\sigma-1}]\left(1+\frac{2}{\alpha(\sigma-1)}\right),$$

so, $\mathbf{E}[\Delta_t]\leq e^{O(1)}\left(\frac{t}{\tau}\right)^{2/\alpha}\mathbf{E}[\Delta_\tau]$. Now, $\Delta_\tau\leq m$, because the graphs G_τ and $\hat{G}_\tau$ differ at most in the last m edges. Therefore $\mathbf{E}[\Delta_t]\leq e^{O(1)}m\left(\frac{t}{\tau}\right)^{2/\alpha}$. □

To apply Azuma's inequality we note first that

$$\left|\mathbf{E}_{G_t}[T_t(u)]-\mathbf{E}_{\hat{G}_t}[T_t(u)]\right|=\left|\mathbf{E}\left[\sum_{\rho=\tau}^{t}\sum_{i=1}^{m}(F(|u-u_i^\rho|)-F(|u-\hat{u}_i^\rho|))\right]\right|\leq\mathbf{E}[\Delta_t(u)], \tag{21}$$

and from Lemma 3

$$\sum_{\tau=1}^{t}\mathbf{E}[\Delta_t(u)]^2\leq(e^{O(1)}mI)^2t^{4/\alpha}\sum_{\tau=1}^{t}\tau^{-4/\alpha}=O\left(I^2m^2(t\ln t+t^{4/\alpha})\right)$$

Therefore, there is C_1 such that

$$\mathbf{Pr}\left[|T_t(u)-\mathbf{E}[T_t(u)]|\geq C_1Im(t^{2/\alpha}+t^{1/2}\ln t)(\ln n)^{1/2}\right]\leq e^{-2\ln n}=n^{-2}.\square$$

3.3 Concentration of $d_k(t)$

We follow the proof of Lemma 3, replacing $T_t(u)$ by $d_k(t)$ and using the same coupling, When we reach (21) we find that $\left|\mathbf{E}_{G_t}[d_k(t)] - \mathbf{E}_{\hat{G}_t}[d_k(t)]\right| \le 2\mathbf{E}[\Delta_t]$, the rest is the same.

This proves (1) and completes the proof of Theorem 1(a) .

4 Small separators

In this section we prove Theorem 1(b). For this, we assume $\alpha > 0$ and $m_0 \le m$ where m_0 is a sufficiently large constant and $\varphi, \eta = o(1)$ are such that $\eta n \to \infty$ and $J_\eta \le \varphi I$.

We use the geometry of the instance to obtain a sparse cut. Consider partitioning the vertices in V_n using a great circle of S. This will divide V_n into sets T and $\bar{T}$ which each contain about $n/2$ vertices. More precisely, we have

$$\mathbf{Pr}\left[|T| < (1-\xi)n/2\right] = \mathbf{Pr}\left[|\bar{T}| < (1-\xi)n/2\right] \le e^{-\xi^2 n/5}.$$

To bound $e(T, \bar{T})$, the number of edges crossing the cut, we divide the edges into two types. We call an edge $\{u, v\}$ in G_n long if $|u - v| \ge \eta$, otherwise we call it short. We will show that **whp** the number of long edges is small, and therefore we just need to consider short edges in a cut. Let Z denote the number of long edges. Then

$$\begin{aligned}
\mathbf{E}[Z] &\le mt_0 + m \sum_{t \ge t_0} \sum_{v \in V_t} \frac{\deg_t(v) J_\eta}{\alpha m I t} \\
&\le mt_0 + m \sum_{t \ge t_0} \frac{J_\eta}{\alpha I} \\
&= mt_0 + O(mn\varphi).
\end{aligned}$$

□

Now **whp** there are at most $\mathbf{E}[Z]/\varphi^{1/2}$ long edges. Apart from these, edges only appear between vertices within distance η, so only edges incident with vertices appearing in the strip within distance η of the great circle can appear in the cut. Since $\eta = o(1)$, this strip has area less than $3\eta\sqrt{\pi}$, and, letting U denote the vertices appearing in this strip, we have

$$\mathbf{Pr}\left[|U| \ge 4\sqrt{\pi}\eta n\right] \le e^{-\sqrt{\pi}\eta n/9} = o(1).$$

Even if every one of the vertices chooses its m neighbors on the opposite side of the cut, this will yield at most $4\sqrt{\pi}\eta n m$ edges **whp**. So the graph has a cut with

$$e(T, \bar{T}) = \tilde{O}((\eta + \varphi^{1/2})mn)$$

with probability at least $1 - o(1)$.

