The emergence of a giant component in random subgraphs of pseudo-random graphs

Alan Frieze* Michael Krivelevich† Ryan Martin‡

July 28, 2003

Abstract

Let G be a d-regular graph G on n vertices. Suppose that the adjacency matrix of G is such that the eigenvalue λ which is second largest in absolute value satisfies $\lambda = o(d)$. Let G_p with $p = \frac{\alpha}{d}$ be obtained from G by including each edge of G independently with probability p. We show that if $\alpha < 1$ then **whp** the maximum component size of G_p is $O(\log n)$ and if $\alpha > 1$ then G_p contains a unique giant component of size $\Omega(n)$, with all other components of size $O(\log n)$.

1 Introduction

Pseudo-random graphs (sometimes also called quasi-random graphs) can be informally defined as graphs whose edge distribution resembles closely that of truly random graphs on the same number of vertices and with the same edge density. Pseudo-random graphs, their constructions and properties have been a subject of intensive study for the last fifteen years (see [2], [7], [11], [10], [12], to mention just a few).

For the purposes of this paper, a pseudo-random graph is a d-regular graph $G = (V, E)$ with vertex set $V = [n] = \{1, \dots, n\}$, all of whose eigenvalues but the first one are significantly smaller than d in their absolute values. More formally, let $A = A(G)$ be the adjacency matrix of G. This is an n-by-n matrix such that $A_{ij} = 1$ if $(i, j) \in E(G)$ and $A_{ij} = 0$ otherwise. Then A is a real symmetric matrix with non-negative values of its entries. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ be the eigenvalues of A, also called the eigenvalues of G. It follows from the Perron-Frobenius theorem that $\lambda_1 = d$ and $|\lambda_i| \leq d$ for all $2 \leq i \leq n$. We thus denote

*Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh PA15213, U.S.A. Supported in part by NSF grant CCR-9818411.

†Department of Mathematics, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel. E-mail: krivelev@math.tau.ac.il. Research supported in part by USA-Israel BSF Grant 99-0013, by grant 64-01 from the Israel Science Foundation and by a Bergmann Memorial Grant.

‡Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA 15213. Supported in part by NSF VIGRE grant DMS-9819950.

$\lambda = \lambda(G) = \max_{2 \le i \le n} |\lambda_i|$. The reader is referred to a monograph of Chung [6] for further information on spectral graph theory.

It is known (see, e.g. [1]) that the greater is the so-called spectral gap (i.e. the difference between d and λ) the more tightly the distribution of the edges of G approaches that of the random graph $G(n, d/n)$. We will cite relevant quantitative results later in the text (see Lemma 1), for now we just state informally that a spectral gap ensures pseudo-randomness.

In this paper we study certain properties of a random subgraph of a pseudo-random graph. Given a graph $G = (V, E)$ and an edge probability $0 \le p = p(n) \le 1$, the *random subgraph* G_p is formed by choosing each edge of G independently and with probability p. We will also need to consider the related random graph G_m whose edge set is a random m-subset of E.

The most studied random graph is the so called binomial random graph $G(n, p)$, formed by choosing the edges of the complete graph on n labeled vertices independently with probability p. Here rather than studying random subgraphs of one particular graph, we investigate the properties of random subgraphs of graphs from a wide class of regular pseudo-random graphs. As we will see, all such subgraphs viewed as probability spaces share certain common features.

Our concern here is with the existence of a giant component in the case $p = \frac{\alpha}{d}$ or $m = \frac{1}{2}\alpha n$ where $\alpha \neq 1$ is an absolute constant. These two models are sufficiently similar so that the results we prove in G_p immediately translate to G_m and vice-versa. The needed formal relations in the case where $G = K_n$ are given in [5] or [9] and they generalise easily to our case.

As customary when studying random graphs, asymptotic conventions and notations apply. In particular, we assume where necessary the number of vertices n of the base graph G to be as large as needed. Also, we say that a graph property $\mathcal{A}$ holds *with high probability*, or **whp** for brevity, in G_p if the probability that G_p has $\mathcal{A}$ tends to 1 as n tends to infinity. Monographs [5], [9] provide a necessary background and reflect the state of affairs in the theory of random graphs.

