A geometric preferential attachment model of networks

Abraham D. Flaxman, Alan M. Frieze*, Juan Vera
Department of Mathematical Sciences,
Carnegie Mellon University,
Pittsburgh PA15213,
U.S.A.

July 28, 2004

Abstract

We study a random graph $G_n$ that combines certain aspects of geometric random graphs and preferential attachment graphs. 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 $R^d$. After generating $x_1$, we randomly connect it to $m$ points from those points in $x_1, x_2, \ldots, x_{n-1}$ which are within distance $r$. Neighbours are chosen with probability proportional to their current degree. We show that if $m$ is sufficiently large and if $r \geq \log n/n^{1/2-\beta}$ for some constant $\beta$, then w.h.p at time $n$ the number of vertices of degree $k$ follows a power law with exponent 3. Unlike the preferential attachment graph, this geometric preferential attachment graph has small separators, similar to experimental observations of [7]. We further show that if $m \geq K \log n$, $K$ sufficiently large, then $G_n$ is connected and has diameter $O(m/r)$ w.h.p.

1 Introduction

Recently there has been much interest in understanding the properties of real-world large-scale networks such as the structure of the Internet and the World Wide Web. For a general introduction to this topic, see Bollobás and Riordan [8], Hayes [21], Watts [32], or Aiello, Chung and Lu [2]. One approach is to model these networks by random graphs. Experimental studies by Albert, Barabási, and Jeong [3], Broder et al [12], and Faloutsos, Faloutsos, and Faloutsos [20] have demonstrated that in the World Wide Web/Internet the proportion of vertices of a given degree follows an approximate inverse power law i.e. the proportion of vertices of degree $k$ is approximately $Ck^{-\alpha}$ for some constants $C, \alpha$. The classical models of random graphs introduced by Erdős and Rényi [18] do not have power law degree sequences, so they are not suitable for modeling these networks. This has driven the development of various alternative models for random graphs.

One approach is to generate graphs with a prescribed degree sequence (or prescribed expected degree sequence). This is proposed as a model for the web graph by Aiello, Chung,

*Supported in part by NSF grant CCR-0200945.
and Lu in [1]. Mihail and Papadimitriou also use this model [27] in their study of large eigenvalues, as do Chung, Lu, and Vu in [14].

An alternative approach, which we will follow in this paper, is to sample graphs via some generative procedure which yields a power law distribution. There is a long history of such models, outlined in the survey by Mitzenmacher [29]. We will use an extension of the preferential attachment model to generate our random graph. The preferential attachment model has been the subject of recently revived interest. It dates back to Yule [33] and Simon [31]. It was proposed as a random graph model for the web by Barabási and Albert [4], and their description was elaborated by Bollobás and Riordan [9] who showed that at time n, \( \text{whp} \) the diameter of a graph constructed in this way is asymptotic to \( \frac{\log \log n}{\log n} \). Subsequently, Bollobás, Riordan, Spencer and Tusnády [11] proved that the degree sequence of such graphs does follow a power law distribution.

The random graph defined in the previous paragraph has good expansion properties. For example, Mihail, Papadimitriou and Saberi [28] showed that \( \text{whp} \) the preferential attachment model has conductance bounded below by a constant. This is in contrast to what has sometimes been found experimentally, for example by Blandford, Blelloch and Kash [7]. Their results seem to suggest the existence of smaller separators than implied by random graphs with the same average degree. The aim of this paper is to describe a random graph model which has both a power-law degree distribution and which has small separators.

We study here the following 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 \( e_t \) edges. Here \( V_t \) is a subset of \( S \), the surface of the sphere in \( \mathbb{R}^3 \) of radius \( \frac{1}{\sqrt{\pi}} \) (so that \( \text{area}(S) = 1 \)).

For \( u \in S \) and \( r > 0 \) we let \( B_r(u) \) denote the spherical cap of radius \( r \) around \( u \) in \( S \). More precisely, \( B_r(u) = \{ x \in S : \|x - u\| \leq r \} \).

1.1 The random process

- **Time step 1:** To initialize the process, we start with \( G_1 \) containing a single vertex \( x_1 \) chosen at random in \( S \). The edge (multi)set consists of \( m \) loops at \( x_1 \).

- **Time step \( t + 1 \):** We choose vertex \( x_{t+1} \) uniformly at random in \( S \) and add it to \( G_t \). If \( V_t \cap B_r(x_{t+1}) \) is nonempty, 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 from \( V_t \cap B_r(x_{t+1}) \) and for \( x \in V_t \cap B_r(x_{t+1}) \),

\[
\Pr(y_i = x) = \frac{\text{deg}_G(x)}{D_t(B_r(x_{t+1}))},
\]

where \( \text{deg}_G(x) \) denotes the degree of vertex \( x \) in \( G_t \) and \( V_t(U) = V_t \cap U \) and \( D_t(U) = \sum_{x \in V_t(U)} \text{deg}_G(x) \).

If \( V_t \cap B_r(x_{t+1}) \) is empty then we add \( m \) loops at \( x_{t+1} \).

Let \( d_k(t) \) denote the number of vertices of degree \( k \) at time \( t \).
We will prove the following:

Theorem 1

(a) If $0 < \beta < 1/2$ is constant and $r \geq n^{\beta-1/2}\log n$ and $m$ is a sufficiently large constant then there exists a constant $c > 0$ such that \textbf{w.h.p}

$$d_k(n) = \frac{cn}{k(k+1)(k+2)} + O(n^{1-\gamma})$$

for some $0 < \gamma < 1$.

(b) If $r = o(1)$ then \textbf{w.h.p} $V_n$ can be partitioned into $T, \bar{T}$ such that $|T|, |ar{T}| \sim n/2$, and there are at most $4\sqrt{\pi nm}$ edges between $T$ and $\bar{T}$.