5 Connectivity and Diameter

Here we prove Theorem 2. Let μ be such that F is smooth for μ, and let $\rho = \rho(\mu)$. Fix $u \in S$ let

$$B_\rho = \{v \in S : \ |v - u| \le \rho\}$$

and let $A_\rho = \int_{v \in B_\rho} dv \in [c_1\rho^2, c_2\rho^2]$ denote the area of B_r. Here c_1, c_2 are some absolute constants, independent of ρ.

We denote the diameter of G by $\text{diam}(G)$, and follow the convention of defining $\text{diam}(G) = \infty$, when G is disconnected. In particular, when we say that a graph has finite diameter this implies it is connected.

Let

$$T = \frac{K_1 \ln n}{A_\rho} \leq \frac{K_1 n}{c_1 L}$$

where K_1 is sufficiently large, and $L^{2/3} \ll K_1 \ll K, L$.

Lemma 4

$$\mathbf{Pr}\left[\text{diam}(G_n(B_\rho)) \geq 2(K_1+1)\ln n\right] = O(n^{-1})$$

where $G_n(B_\rho)$ is the induced subgraph of G_n in B_ρ.

Proof Let $N = |G_n(B_\rho)|$ and let $V(G_n(B_\rho)) = \{x_{t_1}, \ldots, x_{t_N}\}$, where $t_s < t_{s+1}$ for all $s < N$ and $t_N \leq n$. For $s = 1, \ldots, N$ let $H_s = G_{t_s}(B_\rho)$. We concentrate our attention to the evolution of H_s.

Notice that s, is the number of steps for which $x_t \in B_\rho$ with $t \leq t_s$, and so $s \sim \text{Bi}(t_s, A_\rho)$. By the Chernoff bound we have that if $t_s \geq T$,

$$\mathbf{Pr}\left[\frac{1}{2} < \frac{t_s A_\rho}{s} < \frac{3}{2}\right] \geq 1 - n^{-K_1/13}.$$

Therefore, if N_0 is the number of vertices in B_ρ at time T, we may assume for all $s \geq N_0$, $s/2 < t_s A_\rho < 3s/2$. In particular, $N \geq 2nA_\rho/3 \geq c_1 L \ln n/2$ and $N_0 \leq 2TA_\rho \leq 2K_1 \ln n$.

Let X_s be the number of connected components of H_s. Then

$$X_{s+1} = X_s - Y_s + 1, \qquad X_0 = 0$$

where $Y_s \geq 0$ is the number of components conected to x_{t_s}.

B_ρ is contained in $B_{2\rho}(x_{t_s})$ the ball of radius 2ρ centered at x_{t_s}. Therefore if $v \in B_\rho \cap V_{t_s}$ and $t_s > T$,

$$\mathbf{Pr}\left[x_{t_s} \text{ chooses } v\right] \geq \frac{\deg_{t_s}(v) F(|x_{t_s} - v|)}{\alpha m I t_s} \geq \frac{F(2\rho)}{\alpha I t_s} \geq \frac{2A_\rho F(2\rho)}{3\alpha I s} \geq \frac{2c_1\rho^2 F(2\rho)}{3\alpha I s} \geq \frac{2c_1c_3}{3\alpha s}.$$

Now, we can bound the probability of generating a new component,

$$\mathbf{Pr}\left[Y_s = 0 | H_{s-1}\right] = \left(1 - \sum_{v \in H_{s-1}} \mathbf{Pr}\left[x_{t_s} \text{ chooses } v\right]\right)^m$$

$$\leq \left(1 - \frac{2c_1c_3}{3\alpha}\right)^m \leq \exp\left(-\frac{2c_1c_3 m}{3\alpha}\right) \leq n^{-10}$$

If $s < 2K_1 \ln n$, as $m \geq K \ln n$, we can bound the probability of not collapsing components,

$$\begin{aligned}\mathbf{Pr}\left[Y_s = 1 | X_s \geq 2\right] &\leq \mathbf{Pr}\left[Y_s = 1 | X_s \geq 2,\, Y_s > 0\right] + \mathbf{Pr}\left[Y_s = 0 | X_s \geq 2\right] \\ &\leq 2\left(1 - \frac{2c_1c_3}{3\alpha s}\right)^m + n^{-10} \\ &\leq 2\exp\left(-\frac{2mc_1c_3}{3\alpha s}\right) + n^{-10} \leq 1/10\end{aligned}$$