For $\alpha > 1$ we define $\bar{\alpha} < 1$ to be the unique solution (other than α) of the equation $xe^{-x} = \alpha e^{-\alpha}$. We assume from now on that

$$d \to \infty \text{ and } \lambda = o(d). \tag{1}$$

These requirements are quite minimal.

In analogy to the classical case $G = K_n$, studied already by Erdős and Rényi [8],

Theorem 1. *Assume that* (1) *holds.*

- *(a) If* $\alpha < 1$ *then* **whp** *the maximum component size is* $O(\log n)$.
- *(b) If* $\alpha > 1$ *then* **whp** *there is a unique giant component of asymptotic size* $\left(1 - \frac{\bar{\alpha}}{\alpha}\right) n$ *and the remaining components are of size* $O(\log n)$.

One can also prove tighter results on the size and structure of the small components. They correspond nicely to the case where $G = K_n$.

We will use the notation $f(n) \gg g(n)$ to mean $f(n)/g(n) \to \infty$ with n. Similarly, $f(n) \ll g(n)$ means that $f(n)/g(n) \to 0$.

Theorem 2. *Let $\omega = \omega(n) \to \infty$ with n be arbitrary.*

(a) *If $d \gg (\log n)^2$ then **whp** G_p contains no isolated trees of size $\zeta(\log n - \frac{5}{2}\log\log n) + \omega$, where $\zeta^{-1} = \alpha - 1 - \log\alpha > 0$.*

(b) *If $d \gg \log^2 n$ then **whp** G_p contains an isolated tree of size at least $\zeta(\log n - \frac{5}{2}\log\log n) - \omega$.*

(c) *If $d = \Omega(n)$ then **whp** G_p contains $\leq \omega$ vertices on unicyclic components.*

(d) *Let $d \gg \sqrt{n}$. If $\alpha < 1$ then **whp** G_p contains no component with k vertices and with more than k edges.*

(e) *Let $d \gg \sqrt{n}$. If $\alpha > 1$ then **whp** G_p contains no component with $k = o(n)$ vertices and with more than k edges.*

2 Properties of d-regular graphs

In this section we put together those properties needed to prove Theorem 1. For $B, C \subseteq V$ let $e(B, C)$ denote the number of **ordered** pairs (u, v) such that $u \in B$, $v \in C$ and $\{u, v\} \in E$.

Lemma 1. *Suppose $B, C \subseteq V$ and $|B| = bn$ and $|C| = cn$. Then*

$$|e(B, C) - bcdn| \leq \lambda n\sqrt{bc}.$$

This is Corollary 9.2.8 of [3]. Note that $B = C$ is allowed here. Then $e(B, B)$ is twice the number of edges of G in the graph induced by B.

Now let t_k denote the number of k-vertex trees that are contained in G.

Lemma 2.

$$n\frac{k^{k-2}(d-k)^{k-1}}{k!} \leq t_k \leq n\frac{k^{k-2}d^{k-1}}{k!}$$

This is Lemma 2 of [4].

3 Proof of Theorem 1

Let $p = \frac{\alpha}{d}$ and let C_k denote the number of vertices of V that are contained in components of size k in G_p and let $T_k \leq C_k$ denote the number of vertices which are contained in isolated trees of size k.

Lemma 3.