(c) If $r \geq n^{-1/2}\log n$ and $m \geq K\log n$ and $K$ is sufficiently large then \textbf{w.h.p} $G_n$ is connected.

(d) If $r \geq n^{-1/2}\log n$ and $m \geq K\log n$ and $K$ is sufficiently large then \textbf{w.h.p} $G_n$ has diameter $O(\log n/r)$.

We note that geometric models of trees with power laws have been considered in [19], [5] and [6].

1.2 Some definitions

There exists some constant $c_0$ such that for any $u \in S$, we have

$$A_r = \text{Area}(B_r(u)) = c_0 n^{2\beta-1}(\log n)^2.$$ 

Given $u \in S$, we define

$$V_t(u) = V_t(B_r(u))$$

and

$$D_t(u) = D_t(B_r(u)).$$

Given $v \in V_t$, we have

$$\deg_t(v) = m + \deg_t^-(v),$$

where $\deg_t^-(v)$ is the number of edges of $G_t$ that are incident to $v$ and were added by vertices that chose $v$ as a neighbor.

Given $U \subseteq S$, let $D_t^-(U) = \sum_{v \in V_t(U)} \deg_t^-(v)$. We also define $D_t^-(u) = D_t^-(B_r(u))$.

Notice that $D_t(U) = m|V_t(U)| + D_t^-(U)$.

We write $d_k(t)$ to denote the expectation of $d_k(t)$. We also localize these notions: given $U \subseteq S$ and $u \in S$ we define $d_k(t, U)$ to be the number of vertices of degree $k$ at time $t$ in $U$ and $d_k(t, u) = d_k(t, B_r(u))$.

---

\textsuperscript{1}Asymptotics are taken as $n \to \infty$
2 Small separators

Theorem 1 part (b) is the easiest part to prove. We use the geometry of the instance to obtain a sparse cut. Consider partitioning the vertices using a great circle of $S$. This will divide $V$ into sets $T$ and $T'$ which each contain about $n/2$ vertices. More precisely, we have

$$\Pr [\|T\| < (1 - \epsilon)n/2] = \Pr [\|T'\| < (1 - \epsilon)n/2] \leq e^{-c^2 n/6}.$$ 

Since edges only appear between vertices within distance $r$, only vertices appearing in the strip within distance $r$ of the great circle can appear in the cut. Since $r = o(1)$, this strip has area less than $3r\sqrt{\pi}$, so, letting $U$ denote the vertices appearing in this strip, we have

$$\Pr [\|U\| \geq 4\sqrt{\pi}r_n] \leq e^{-\sqrt{\pi}r_n/9}.$$

vertices. Even if every one of the vertices chooses its $m$ neighbors on the opposite side of the cut, this will yield at most $4\sqrt{\pi} r m$ edges w.h.p. So the graph has a cut with $\frac{e(T, T')}{\|T\| \|T'\|} \leq \frac{17\sqrt{\pi} r m}{n}$ with probability at least $1 - e^{-\Omega(rm)}$.

3 Proving a power law

3.1 Establishing a recurrence for $D_k(t)$

Our approach to proving Theorem 1 part (a) is to find a recurrence for $D_k(t)$.

We define $d_m(t) = 0$ for all integers $t$ with $t > 0$. Let $\eta_1(t)$ denote the probability that $V_t \cap B_r(x_{t+1}) = \emptyset$ so $\eta_1(t) = (1 - A_r)^t$. Let $\eta_2(t)$ denote the probability that a parallel edge is created. Thus

$$\eta_2(t) = O\left( \sum_{i=m}^{k} d_i(t, x_{t+1})t^2 / D_t(x_{t+1})^2 \right) = O(k / D_t(x_{t+1})).$$

Then for $k \geq m$,

$$\mathbb{E}[d_k(t + 1) \mid G_t, x_{t+1}] = d_k(t) + md_{k-1}(t, x_{t+1}) \frac{k - 1}{D_t(x_{t+1})}$$

$$-md_k(t, x_{t+1}) \frac{k}{D_t(x_{t+1})} + 1_{k=m} + O(\eta_1(t) + \eta_2(t)). \quad (2)$$

Let

$$\alpha = \frac{1}{400} \quad \text{and} \quad \gamma = \frac{\alpha^2}{2(1 - \alpha^2)},$$

and let $A_t$ be the event

$$\{ |D_t(x_{t+1}) - 2mA_t| < A_{r, t}1^{1-\gamma} \}.$$
Then, because $\mathbb{E}[d_k(t, x_{t+1})] \leq k^{-1}\mathbb{E}[m | V_t(B_2(x_{t+1}))] \leq k^{-1}m(4A_r t)$ and $d_k(t, x_{t+1}) \leq k^{-1}D_t(x_{t+1})$, we have

$$
\mathbb{E} \left[ \frac{d_k(t, x_{t+1})}{D_t(x_{t+1})} \right] = \mathbb{E} \left[ \frac{d_k(t, x_{t+1})}{D_t(x_{t+1})} \right] \mathbf{P}[A_t] + \mathbb{E} \left[ \frac{d_k(t, x_{t+1})}{D_t(x_{t+1})} \right] \mathbf{P}[-A_t]
$$

$$
= \mathbb{E} \left[ \frac{d_k(t, x_{t+1})}{2mA_r t} \right] \mathbf{P}[A_t] + O \left( \frac{t^{-\gamma}}{k} \right) + O \left( \frac{1}{k} \mathbf{P}[-A_t] \right)
$$

$$
= \mathbb{E} \left[ \frac{d_k(t, x_{t+1})}{2mA_r t} \right] + O \left( \frac{t^{-\gamma}}{k} \right) + O \left( \frac{1}{k} + \frac{1}{A_r} \mathbf{P}[-A_t] \right)
$$