Therefore, X_s is stochastically dominated by the random variable $\max\{1, N_0 - Z_s\}$ where $Z_s \sim \mathrm{Bi}(s, 9/10)$. We then have

$$\mathbf{Pr}\left[X_{4K_1 \ln n} > 1\right] \leq \mathbf{Pr}\left[Z_{4K_1 \ln n} < N_0\right] \leq \mathbf{Pr}\left[Z_{4K_1 \ln n} < 2K_1 \ln n\right] \leq n^{-3}.$$

And therefore

$$\mathbf{Pr}\left[H_{4K_1 \ln n} \text{ is not connected}\right] \leq n^{-3}.$$

Now, to obtain an upper bound on the diameter, we run the process of construction of H_N by rounds. The first round consists of $4K_1 \ln n$ steps and in each new round we double the size of the graph, i.e. it consists of as many steps as the total number of steps of all the previous rounds. Notice that we have less than $\log_2 n$ rounds in total. Let $\mathcal{A}$ be the event for all $i > 0$ every vertex created in the $(i+1)$th round is adjacent to a vertex in $H_{2^{i+1}K_1 \ln n}$, the graph at the end of the ith round.

On the event $\mathcal{A}$, every vertex in H_N is at distance at most $\log -2n$ of $H_{2K_1 \ln n}$ whose diameter is not greater than $2K_1 \ln n$. Thus, the diameter of H_N is smaller than $2(K_1 + 2) \ln n$.

Now, if v is created in the $(i+1)$st round,

$$\mathbf{Pr}\left[v \text{ is not adjacent to } H_{2^{i-1}K_1 \ln n}\right] \leq \left(1 - \frac{2c_1c_3}{3\alpha}\right)^m.$$

Therefore

$$\mathbf{Pr}\left[\neg\mathcal{A}\right] \leq \left(1 - \frac{2c_1c_3}{3\alpha}\right)^m n(\ln n) \leq n^{1+o(1)-2Kc_1c_3/(3\alpha)}.$$

□

To finish the proof of connectivity and the diameter, let u, v be two vertices of G_n. Let $C_1, C_2, \ldots, C_M$, $M = O(1/\rho)$ be a sequence of spherical caps of radius ρ such that u is the center of C_1, v is the center of C_M and such that the centers of C_i, C_{i+1} are distance $\leq \rho/2$ apart. The intersections of C_i, C_{i+1} have area at least $A_\rho/10$ and so **whp** each intersection contains a vertex. Using Lemma 4 we deduce that **whp** there is a path from u to v in G_n of size at most $O(\ln n/\rho)$.

6 Proof of Theorem 3

For a set $K \subseteq V_n$ we define $\deg_n(K) = \sum_{v \in K} \deg_n(v)$.

Lemma 5 *There is an absolute constant* $0 < \xi < 1/4$ *such that*

$$\mathbf{Pr}(\exists K \subseteq V_n, |K| \geq (1-\xi)n:\ \deg_n(K) \leq (1+\xi)mn) = o(n^{-3}).$$

Proof Let ζ be a small positive constant and divide V_n into approximately $1/\zeta$ sets $S_1, S_2, \ldots$ of size $s = \lceil \zeta n \rceil$ plus a set of $n - \lfloor 1/\zeta \rfloor s$ where $S_i = \{x_{(i-1)s+1}, \ldots, x_{is}\}$, $i = 1, 2, \ldots,$. We put a high probability upper bound on $\deg_n(S_1)$. Now consider the random variables $\beta_k, k = 2, \ldots$ where $\beta_k = \deg_{\tau_k}(S_2 \cup \cdots S_k)/ms$ and $\tau_k = ks$. Now $\beta_2 \geq ms$ and conditional on the value of $\beta_k \geq (k-1)ms$

$$\beta_{k+1}ms \text{ dominates } ms + \beta_k ms + \text{Bi}\left(ms, \frac{\beta_k \lambda}{2(k+1)}\right)$$

where $\lambda = C_1/C_2$.