(a)

$$\mathbf{E}C_k \leq n\frac{k^{k-1}}{k!}\alpha^{k-1}e^{-\alpha k(1-\xi_k)}$$

where

$$\xi_k = \min\left\{\frac{k}{d}, \frac{k}{n} + \frac{\lambda}{d}\right\}.$$

(b) For $k \ll d$,

$$\mathbf{E}T_k \geq n\frac{k^{k-1}}{k!}\alpha^{k-1}e^{-\alpha k(1+\eta_k)}$$

where

$$\eta_k = \frac{2k}{d} + \frac{2k}{\alpha d} + \frac{\alpha}{d}.$$

Proof

(a) Let $\mathcal{T}_k$ denote the set of trees of size k in G. Then

$$\mathbf{E}C_k \leq \sum_{T\in\mathcal{T}_k} kp^{k-1}(1-p)^{e_T}$$

where $e_T = e\left(V(T), \overline{V(T)}\right)$. Now Lemma 1 implies that

$$e_T = kd - e(V(T), V(T)) \geq a_k \stackrel{\text{def}}{=} kd - \frac{k^2 d}{n} - \lambda k \tag{2}$$

and we also have the simple inequality

$$e_T \geq b_k \stackrel{\text{def}}{=} kd - k(k-1)$$

which is true for an arbitrary d-regular graph.

Thus,

$$\mathbf{E}C_k \leq kt_k p^{k-1}(1-p)^{\max\{a_k,b_k\}} \tag{3}$$

and (a) follows from Lemma 2 and some straightforward estimations.

(b) Similarly,

$$\mathbf{E}T_k \geq \sum_{T\in\mathcal{T}_k} kp^{k-1}(1-p)^{e_T+k^2}$$

where we crudely bound by k^2, the number of edges contained in $V(T)$ which must be absent to make T an isolated tree component of G_p. Now we can simply use

$$e_T \leq kd$$

and Lemma 2. We also use $1-p \geq e^{-p-p^2}$ for p small and

$$(d-k)^{k-1} > d^{k-1}\left(1-\frac{k}{d}\right)^k \geq d^{k-1}\exp\left\{-\frac{k^2}{d}-\frac{k^3}{d^2}\right\},$$

for k/d small and make some straightforward estimations.

□

Now choose $\gamma = \gamma(\alpha)$ such that

$$\alpha e^{1-\alpha+2\alpha\gamma} = 1.$$

(Note that $\alpha e^{1-\alpha} < 1$ for $\alpha \neq 1$.)

Lemma 4. Whp, $C_k = 0$ *for* $k \in I = \left[\frac{1}{\alpha\gamma}\log n, \gamma n\right]$

Proof First assume that $k \leq \gamma d$ and observe that $\xi_k \leq \gamma$ in this range. Then from Lemma 3(a) and $k! \geq \left(\frac{k}{e}\right)^k$ we see that

$$\begin{aligned}
\sum_{k=\frac{1}{\alpha\gamma}\log n}^{\gamma d} \mathbf{E}\, C_k &\leq \frac{n}{\alpha}\sum_{k=\frac{1}{\alpha\gamma}\log n}^{\gamma d} k^{-1}(\alpha e^{1-\alpha+\alpha\xi_k})^k \\
&\leq \frac{n}{\alpha}\sum_{k=\frac{1}{\alpha\gamma}\log n}^{\gamma d} k^{-1}(\alpha e^{1-\alpha+\alpha\gamma})^k \\
&= \frac{n}{\alpha}\sum_{k=\frac{1}{\alpha\gamma}\log n}^{\gamma d} k^{-1}e^{-\alpha\gamma k} \\
&\leq \frac{\gamma n}{\log n}\sum_{k=\frac{1}{\alpha\gamma}\log n}^{\infty} e^{-\alpha\gamma k} \\
&= o(1). \qquad (4)
\end{aligned}$$

Now assume that $\gamma d \leq k \leq \gamma n$ and observe that (1) implies $\xi_k \leq \gamma + o(1)$ in this range. Then

$$\begin{aligned}
\sum_{k=\frac{1}{\alpha\gamma}\log n}^{\gamma n} \mathbf{E}\, C_k &\leq \frac{n}{\alpha}\sum_{k=\frac{1}{\alpha\gamma}\log n}^{\gamma n} k^{-1}(\alpha e^{1-\alpha+\alpha\gamma+o(1)})^k \\
&\leq \frac{n}{\alpha}\sum_{k=\frac{1}{\alpha\gamma}\log n}^{\gamma n} k^{-1}e^{-(\alpha\gamma-o(1))k} \\
&= o(1). \qquad (5)
\end{aligned}$$