In Lemmas 1 and 3 below we prove that $\mathbb{E}[d_k(t, x_{t+1})] = A_r \overline{d}_k(t)$ and that if $t \geq n^{1-\alpha}$ then $\mathbf{P}[-A_t] = O \left( n^{-2} \right)$. Therefore if $t \geq n^{1-\alpha}$ then

$$
\mathbb{E} \left[ \frac{d_k(t, x_{t+1})}{2mt} \right] = \overline{d}_k(t) + O \left( \frac{t^{-\gamma}}{k} \right). \quad (3)
$$

In a similar way

$$
\mathbb{E} \left[ \frac{d_k(t, x_{t+1})}{2mA_r t} \right] = \mathbb{E} \left[ \frac{d_{k-1}(t, x_{t+1})}{2mt} \right] + O \left( \frac{t^{-\gamma}}{k^{-1}} \right). \quad (4)
$$

Now note that

$$
\eta_2(t) \leq \mathbf{P}[-A_t] + O(k/tA_r).
$$

Taking expectations on both sides of Eq. (2) and using Eq. (3) and Eq. (4), we see that if $k \leq n^{1-\alpha} \leq t$ then

$$
\overline{d}_k(t + 1) = \overline{d}_k(t) + \frac{k}{2t} \overline{d}_{k-1}(t) - \frac{k}{2t} \overline{d}_k(t) + 1_{k=m} + O \left( t^{-\gamma} \right) \quad (5)
$$

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

$$
f_k = 1_{k=m} + \frac{k-1}{2} f_{k-1} - \frac{k}{2} f_k
$$

which has solution

$$
f_k = f_m \prod_{i=m+1}^{k} \frac{i-1}{k+2} = f_m \frac{m(m+1)(m+2)}{k(k+1)(k+2)}
$$
Let \( t_0 = n^{1-\alpha} \). We finish the proof of Theorem 1(a) by showing that there exists a constant \( M > 0 \) such that
\[
|\bar{d}_k(t) - f_k t| \leq M(t_0 + t^{1-\gamma})
\]for all \( k \) with \( m \leq k \leq n^{1-\alpha} \) and all \( t > 0 \). For \( k > n^{1-\alpha} \) we use the fact that \( \bar{d}_k(t) \leq 2mt/k \).

Let \( \Theta_k(t) = \bar{d}_k(t) - f_k t \). Then for \( m \leq k \leq n^\alpha \) and \( t \geq t_0 \),
\[
\Theta_k(t + 1) = \frac{k - 1}{2t} \Theta_k(t - 1) - \frac{k}{2t} \Theta_k(t) + O(t^{-\gamma}).
\]

Let \( L \) denote the hidden constant in \( O(t^{-\gamma}) \) of (7). Our inductive hypothesis \( H_k \) is that
\[
|\Theta_k(t)| \leq M(t_0 + t^{1-\gamma})
\]
for every \( m \leq k \leq n^{1-\alpha} \). It is trivially true for \( t \leq t_0 \). So assume that \( t \geq t_0 \). Then, from (7),
\[
|\Theta_k(t + 1)| \leq M(t_0 + t^{1-\gamma}) + Lt^{-\gamma} \leq M(t_0 + (t + 1)^{1-\gamma}).
\]

This verifies \( H_{k+1} \) and completes the proof by induction.

3.2 Expected Value of \( d_k(t, u) \)

**Lemma 1** Let \( u \in S \) and let \( k \) and \( t \) be positive integers. Then \( \mathbb{E}[d_k(t, u)] = A_r \bar{d}_k(t) \)

**Proof** By symmetry for any \( w \in S \), \( d_k(t, u) \sim d_k(t, w) \). Then
\[
\mathbb{E}[d_k(t, u)] = \int_S \mathbb{E}[d_k(t, u)] \, dw = \int_S \mathbb{E}[d_k(t, w)] \, dw
\]
\[
= \mathbb{E} \left[ \int_S d_k(t, w) \, dw \right] = \mathbb{E} \left[ \int_S \sum_{v \in V_t} 1_{\deg v = k} 1_{v \in B_r(w)} \, dw \right]
\]
\[
= \mathbb{E} \left[ \sum_{v \in V_t} 1_{\deg v = k} \int_S 1_{w \in B_r(v)} \, dw \right] = \mathbb{E} \left[ \sum_{v \in V_t} 1_{\deg v = k} A_r \right]
\]
\[
= A_r \mathbb{E}[d_k(t)]
\]
\[\square\]

**Lemma 2** Let \( u \in S \) and \( t > 0 \) then \( \mathbb{E}[D_t(u)] = 2A_r mt \)

**Proof**
\[
\mathbb{E}[D_t(u)] = \sum_{k > 0} \mathbb{E}[d_k(t, u)] = A_r \sum_{k > 0} \mathbb{E}[d_k(t)] = A_r \mathbb{E} \left[ \sum_{k > 0} d_k(t) \right] = 2A_r mt
\]
\[\square\]
3.3 Concentration of $D_t(u)$

In this section we prove

**Lemma 3** Let $\alpha = 1/400$ and $\gamma = \frac{\alpha^2}{3(1-\alpha)}$ and $n_0 = n^{1-2\alpha}$. If $t > n^{1-\alpha}$ and $u \in S$ then

$$\Pr \left[ |D_t(u) - E[D_t(u)]| \geq A_r m t^{1-\gamma} \right] = O\left(n^{-2}\right).$$

**Proof** We think of every edge added as two directed edges. Then choosing a vertex by preferential attachment is equivalent to choosing one of these directed edges uniformly, and taking the vertex pointed to by this edge as the chosen vertex. So the $i$th step of the process is defined by a tuple of random variables $T = (X, Y_1, \ldots, Y_m) \in S \times E_i^n$ where $X$ is the location of the new vertex, a randomly chosen point in $S$, and $Y_j$ is an edge chosen u.a.r. among the edges directed into $B_r(X)$ in $G_{i-1}$. The process $G_t$ is then defined by a sequence $\langle T_1, \ldots, T_t \rangle$, where each $T_i \in S \times E_i^n$.