So, there exist constants γ_1, γ_2 (independent of ζ) such that

$$\mathbf{Pr}\left(\frac{\beta_{k+1}}{ms} \leq 1 + (1+\gamma_2)\frac{\beta_k}{ms}\right) \leq e^{-m\gamma_1 n}.$$

So, after some calculations, we find that with probability $1 - O(e^{-m\gamma_1 n})$,

$$\deg_n(V_n \setminus S_1) \geq ms(1+\gamma_2)\gamma_2^{-1}((1+\gamma_2)^{\lfloor 1/\zeta \rfloor - 3} - 1) \geq mn(1+\zeta/2)$$

for small enough ζ.

Now $\deg_n(S_1)$ dominates $\deg_n(L)$ for any set L of size $\lceil \zeta n \rceil$. So, if $m > 1/\gamma_1$ then the probability there is a set of size $\lceil \zeta n \rceil$ which has total degree exceeding $mn(1 - \gamma_2)$ is exponentially small ($\leq \binom{n}{\lceil \zeta n \rceil} e^{-n}$)). In which case, every set K of size at least $n - \lceil \zeta n \rceil$ has total degree $\deg_n(K) \geq mn(1+\gamma_2/2)$ and the lemma follows by taking $\xi = \min\{\zeta, \gamma_2/2, 1/4\}$. □

We have to estimate $\Phi(K)$ for all K with $\deg_n(K) \leq mn$. The above lemma shows that we can restrict our attention to sets K with $|K| \leq (1-\xi)n$.

We now observe that for $K \subseteq V_n$,

$$\deg_n(K) = m|K| + |E(K : \bar{K})|$$

and so to bound $\Phi(K)$, it suffices to prove lower bounds $|E(K : \bar{K})| \geq \eta m|K|$ for some positive constant η.

Lemma 6 *If* $m \geq C \ln n$ *where* C *is sufficiently large then there exists an absolute constant* $\kappa > 0$ *such that*

$$\mathbf{Pr}\left(\Phi(G_n) < \kappa\right) = O(n^{-3}).$$

Proof

6.1 $1 \le |K| \le A_0 n$.

Here A_0 is a sufficiently small constant. Let $K_1 = K \cap V_{n/2}$ and $K_2 = K \setminus K_1$. Let $W_1 = V_{n/2} \setminus K_1$ and $W_2 = V_n \setminus (V_{n/2} \cup K_2)$. The number of edges between K_1 and W_2 dominates $\mathrm{Bi}(m(n/2 - |K_2|), \lambda|K_1|/(\alpha n))$. This is because each edge chosen by $V_j, j \in W_2$ has probability at least $m\lambda|K_1|/(\alpha mn)$ of being in K_1. Similarly, the number of edges between K_2 and W_1 dominates $\mathrm{Bi}(m|K_2|, \lambda(n/2 - |K_1|)/(\alpha n))$. Thus $\mathbf{E}\left[|E(K : \bar{K})|\right] \ge m\lambda|K|/(3\alpha)$ and so by Hoeffdings inequality we see that $|E(K : \bar{K})| \ge m\lambda|K|/(4\alpha)$ with probability $1 - e^{-cm\lambda|K|}$ for some constant $c = c(\alpha)$. Thus

$$\mathbf{Pr}(\exists K,\, 1 \le |K| \le A_0 n,\, |E(K : \bar{K})| < m\lambda|K|/(4\alpha)) \le \sum_{k=1}^{A_0 n} \binom{n}{k} e^{-cC\lambda k \ln n} = o(1)$$

if $C \ge 2/(c\lambda)$.

6.2 $A_0 n \le |K| \le (1 - \xi)n$.

Here ξ is as in Lemma 5. Let K_1, K_2, W_1, W_2 be as in Section 6.1. Let $q = |K_1|$ and $r = |K_2|$. We calculate the expected number of edges $\mu(K_1, K_2)$ of $L = (K_2 \times W_1 \cup W_2 \times K_1)$ generated at steps τ, $n/2 \le \tau \le n$ which are directed into K. At step τ the number of such edges falling in L is an independent random variable with distribution dominating

$$1_{\tau \in W_2} \mathrm{Bi}\left(m, \frac{\lambda q}{\alpha \tau}\right) + 1_{\tau \in K_2} \mathrm{Bi}\left(m, \frac{\lambda(n/2 - q)}{\alpha \tau}\right).$$