□

Now let us show that there are many vertices on small isolated trees. Let

$$f(\alpha)=\sum_{k=1}^{\infty}\frac{k^{k-1}}{k!}\alpha^{k-1}e^{-\alpha k}.$$

It is known, see for example Erdős and Rényi [8] that

$$f(\alpha)=\begin{cases}1 & \alpha\leq 1.\\ \frac{\bar{\alpha}}{\alpha} & \alpha>1.\end{cases}$$

Lemma 5. *Let* $k_0=d^{1/3}$*. Then*

$$\mathbf{Pr}\left(\left|\sum_{k=1}^{k_0}C_k-nf(\alpha)\right|\geq n^{5/6}\log n\right)=o(1).$$

Proof Note that $k\xi_k=O(d^{-1/3})$ and $k\eta_k=O(d^{-1/3})$ for $k\leq k_0$. Thus from Lemma 3(a) we have

$$\mathbf{E}\sum_{k=1}^{k_0}C_k\leq(1+O(d^{-1/3}))n\sum_{k=1}^{k_0}\frac{k^{k-1}}{k!}\alpha^{k-1}e^{-\alpha k}=(1+O(d^{-1/3}))nf(\alpha).\quad(6)$$

On the other hand, Lemma 3(b) implies,

$$\mathbf{E}\sum_{k=1}^{k_0}C_k\geq\mathbf{E}\sum_{k=1}^{k_0}T_k\geq(1-O(d^{-1/3}))n\sum_{k=1}^{k_0}\frac{k^{k-1}}{k!}\alpha^{k-1}e^{-\alpha k}=(1-O(d^{-1/3}))nf(\alpha).\quad(7)$$

We now use the Azuma-Hoeffding martingale tail inequality [3] to show that the random variable $Z=\sum_{k=1}^{k_0}C_k$ is sharply concentrated. We switch to the model $G_m, m=\frac{1}{2}\alpha n$. Changing one edge can only change Z by at most $2k_0$ and so for any $t>0$

$$\mathbf{Pr}(|Z-\mathbf{E}Z|\geq t)\leq 2\exp\left\{-\frac{2t^2}{4mk_0^2}\right\}.$$

Putting $t=n^{1/2}k_0\log n$ yields the lemma, in conjunction with (6), (7) and $d\to\infty$. □

The first part of Theorem 1 now follows easily. Since $\alpha<1$ here, we have $f(\alpha)=1$ and so by Lemma 4 and Lemma 5 **whp** there are at least $n-n^{5/6}\log n$ vertices in components of size at most $\frac{1}{\alpha\gamma}\log n$. Applying Lemma 4 again, we see that **whp** the remaining vertices X must be in components of size at least γn. So if $X\neq\emptyset$ then $|X|\geq\gamma n$. But we know that **whp** $|X|\leq n^{5/6}\log n$ and so $X=\emptyset$ **whp**.

For the second part of the theorem where $\alpha>1$ we see that **whp** there are $\frac{\bar{\alpha}}{\alpha}n+O(n^{5/6}\log n)$ vertices on components of size $\leq\frac{1}{\alpha\gamma}\log n$ and the remaining vertices lie in *large* components of size at least γn. This statement remains true if we consider $G_{m-\log n}$. Let $S_1, S_2, \ldots, S_s$ be the large components of $G_{m-\log n}$, where $s \leq 1/\gamma$. We now show that **whp** adding the remaining $\log n$ random edges Y puts $S_1, S_2, \ldots, S_s$ together in one giant component of size $\left(1 - \frac{\bar{\alpha}}{\alpha}\right) n + O(n^{5/6} \log n)$. We also **whp** have $\frac{\bar{\alpha}}{\alpha} n + O(n^{5/6} \log n)$ vertices on components of size $\leq \frac{1}{\alpha\gamma} \log n$ and Lemma 4 shows that this accounts for all the vertices.