Let $s$ be a sequence $s = \langle s_1, \ldots, s_t \rangle$ where $s_i = (x_i, y_{(i-1)m+1}, \ldots, y_{im})$ with $x_i \in S$ and $y_j \in E_{\lfloor s_j \rfloor}$. We say $s$ is acceptable if for every $j$, $y_j$ is an edge entering $B_r(x_{\lfloor s_j \rfloor})$. Notice that non-acceptable sequences have probability 0 of being observed.

In what follows we condition on the event

$$E = \{\text{for all } s \text{ with } n_0 < s \leq n \text{ we have } D_s(u) > (1 + \alpha) A_r m s\},$$

where $\alpha > 0$ is an appropriate constant that will be chosen later.

Fix $t > 0$. Fix an acceptable sequence $s = \langle s_1, \ldots, s_t \rangle$, and let $A_k(s) = \{z \in S \times E_k^n : \langle s_1, \ldots, s_k-1, z \rangle \text{ is acceptable}\}$. For any $k$ with $1 \leq k \leq t$ and any $z \in A_k(s)$ let

$$g_k(z) = E[D_k(u) \mid T_1 = s_1, \ldots, T_{k-1} = s_{k-1}, T_k = z, E],$$

let $r_k(s) = \sup\{|g_k(z_1) - g_k(z_2)| : z_1, z_2 \in A_k(s)\}$ and let $R^2(s) = \sum_{k=1}^t r_k(s)^2$. Finally define $\bar{r}^2 = \sup_s R^2(s)$, where the supremum is taken over all acceptable sequences.

By Thm 3.7 of [26] we know that for all $\lambda > 0$,

$$\Pr \left[ |D_t(u) - E[D_t(u)]| \geq \lambda \right] < 2e^{-2\lambda^2/\bar{r}^2} + \Pr [-E]. \quad (8)$$

Now, fix $k$, with $1 \leq k \leq t$. Our goal now is to bound $r_k(s)$ for any acceptable sequence $s$.

Fix $z, z' \in A_k(s)$. We define $\Omega(G_t, G'_t)$, the following coupling between $G_t = G_t(s_1, \ldots, s_{k-1}, z)$ and $G'_t = G_t(s_1, \ldots, s_{k-1}, z')$

- **Step k**: Start with the graph $G_k(s_1, \ldots, s_{k-1}, z)$ and $G'_k(s_1, \ldots, s_{k-1}, z')$ respectively.

- **Step $\tau$ ($\tau > k$)**: Choose the same point $x_{\tau} \in S$ in both processes. Let $E_\tau$ (resp. $E'_\tau$) be the edges pointing to the vertices in $B_r(x_{\tau})$ in $G_{\tau-1}$ (resp. $G'_{\tau-1}$). Let $C_\tau = E_\tau \cap E'_\tau$, $R_\tau = E_\tau \setminus E'_\tau$, and $L_\tau = E'_\tau \setminus E_\tau$.

Let $D_\tau = |E_\tau|$ and $D'_\tau = |E'_\tau|$. Without loss of generality assume that $D_\tau \leq D'_\tau$.

Note that $D_\tau = 0$ if $V_\tau \cap B_r(x_{\tau}) = 0$, in which case $D'_\tau = 0$ as well. Assume for now that $D_\tau > 0$. Let $p = 1/D_\tau$ and let $p' = 1/D'_\tau$. Construct $G_\tau$ choosing $m$ edges u.a.r. $e_1, \ldots, e_m$ in $E_\tau$, and joining $x_{\tau}$ to the end point of them. For each of the $m$ edges $e_i = e_i^\tau$, we define $\hat{e}_i = \hat{e}_i^\tau$ by
If \( e_i \in C_{\tau} \) then, with probability \( p'/p \), \( \hat{e}_i = e_i \). With probability \( 1 - p'/p \), \( \hat{e}_i \) is chosen from \( L_{\tau} \) u.a.r.

- If \( e_i \in R_{\tau} \), \( \hat{e}_i \in L_{\tau} \) is chosen u.a.r.

Notice that for every \( i = 1, \ldots, m \) and every \( e \in E'_\tau \), \( \Pr[\hat{e}_i = e] = p' \). To finish, in \( G'_\tau \) join \( x_\tau \) to the \( m \) vertices pointed to by the edges \( \hat{e}_i \).

**Lemma 4** Let \( \Delta_\tau = \Delta_\tau(k, s, z, z', u) = |E_{G_\tau}(B_r(u)) \Delta E_{G'_\tau}(B_r(u))| \), the discrepancy between the edge-sets incident to \( V_\tau(u) \) in the two coupled graphs. Then \( |g_k(z) - g_k(z')| \leq E[\Delta_1 \mathcal{E}] / \Pr[\mathcal{E}] \).

**Proof**

\[
|g_k(z) - g_k(z')| = |E_{G_\tau}[D_k(u) \mid \mathcal{E}] - E_{G'_\tau}[D_k(u) \mid \mathcal{E}]| \\
\leq E_{\Omega(G_\tau, G'_\tau)}[D'_k(u) - D_k(u) \mid \mathcal{E}] \\
\leq E_{\Omega(G_\tau, G'_\tau)}[\Delta_\tau \mid \mathcal{E}] \\
= E_{\Omega(G_\tau, G'_\tau)}[\Delta_1 \mathcal{E}] / \Pr[\mathcal{E}] .
\]

Recall that \( A_\tau = \text{Area}(B_r(u)) = c_0 n^{2\beta-1}( \log n )^2 \) and \( n_0 = n^{1-2\alpha} \) and we have fixed \( k \) to be an integer with \( 1 \leq k \leq t \).