Thus

$$\begin{aligned} \mu(K_1, K_2) &\ge \frac{m\lambda q}{\alpha} \sum_{\tau \in W_2} \frac{1}{\tau} + \frac{m\lambda(n/2 - q)}{\alpha} \sum_{\tau \in K_2} \frac{1}{\tau} \\ &= \frac{m\lambda}{\alpha} \left((k - r) \sum_{\tau \in W_2} \frac{1}{\tau} + (n/2 - (k - r)) \sum_{\tau \in K_2} \frac{1}{\tau} \right). \end{aligned}$$

Let $\mu(k) = \min_{K_1, K_2} \mu(K_1, K_2)$. Then 'somewhat crudely'

$$\begin{aligned} \sum_{\tau \in W_2} \frac{1}{\tau} &\ge \ln \frac{n}{n/2 + r} \\ \sum_{\tau \in K_2} \frac{1}{\tau} &\ge \ln \frac{n}{n - r}. \end{aligned}$$

Thus

$$\mu(k) \ge \frac{m\lambda}{\alpha} \left((k - r) \ln \frac{2n}{n + 2r} + \left(\frac{n}{2} - (k - r)\right) \ln \frac{n}{n - r} \right).$$

Putting $k = \kappa n$ and $r = \rho n$ we see that

$$\mu(k) \ge \frac{\lambda mn}{\alpha} g(\kappa, \rho)$$

where

$$g(\kappa,\rho)=(\kappa-\rho)\ln\frac{2}{1+2\rho}+\left(\frac{1}{2}-\kappa+\rho\right)\ln\frac{1}{1-\rho}.$$

We put a lower bound on g:

$$\rho\leq\frac{\xi}{2}\text{ implies }\kappa-\rho\geq\frac{\xi}{2}\text{ and so }g(\kappa,\rho)\geq\frac{\xi}{2}\ln\frac{2}{1+\xi}.$$

So we can assume that $\rho\geq\xi/2$. Then

$$\begin{aligned}
\kappa-\rho\leq\frac{1-\xi}{2} \quad & \text{implies} \quad g(\kappa,\rho)\geq\frac{\xi}{2}\ln\frac{2}{2-\xi}.\\
\kappa-\rho>\frac{1-\xi}{2}\text{ and }\rho\leq\frac{1-\xi}{2} \quad & \text{implies} \quad g(\kappa,\rho)\geq\frac{1-\xi}{2}\ln\frac{2}{2-\xi}.\\
\kappa-\rho>\frac{1-\xi}{2}\text{ and }\rho>\frac{1-\xi}{2} \quad & \text{implies} \quad \kappa>1-\xi.
\end{aligned}$$

We deduce that within our range of interest,

$$\mu(k)\geq\eta mn$$

for some absolute constant η.

Let Z be the number of edges generated within L, so that Z counts a subset of the edges between K and $\overline{K}$. Then

$$\mathbf{Pr}\left(\exists K_1,K_2\subseteq N:\ Z\leq\frac{1}{2}\eta mn\right)\leq 2^n e^{-\eta mn/8}\leq e^{-\eta mt/10}=o(1).$$

This completes the proof of Theorem 3(a). Part (b) is an immediate consequence of Part (a).

To prove part (c) we need to prove some vertex expansion properties of G_n. So fix $K\subseteq V_n$ with $1\leq|K|\leq A_0n$ and go back to Section 6.1. We see that the number of neighbors of K_1 in W_2 dominates $B_1=\mathrm{Bi}(n/2-|K_2|,1-(1-\lambda|K_1|/(\alpha n))^m)$ and the number of neighbours of K_2 in W_1 dominates $B_2=\mathrm{Bi}(n/2-|K_1|,1-(1-\lambda/(\alpha n))^{m|K_2|})$. So, for $i=1,2$,

$$\mathbf{E}\left[B_i\right]\geq\begin{cases}\frac{\lambda m|K_i|}{3\alpha} & if\ \frac{\lambda m|K_i|}{\alpha n}\leq\frac{1}{10}\\ \frac{n}{60} & otherwise\end{cases}$$