So let us show that

$$\Pi = \mathbf{Pr}(\exists\, 1 \leq i < j \leq s : \ Y \text{ contains no edge joining } S_i \text{ and } S_j) = o(1), \qquad (8)$$

completing the proof of Theorem 1. Now by Lemma 1, G contains at least $(1 - o(1))\gamma^2 dn$ edges between S_i and S_j, and the probability that Y contains none of these is at most $\left(1 - \frac{(1-o(1))\gamma^2 dn}{\frac{1}{2}dn}\right)^{\log n} \leq n^{-2\gamma^2+o(1)}$. So $\Pi \leq \gamma^{-2} n^{-2\gamma^2+o(1)} = o(1)$, proving (8). □

4 Proof of Theorem 2

Let $k_{\pm} = \zeta(\log n - \frac{5}{2}\log\log n) \pm \omega$. Let N_k denote the number of tree components of size k in G_p.

Assume that $k_- \leq k \leq \frac{1}{\alpha\gamma}\log n$. Then from *the proof of* Lemma 3, (notice k^{k-2} in place of k^{k-1}, we are counting trees, not vertices on trees),

$$\begin{aligned} \mathbf{E} N_k &\leq n\frac{k^{k-2}}{k!}\alpha^{k-1} e^{-\alpha k(1-\xi_k)} \\ &= (1+o(1))\frac{n}{\alpha k^{5/2}\sqrt{2\pi}} e^{-\zeta^{-1}k} \end{aligned} \qquad (9)$$

and when $k = o(d^{1/2})$

$$\mathbf{E} N_k = (1+o(1))\frac{n}{\alpha k^{5/2}\sqrt{2\pi}} e^{-\zeta^{-1}k} \qquad (10)$$

(a) Let γ be as in Lemma 4. Using (9),

$$\sum_{k=k_+}^{\frac{1}{\alpha\gamma}\log n} \mathbf{E} N_k = O\left(\sum_{k=k_+}^{\frac{1}{\alpha\gamma}\log n} e^{-\zeta^{-1}(\omega+k-k_+)}\right) = o(1)$$

and part (a) will follow once we verify that when $\alpha > 1$, the giant component is not a tree. However, the number of edges in the giant is asymptotically

$$\begin{aligned} \frac{\alpha n}{2} - \frac{n}{\alpha}\sum_{k=1}^{\infty}\frac{(k-1)k^{k-2}}{k!}(\alpha e^{-\alpha})^k &= \alpha n\left(\frac{1}{2} - \frac{1}{\alpha^2}\sum_{k=1}^{\infty}\frac{(k-1)k^{k-2}}{k!}(\bar{\alpha} e^{-\bar{\alpha}})^k\right) \\ &= \alpha n\left(\frac{1}{2} - \frac{\bar{\alpha}^2}{2\alpha^2}\right). \end{aligned}$$

Note that

$$\frac{n}{\bar{\alpha}}\sum_{k=1}^{\infty}\frac{(k-1)k^{k-2}}{k!}(\bar{\alpha}e^{-\bar{\alpha}})^k = \frac{\bar{\alpha}}{2}n$$

which can be seen from the fact that the LHS is asymptotically equal to the expected number of edges of $G_{n,\frac{\bar{\alpha}}{n}}$ which lie on trees. So, the ratio of edges to vertices for the giant is asymptotically equal to $\frac{\alpha+\bar{\alpha}}{2} > 1$.

(b) Now let $k = k_-$. Then from (10),

$$\mathbf{E}N_k = \Omega(e^{\zeta^{-1}\omega}) \to \infty.$$

Bounding the number of G-edges inside and between two disjoint subtrees by $3k^2$ we estimate

$$\begin{aligned}\mathbf{E}N_k^2 &\leq t_k^2 p^{2k-2}(1-p)^{2dk-3k^2}\\ &= (1+o(1))(\mathbf{E}N_k)^2\end{aligned}$$

and (b) follows from the Chebychev inequality.