**Lemma 5** Let \( k_0 = \max\{k, n_0\} \) and \( \epsilon_1 \) be a small positive constant, then

\[
E[\Delta_1 \mathcal{E}] \leq 16 m A_\tau n^{\alpha \epsilon_1} \left( \frac{k_0}{k} \right)^{1/(1-2\epsilon_1)} \left( \frac{t}{k_0} \right)^{1/(1+\alpha)}.
\]

**Proof** Notice that for \( \tau > k \), \( \Delta_\tau = \Delta_{\tau-1} + Y_\tau \), where \( Y_\tau \) is the edge discrepancy created at time \( \tau \). \( Y_\tau \leq 2 \sum_{i=1}^m 1_{e_i \not\in \hat{e}_i} \) where \( 1_{e_i \not\in \hat{e}_i} \) is the indicator variable of the \( t \)th edge chosen at time \( \tau \) not being the same for both processes. For every \( i = 1, \ldots, m \)

\[
\Pr[e_i \not\in \hat{e}_i \mid G_{\tau-1}, G'_{\tau-1}, x_\tau] = 1 - \frac{|C_{\tau}| p'}{|E_{\tau}|} = 1 - \frac{|C_{\tau}|}{|E_{\tau}'|} = \frac{|L_{\tau}|}{\max\{D_{\tau}, D'_{\tau}\}}.
\]

Therefore

\[
E[\Delta_\tau \mid G_{\tau-1}, G'_{\tau-1}, x_\tau] = \Delta_{\tau-1} + 2m \frac{|L_{\tau}|}{\max\{D_{\tau}, D'_{\tau}\}}.
\]

For each \( e \in E(G'_\tau) \setminus E(G_\tau) \), \( e \in L_{\tau} \) implies \( x_{\tau+1} \) is in the ball of radius \( r \) centered at the end point of \( e \). Thus \( \Pr[e \in L_{\tau} \mid G_{\tau-1}, G'_{\tau-1}] \leq A_\tau \) and therefore

\[
E[|L_{\tau}| \mid G_{\tau-1}, G'_{\tau-1}] \leq A_\tau \Delta_{\tau-1}.
\]

\[
E[|L_{\tau}|] \leq E[A_\tau \Delta_{\tau-1}/2],
\]

where the factor of \( 1/2 \) in the unconditional expectation comes from symmetry.

Let \( k_1 = \epsilon_1 n^{2\beta-1}/\log n \).
Case 1 $\tau \geq k_0$.
If $E$ holds then $D_{\tau} > (1 + \alpha)m A_r \tau$ and so, by using (10), we have

$$E[\Delta_{\tau+1|E}] \leq E[\Delta_{\tau|E}] + 2m \frac{E[|L_{\tau}|]}{(1 + \alpha)m A_r \tau} \leq E[\Delta_{\tau|E}] \left( 1 + \frac{1}{(1 + \alpha)\tau} \right) .$$

(11)

Case 2 $k \leq \tau < k_1$.
If $|L_{\tau}| > 0$ then $\max\{D_{\tau}, D'_{\tau}\} \geq 1$. So we have

$$E[\Delta_{\tau|E}] \leq E[\Delta_{\tau-1|E}] (1 + m A_r).$$

Case 3 $\max\{k, k_1\} \leq \tau < k_0$.
We write

$$E\left[ \frac{|L_{\tau}|}{\max\{D_{\tau}, D'_{\tau}\}} \right] \leq \frac{E\left[ \frac{|L_{\tau}|}{m(1 - \epsilon_1)A_r \tau} \right] |V_r(x_{\tau+1})| \geq (1 - \epsilon_1)A_r \tau}{Pr[|V_r(x_{\tau+1})| \geq (1 - \epsilon_1)A_r \tau]} + \frac{E\left[ \frac{|L_{\tau}|}{m(1 - \epsilon_1)A_r \tau} \right] |V_r(x_{\tau+1})| < (1 - \epsilon_1)A_r \tau}{Pr[|V_r(x_{\tau+1})| < (1 - \epsilon_1)A_r \tau]} \leq \frac{A_r E[\Delta_{\tau-1|E}]/2}{2m(1 - 2\epsilon_1)\tau} + A_r E[\Delta_{\tau-1|E} n^{-\Omega(\epsilon^2)}]$$

(After using (9) and the Chernoff bounds)

$$\leq \frac{E[\Delta_{\tau-1|E}]}{m(1 - 2\epsilon_1)\tau}.$$

Putting all 3 cases together, we have

$$E[\Delta_{\tau}] \leq E[\Delta_{k}] n^{\alpha_1} \prod_{\tau = k_0 + 1}^{k_0} \left( 1 + \frac{1}{(1 - 2\epsilon_1)\tau} \right) \prod_{\tau = k_0 + 1}^{k} \left( 1 + \frac{1}{(1 + \alpha)\tau} \right) \leq E[\Delta_{k}] n^{\alpha_1} \left( \frac{k_0}{k} \right) \frac{1}{(1 - 2\epsilon_1)} \left( \frac{t}{k_0} \right) \frac{1}{1 + \alpha} .$$

Now, $\Delta_k = \Delta_k(k, s, z, z') = |E_{G_k}(B_r(u)) \triangle E_{G'_k}(B_r(u))| \leq 2m$, because the graphs $G_k$ and $G'_k$ differ at most in the last $m$ edges.