Therefore, using the Chernoff bounds, we have

$$\begin{aligned}
\mathbf{Pr}\left(\exists K,i:\ 1\leq|K_i|\leq\frac{\alpha n}{10\lambda m}\ and\ B_i\leq\frac{\lambda m|K_i|}{6\alpha}\right) &\leq \sum_{k=1}^{\alpha n/(10\lambda m)}\binom{n}{k}e^{-\lambda mk/(24\alpha)}\\
&= o(1). \qquad (22)
\end{aligned}$$

$$\begin{aligned}
\mathbf{Pr}\left(\exists K,i:\ \frac{\alpha n}{10\lambda m}\leq|K_i|\leq A_0n\ and\ B_i\leq\frac{n}{120}\right) &\leq \sum_{k=1}^{A_0n}\binom{n}{k}e^{-n/1000}\\
&= o(1). \qquad (23)
\end{aligned}$$

Now fix $x, y \in V_n$ and then for $a = x, y$ let $S_{i,a} = \{z \in V_n : \text{dist}(a, z) = i\}$. Here $\text{dist}(a, z)$ is the graph distance between a and z in G_n. It follows from (22) and (23) that there exists $j_a = O(\log_m n)$ such that $|S_{j,a}| \geq n/120$. It follows from the proof of Lemma 6 that if $|S_{j_a}| \leq (1 - \xi)n$ then $|E(S_{j_a} : \bar{S}_{j_a}| \geq \eta mn/120$. It follows that there exists $l_a \leq 240/\eta$ such that $|S_{j_a+l_a}| \geq (1-\xi)n \geq 3n/4$. It follows that $S_{j_x+l_x} \cap S_{j_y+l_y} \neq \emptyset$ and $\text{dist}(x, y) \leq j_x + j_y + l_x + l_y = O(\log_m n)$. This completes the proof of Theorem 3. □

Acknowledgement: We thank Henri van den Esker for pointing out some errors and for some very useful comments.

References

[1] W. Aiello, F. R. K. Chung, and L. Lu, A random graph model for massive graphs, *Proc. of the 32nd Annual ACM Symposium on the Theory of Computing*, (2000) 171–180.

[2] N. Alon and J. Spencer, The Probabilistic Method, Second Edition, Wiley-Interscience 2000.

[3] W. Aiello, F. R. K. Chung, and L. Lu, Random Evolution in Massive Graphs, *Proc. of IEEE Symposium on Foundations of Computer Science*, (2001) 510–519.

[4] R. Albert, A. Barabási, and H. Jeong, Diameter of the world wide web, *Nature* 401 (1999) 103–131.

[5] A. Barabasi and R. Albert, Emergence of scaling in random networks, *Science* 286 (1999) 509–512.

[6] N. Berger, B. Bollobas, C. Borgs, J. Chayes, and O. Riordan, Degree distribution of the FKP network model, *Proc. of the 30th International Colloquium of Automata, Languages and Programming*, (2003) 725–738.

[7] N. Berger, C. Borgs, J. Chayes, R. D'Souza, and R. D. Kleinberg, Competition-induced preferential attachment, Proceedings of the 31st International Colloquium on Automata, Languages and Programming (ICALP), 208-221, Lecture Notes in Computer Science 3142 (2004).

[8] D. Blandford, G. E. Blelloch, and I. Kash, Compact Representations of Separable Graphs, *Proc. of ACM/SIAM Symposium on Discrete Algorithms* (2003) 679–688.

[9] B. Bollobás and O. Riordan, Mathematical Results on Scale-free Random Graphs, in *Handbook of Graphs and Networks*, Wiley-VCH, Berlin, 2002.

[10] B. Bollobás and O. Riordan, The diameter of a scale-free random graph, *Combinatorica*, 4 (2004) 5–34.

[11] B. Bollobás and O. Riordan, Coupling scale free and classical random graphs, *Internet Mathematics* 1 (2004), no. 2, 215–225.

[12] B. Bollobás, O. Riordan, J. Spencer and G. Tusanády, The degree sequence of a scale-free random graph process, *Random Structures and Algorithms* 18 (2001) 279–290.

[13] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener, Graph structure in the web, *Proc. of the 9th Intl. World Wide Web Conference* (2002) 309–320.

[14] G. Buckley and D. Osthus, Popularity based random graph models leading to a scale-free degree distribution, *Discrete Mathematics* 282 (2004) 53–68.

[15] F.R.K. Chung, L. Lu, and V. Vu, Eigenvalues of random power law graphs, *Annals of Combinatorics* 7 (2003) 21–33.

[16] F.R.K. Chung, L. Lu, and V. Vu, The spectra of random graphs with expected degrees, *Proceedings of national Academy of Sciences* 100 (2003) 6313–6318.