(c) Let U_k denote the number of isolated unicyclic components in G_p of size k. Then

$$\begin{aligned}\sum_{k=3}^{n}\mathbf{E}(kU_k) &\leq \sum_{k=3}^{n} kt_k\binom{k}{2}p^k(1-p)^{dk-k^2}\\ &\leq (1+o(1))\frac{n}{2d}\sum_{k=3}^{n}\frac{k^{k+1}}{k!}(\alpha e^{-\alpha+o(1)})^k\\ &\leq (1+o(1))\frac{n}{2d}\sum_{k=3}^{n}\frac{k^{1/2}}{\sqrt{2\pi}}(\alpha e^{1-\alpha+o(1)})^k\\ &= O(1)\end{aligned}$$

since we are assuming that $d = \Omega(n)$ here. Part (c) follows from the Markov inequality.

(d), (e) Let $COMP_k$ denote the number of components with k vertices and at least $k+1$ edges. We can restrict our attention to $4 \leq k \leq \gamma n$ since if $\alpha < 1$ there are no larger components **whp**. Then, as in (3),

$$\begin{aligned}\mathbf{E}\sum_{k=4}^{\gamma n} COMP_k &\leq \sum_{k=4}^{\gamma n} t_k\binom{k}{2}^2 p^{k+1}(1-p)^{\max\{a_k,b_k\}}\\ &\leq (1+o(1))\frac{n\alpha}{4\sqrt{2\pi}d^2}\sum_{k=4}^{\gamma n}k^{3/2}(\alpha e^{1-\alpha+\alpha\gamma+o(1)})^k.\\ &\leq (1+o(1))\frac{n\alpha}{4\sqrt{2\pi}d^2}\sum_{k=4}^{\gamma n}k^{3/2}e^{-(\alpha\gamma-o(1))k}\\ &= o(1)\end{aligned}$$

since $n/d^2 \to 0$.

This completes the proof of (d), (e). $\square$

References

[1] N. Alon and F. R. K. Chung, Explicit construction of linear sized tolerant networks, *Discrete Mathematics* 72 (1988), 15–19.

[2] N. Alon, M. Krivelevich and B. Sudakov, List coloring of random and pseudo-random graphs, *Combinatorica* 19 (1999), 453–472.

[3] N. Alon and J. Spencer, **The probabilistic method**, Wiley, New York 1992.

[4] A. Beveridge, A.M. Frieze and C.McDiarmid, Random minimum length spanning trees in regular graphs, *Combinatorica* 18 (1998) 311-333.

[5] B. Bollobás, **Random Graphs**, (2nd Edition) Cambridge University Press (2001).

[6] F. R. K. Chung, **Spectral graph theory**, Regional Conference Series in Mathematics 92, Amer. Math. Soc., Providence, 1997.

[7] F. R. K. Chung, R. L. Graham and R. M. Wilson, Quasi-random graphs, *Combinatorica* 9 (1989), 345–362.

[8] P. Erdős and A. Rényi, On the evolution of random graphs, *Publ. Math. Inst. Hungar. Acad. Sci.* 5 (1960) 17-61.

[9] S. Janson, T. Łuczak and A. Ruciński, **Random graphs**, Wiley, New York, 2000.

[10] M. Simonovits and V. T. Sós, Hereditary extended properties, quasi-random graphs and not necessarily induced subgraphs, *Combinatorica* 17 (1997), 577–596.

[11] A. Thomason, Pseudorandom graphs, In: Random graphs'85 (Poznań 85), North-Holland Math. Stud. 144, North Holland, Amsterdam, 1987, pp. 307–331.

[12] A. Thomason, Random graphs, strongly regular graphs and pseudorandom graphs, In: Surveys in Combinatorics 1987, London Math. Soc. Lecture Note Ser. 123, Cambridge Univ. Press, Cambridge, 1987, pp. 173–195,