Also, if $|E_{G_k}(B_r(u)) \triangle E_{G'_k}(B_r(u))| > 0$ then $u \in B_{2r}(x_k) \cup B_{2r}(x'_k)$. So

$$E[\Delta_{k|E}] \leq 2m Pr \left[ u \in B_{2r}(x_k) \cup B_{2r}(x'_k) \right] \leq 16m A_r .$$

$\square$
By applying Lemma 5, we have that for any acceptable sequence we have

$$R^2(s) = \sum_{k=1}^{t} r_k(s)^2 \leq$$

$$(16A_r m)^2 n^{-2\alpha} \left( \sum_{k=1}^{n_0} \left( \frac{n_0}{k} \right)^{\frac{3}{1-2\alpha}} \left( \frac{t}{n_0} \right)^{\frac{2}{1+\alpha}} + \sum_{k=n_0+1}^{t} \left( \frac{t}{k} \right)^{\frac{3}{1+\alpha}} \right) \Pr [E]^{-2}$$

$$\leq \left( 256A_r^2 m^2 t^{1-\alpha} n^{-2\alpha} \right) \left( \sum_{k=1}^{\infty} k^{2-2/(1-2\alpha)} + \sum_{k=1}^{\infty} k^{-2/(1-2\alpha)} \right) \Pr [E]^{-2}$$

$$= O \left( A_r^2 m^2 t^{1-\alpha} n^{-2\alpha} \right),$$

where the final equality relies on the fact that \( \Pr [E] = 1 - o(1) \), which is proved below.

Therefore, by using Eq. (8), we have

$$\Pr \left[ |D_t(u) - E[D_t(u)]| \geq A_r m t^{1-\alpha} n^{-\frac{1}{1-2\alpha}} \log n \right] \leq e^{-\Omega(\log n^2)} + \Pr [-E].$$

(12)

Now we concentrate in bounding \( \Pr [-E] \).

**Lemma 6** Let \( \alpha = 1/400 \). There is \( c > 0 \) such that \( \Pr [-E] = O(e^{-cm_0 A_0}) \).

**Proof** Let \( W \) be a set of points in \( S \) such that every point in \( S \) is at distance smaller than \( r/2 \) from \( W \). We can construct \( W \) such that \( |W| = O(1/r^2) \).

Let \( w \in W \). We are going to prove that new vertices that fall in \( B_{r/2}(w) \) are likely to choose vertices in \( B_{r/2}(w) \) with positive probability and therefore \( D_\tau^- (B_{r/2}(w)) \) is likely to be a positive proportion of \( m r \tau \).

Suppose we are in step \( \tau + 1 \) of the process. Let \( x_{\tau+1} \) be the chosen point in \( S \), and let \( y_1, \ldots, y_m = y_1', \ldots, y_m' \) be the \( m \) vertices chosen by \( x_{\tau+1} \). Then

$$\Pr [y_i \in B_{r/2}(w) \mid G_{\tau-1}, x_{\tau+1}] = \frac{D_\tau (B_{r/2}(w) \cap B_r (x_{\tau+1}))}{D_\tau (B_r (x_{\tau+1}))}. \quad (13)$$

The key of the proof is the following bound on \( D_\tau (B_R (w)) \), for any \( R > 0 \). It follows from (1) and the fact that any neighbor of a vertex in \( B_R (w) \) lies in \( B_{R+r} (w) \).

$$m |V_\tau (B_R (w))| \leq D_\tau (B_R (w))$$

$$= m |V_\tau (B_R (w))| + D_\tau^- (B_R (w)) \leq 2 m |V_\tau (B_{R+r} (w))|. \quad (14)$$

Notice that if \( x_{\tau+1} \in B_{r/2}(w) \) then \( B_{r/2}(w) \subseteq B_r (w) \cap B_r (x_{\tau+1}) \) and \( B_r (x_{\tau+1}) \subseteq B_{3r/2}(w) \). So using Eq. (13) and Eq. (14) we have

$$\Pr [y_i \in B_{r/2}(w) \mid G_{\tau-1}, x_{\tau+1} \in B_{r/2}(w)] \geq \frac{D_\tau (B_{r/2}(w))}{D_\tau (B_{3r/2}(w))} \geq \frac{|V_\tau (B_{r/2}(w))|}{2 |V_\tau (B_{3r/2}(w))|}, \quad (15)$$

10
Let \( n_1 = n_0/100 \) and let \( \mathcal{E}_1 \) be the event “for all \( s \) with \( n_1 < s \leq n \) we have \(|V_s(B_r/2(w))| \geq A_r s/6 \) and \(|V_s(B_{5r/2}(w))| \leq 8A_r s\)”. Then

\[
\Pr \left[ y_i \in B_r(w) \mid \mathcal{E}_1, x_{\tau+1} \in B_{r/2}(w) \right] \geq \frac{1}{96},
\]

and therefore,

\[
\Pr \left[ y_i \in B_r(w) \mid \mathcal{E}_1 \right] \geq \frac{A_r}{384}.
\]

Writing \( D^-_\tau(B_{r/2}(w)) = \sum_{k=1}^{\tau} \sum_{i=1}^{m} 1_{y_i \in B_{r/2}(w)} \) we get that, conditioned on \( \mathcal{E}_1 \),

\( D^-_\tau(B_{r/2}(w)) \) stochastically dominates \( X \sim \text{Bi} \left( \tau - n_1, \frac{A_r}{384} \right) \). So, if \( \tau > n_0 = 100n_1 \) then

\[
\Pr \left[ D^-_\tau(B_{r/2}(w)) < \frac{m\tau A_r}{400} \right] \leq \Pr \left[ X < \frac{m\tau A_r}{400} \right] + \Pr \left[ -\mathcal{E}_1 \right] 
\leq e^{-cm\tau A_r} + \Pr \left[ -\mathcal{E}_1 \right], \tag{16}
\]

for some \( c > 0 \).