[17] C. Cooper and A. M. Frieze, A General Model of Undirected Web Graphs, *Random Structures and Algorithms*, 22 (2003) 311–335.

[18] C. Cooper, A. M. Frieze, and J. Vera, Random deletions in a scale free random graph process, *Internet Mathematics* 1 (2004), 463–483.

[19] E. Drinea, M. Enachescu, and M. Mitzenmacher, *Variations on Random Graph Models for the Web*, Harvard Technical Report TR-06-01 (2001).

[20] E. Estrada, Spectral scaling and good expansion properties in complex networks, *Europhysics Letters* 73 (4) (2006) 649–655.

[21] P. Erdős and A. Rényi, On random graphs I, *Publicationes Mathematicae Debrecen* 6 (1959) 290–297.

[22] A. Fabrikant, E. Koutsoupias, and C. H. Papadimitriou, Heuristically Optimized Trade-Offs: A New Paradigm for Power Laws in the Internet, *Proc. of 29th International Colloquium of Automata, Languages and Programming* (2002) .

[23] M. Faloutsos, P. Faloutsos, and C. Faloutsos, On Power-law Relationships of the Internet Topology, *ACM SIGCOMM Computer Communication Review* 29 (1999) 251–262.

[24] A. Flaxman, Expansion and lack thereof in randomly perturbed graphs, to appear in *Proc. of the Web Algorithms Workshop* (2006).

[25] A. Flaxman, A. M. Frieze and J. Vera, A Geometric Preferential Attachment Model of Networks, Internet Mathematics, to appear.

[26] J. Gómez-Gardeñes and Y. Moreno, Local versus global knowledge in the Barabási-Albert scale-free network model, *Physical Review E* 69 (2004) 037103.

[27] B. Hayes, Graph theory in practice: Part II, *American Scientist* 88 (2000) 104-109.

[28] J. M. Kleinberg, R. Kumar, P. Raghavan, S. Rajagopalan, and A. S. Tomkins, The Web as a Graph: Measurements, Models and Methods, *Proc. of the 5th Annual Intl. Conf. on Combinatorics and Computing (COCOON)* (1999).

[29] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal, Stochastic Models for the Web Graph, *Proc. IEEE Symposium on Foundations of Computer Science* (2000) 57.

[30] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal, The Web as a Graph, *Proc. 19th ACM SIGACT-SIGMOD-AIGART Symp. Principles of Database Systems (PODS)* (2000) 1–10.

[31] L. Li, D. Alderson, J. C. Doyle, W. Willinger, Towards a Theory of Scale-Free Graphs: Definition, Properties, and Implications, *Internet Mathematics* Vol. 2, No. 4, 431–523.

[32] M. Mihail.

[33] R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins, Trawling the Web for emerging cyber-communities, *Computer Networks* 31 (1999) 1481–1493.

[34] C. J. H. McDiarmid, Concentration, in *Probabilistic methods in algorithmic discrete mathematics*, (1998) 195–248.

[35] M. Mihail and C. H. Papadimitriou, On the Eigenvalue Power Law, *Proc. of the 6th International Workshop on Randomization and Approximation Techniques* (2002) 254–262.

[36] M. Mihail, C. H. Papadimitriou, and A. Saberi, On Certain Connectivity Properties of the Internet Topology, *Proc. IEEE Symposium on Foundations of Computer Science* (2003) 28.

[37] M. Mitzenmacher, A brief history of generative models for power law and lognormal distributions, *Internet Mathematics* 1 (2004), no. 2, 226–251.

[38] M. Newman, A.-L. Barabási, D. J. Watts, *The Structure and Dynamics of Networks*, Princeton University Press (2006).

[39] M. D. Penrose, *Random Geometric Graphs*, Oxford University Press (2003).

[40] H. A. Simon, On a class of skew distribution functions, *Biometrika* 42 (1955) 425-440.

[41] R. van der Hofstad, *Random Graphs and Complex Networks*, unpublished manuscript (2007).

[42] D. J. Watts, *Small Worlds: The Dynamics of Networks between Order and Randomness*, Princeton: Princeton University Press (1999).

[43] G. Yule, A mathematical theory of evolution based on the conclusions of Dr. J.C. Willis, *Philosophical Transactions of the Royal Society of London (Series B)* 213 (1925) 21–87.