Now we are going to bound \( \Pr \left[ -\mathcal{E}_1 \right] \). Let \( s \) be such that \( n_1 < s \leq n \). Let \( \tau \leq s \). \( \Pr \left[ x_{\tau} \in B_{r/2}(w) \right] = A_r/4 \) and \( \Pr \left[ x_{\tau} \in B_{5r/2}(w) \right] = 25A_r/4 \). Then \( V_s(B_{r/2}(w)) \sim \text{Bi}(s, A_r/4) \) and \( V_s(B_{5r/2}(w)) \sim \text{Bi}(s, 25A_r/4) \). By Chernoff’s bound

\[
\Pr \left[ V_s(B_{r/2}(w)) \leq A_r s/6 \right] \leq e^{-A_r s/72}
\]
\[
\Pr \left[ V_s(B_{5r/2}(w)) \geq 8A_r s \right] \leq e^{-A_r s/6}
\]

So,

\[
\Pr \left[ -\mathcal{E}_1 \right] \leq \sum_{\tau=n_1}^{n} \left( e^{-A_r s/72} + e^{-A_r s/6} \right) \leq ne^{-A_r n_1/72} = O \left( e^{-A_r n_1/100} \right)
\]

Now, by using (16) we see that if \( \tau > n_0 \) then

\[
\Pr \left[ D^-_\tau(B_{r/2}(w)) < \frac{m\tau A_r}{400} \right] \leq e^{-c_1 n_0 A_r}, \tag{18}
\]

for some \( c_1 > 0 \).

To extend this to all points in \( S \), note that for any \( u \in S \), there is a \( w \in W \) such that \( \| u - w \| \leq r/2 \). Therefore \( B_{r/2}(w) \subseteq B_r(u) \) and for any \( \tau, D^-_\tau(u) \geq D^-_\tau(B_{r/2}(w)) \). So

\[
\Pr \left[ -\mathcal{E} \right] = \Pr \left[ \exists \tau, n_0 \leq \tau \leq t, \exists u \in S, D^-_\tau(u) \leq m\tau r/400 \right]
\leq \Pr \left[ \exists \tau, n_0 \leq \tau \leq t, \exists w \in W, D^-_\tau(B_{r/2}(w)) \leq m\tau r/400 \right]
\leq \sum_{\tau=n_0}^{t} \sum_{w \in W} \Pr \left[ D^-_\tau(B_{r/2}(w)) \leq m\tau r/400 \right]
= O \left( m^{-2} e^{-c_1 n_0 A_r} \right) = O \left( e^{-c_2 n_0 A_r} \right),
\]

for some \( c_2 > 0 \).

Returning to (12) and taking \( \epsilon_1 \) sufficiently small, we see that there is \( c > 0 \) such that

\[
\Pr \left[ |D_t(u) - E [D_t(u)] | \geq A_r m t^{1-\gamma} \right] \leq n^{-2} + O \left( e^{-c_0 A_r} \right), \tag{19}
\]

which completes the proof of Lemma 3.
4 Connectivity

Here we are going to prove that for $r \geq n^{-1/2} \log n$, $m > K \log n$, and $K$ sufficiently large, \textbf{whp} $G_n$ is connected and has diameter $O(\log n / r)$. Notice that $G_n$ is a subgraph of the graph $G(n, r)$, the intersection graph of the caps $B_r(x_t)$, $t = 1, 2, \ldots, n$ and therefore it is disconnected for $r = o(n^{-1/2} \log n)$ [30]. We denote the diameter of $G$ by diam($G$), and follow the convention of defining diam($G$) = $\infty$, when $G$ is disconnected. In particular, when we say that a graph has finite diameter this implies it is connected.

Let $T = K_1 \log n / A_r = \Theta(n / \log n)$ where $K_1$ is sufficiently large, and $K_1 \ll K$.

**Lemma 7** Let $u \in S$ and let $B = B_{r/2}(u)$. Then

$$\Pr[\text{diam}(G_n(B)) \geq 2(K_1 + 1) \log n] = O(n^{-3})$$

where $G_n(B)$ is the induced subgraph of $G_n$ in $B$.

**Proof**

Given $k_0$ and $N$, we consider the following process which generates a sequence of graphs $H_s = (W_s, F_s)$, $s = 1, 2, \ldots, N$. (The meanings of $N, k_0$ will become apparent soon).

**Time step 1**

To initialize the process, we start with $H_1$ consisting of $k_0$ isolated vertices $y_1, \ldots, y_{k_0}$.

**Time step $s \geq 1$**

We add vertex $y_{s+k_0}$. We then add $m/4000$ random edges incident with $y_{s+k_0}$ of the form $(y_{s+k_0}, w_i)$ for $i = 1, 2, \ldots, m/4000$. Here each $w_i$ is chosen uniformly from $W_s$.

The idea is to couple the construction of $G_n$ with the construction of $H_N$ for $N \sim \text{Bi}(n - T, A_r / 4)$ and $k_0 = \text{Bi}(T, A_r / 4)$ such that \textbf{whp} $H_N$ is a subgraph of $G_n$ with vertex set $V_n(B)$. We are going to show that \textbf{whp} diam($H_N$) $\leq 2(K_1 + 1) \log n$, and therefore diam($G_n(B)$) $\leq 2(K_1 + 1) \log n$.

To do the coupling we use two counters, $t$ for the steps in $G_n$ and $s$ for the steps in $H_N$:

- Given $G_n$, set $s = 0$. Let $W_0 = V_B$. Notice $k_0 = |W_0| \sim \text{Bi}(T, A_r / 4)$ and that $k_0 \leq K_1 \log n$ \textbf{whp}.

- For every $t > T$,
  - If $x_t \notin B$, do nothing in $H_N$.
  - If $x_t \in B$, set $s := s + 1$. Set $y_{s+k_0} = x_t$. As we want $H_N$ to be a subgraph of $G_n$ we must choose the neighbors of $y_{s+k_0}$ among the neighbors of $G_n$. Let $A$ be the set of vertices chosen by $x_t$ in $V_B$. Notice that $|A|$ stochastically dominates $a_t \sim \text{Bi} \left( m, \frac{D_t(B)}{D_t(x_t)} \right)$. If $\frac{D_t(B)}{D_t(x_t)} \geq \frac{1}{100}$, then $a_t$ stochastically dominates $b_t \sim \text{Bi}(m, \frac{1}{200})$ and so \textbf{whp} is at least $m/200$. If $\frac{D_t(B)}{D_t(x_t)} < \frac{1}{100}$ we declare failure, but as we see below this is unlikely to happen — see (20). We can assume that $D_t(B) \leq 3m A_t$ and $k_t = |V_t(B)| \geq A_t t / 5$ and so each vertex of $B$ has probability at least $\frac{m}{3mA_t} \geq \frac{1}{100}$ of being chosen under preferential attachment. Thus, as insightfully observed by Bollobás and Riordan [10] we can legitimately start the addition of $x_t$ in $G_t$ by choosing $\text{Bi}(m, 1/3000)$ random neighbours uniformly in $B$. Observe that $\text{Bi}(m, 1/3000) \geq m/4000$ \textbf{whp}. 


Notice that $N$, the number of times $s$ is increased, is the number of steps for which $x_t \in B$, and so $N \sim \text{Bi}(n - T, A_r/4)$.

Notice also that, by (14), we have

$$\frac{D_t(B)}{D_t(x_t)} \geq \frac{V_t(B)}{2V_t(B_{2r}(x_t))},$$

and therefore, for $t \geq T$,

$$\Pr \left[ \frac{D_t(B)}{D_t(x_t)} \leq \frac{1}{100} \right] \leq \Pr \left[ V_t(B) \leq A_r t/6 \text{ or } V_t(B_{2r}(x_t)) \geq 8A_r t\right] \leq 2n^{-K_1/8}, \tag{20}$$

where the final inequality follows from Chernoff’s bound (see (17)).

Now we are ready to show that $H_N$ is connected \textit{whp}.

Notice that by Chernoff’s bound we get that

$$\Pr \left[ \left| k_0 - \frac{K_1}{4} \log n \right| \geq \frac{K_1}{8} \log n \right] \leq 2n^{-K_1/48}$$

and

$$\Pr \left[ N \leq \frac{1}{3} (\log n)^2 \right] \leq e^{-c (\log n)^2}$$

for some $c > 0$. Therefore, we can assume $\log n \leq k_0 \leq K_1 \log n$ and $N \geq \frac{1}{3} (\log n)^2$.

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

$$X_{s+1} = X_s - Y_s, \quad X_0 = k_0$$

where $Y_s$ is the number of components (minus one) collapsed into one by $y_{s+0}$. Then

$$\Pr [ Y_s = 0 \leq \sum_{i=1}^{X_s} \left( \frac{c_i}{s + k_0} \right)^{m/4000}$$

where the $c_i$ are the component sizes of $H_s$. Therefore, if $s < 2K_1 \log n$ then, since $m \geq K \log n$, we have

$$\Pr [ Y_s = 0 \mid X_s \geq 2 ] \leq 2 \left( 1 - \frac{1}{s + k_0} \right)^{m/4000} \leq e^{-m/(4000(s+k_0))} \leq 1/10.$$

So $X_s$ is stochastically dominated by the random variable $\max(1, k_0 - Z_s)$ where $Z_s \sim \text{Bi}(s, 9/10)$. We get then

$$\Pr [ Z_{2K_1 \log n} > 1 ] \leq \Pr [ Z_{2K_1 \log n} < k_0 ] \leq \Pr [ Z_{2K_1 \log n} < K_1 \log n ] \leq n^{-3}.$$

And therefore

$$\Pr [ H_{2K_1 \log n} \text{ is not connected} ] \leq n^{-3}.$$

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 $2K_1\log n$ 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 \( \log n \) 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 \( H_{2K_1 \log n + (i-1) \log n} \), the graph at the end of the \( i^{th} \) round.

Conditioning in \( \rho A \), every vertex in \( H_N \) is at distance at most \( \log n \) of \( H_{2K_1 \log n} \) whose diameter is not greater than \( 2K_1 \log n \). Thus, the diameter of \( H_N \) is smaller than \( 2(K_1 + 1) \log n \).

Now, we have that if \( v \) is created in the \((i + 1)^{th}\) round,

\[ \Pr [v \text{ is not adjacent to } H_{2K_1 \log n + (i-1) \log n}] \leq \left( \frac{1}{2} \right)^m. \]

Therefore

\[ \Pr [\neg \mathcal{A}] \leq \left( \frac{1}{2} \right)^m n(\log n) \leq \frac{\log n}{nK \log 2 - 1}. \]

\[ \square \]

To finish the proof of connectivity and the diameter, let \( u, v \) be two vertices of \( G_n \). Let \( C_1, C_2, \ldots, C_M, M = O(1/r) \) be a sequence of spherical caps of radius \( r/4 \) such that \( u \) is the center of \( C_1 \), \( v \) is the center of \( v \) and such that the centers of \( C_i, C_{i+1} \) are distance \( \leq r/2 \) apart. The intersections of \( C_i, C_{i+1} \) have area at least \( A_r/40 \) and so \textbf{whp} each intersection contains a vertex. Using Lemma 7 we deduce that \textbf{whp} there is a path from \( u \) to \( v \) in \( G_n \) of size at most \( O(\log n/r) \).

References


[33] 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.