The giant component in a random subgraph of a given graph

Fan Chung[1] ⋆, Paul Horn[1], and Linyuan Lu[2] ⋆⋆

[1] University of California, San Diego
[2] University of South Carolina

Abstract. We consider a random subgraph G_p of a host graph G formed by retaining each edge of G with probability p. We address the question of determining the critical value p (as a function of G) for which a giant component emerges. Suppose G satisfies some (mild) conditions depending on its spectral gap and higher moments of its degree sequence. We define the second order average degree $\tilde{d}$ to be $\tilde{d} = \sum_v d_v^2/(\sum_v d_v)$ where d_v denotes the degree of v. We prove that for any $\epsilon > 0$, if $p > (1+\epsilon)/\tilde{d}$ then asymptotically almost surely the percolated subgraph G_p has a giant component. In the other direction, if $p < (1-\epsilon)/\tilde{d}$ then almost surely the percolated subgraph G_p contains no giant component.

1 Introduction

Almost all information networks that we observe are subgraphs of some host graphs that often have sizes prohibitively large or with incomplete information. A natural question is to deduce the properties that a random subgraph of a given graph must have.

We are interested in random subgraphs of G_p of a graph G, obtained as follows: for each edge in G_p we independently decide to retain the edge with probability p, and discard the edge with probability $1-p$. A natural special case of this process is the Erdős-Rényi graph model $G(n,p)$ which is the special case where the host graph is K_n. Other examples are the percolation problems that have long been studied [13, 14] in theoretical physics, mainly with the host graph being the lattice graph $\mathbb{Z}^k$. In this paper, we consider a general host graph, an example of which being a contact graph, consisting of edges formed by pairs of people with possible contact, which is of special interest in the study of the spread of infectious diseases or the identification of community in various social networks.

A fundamental question is to ask for the critical value of p such that G_p has a giant connected component, that is a component whose volume is a positive fraction of the total volume of the graph. For the spread of disease on contact networks, the answer to this question corresponds to the problem of finding the epidemic threshold for the disease under consideration, for instance.

For the case of K_n, Erdős and Rényi answered this in their seminal paper [11]: if $p = \frac{c}{n}$ for $c < 1$, then almost surely G contains no giant connected component and all components are of size at most $O(\log n)$, and if $c > 1$ then, indeed, there is a giant component of size ϵn. For general host

⋆ This author was supported in part by NSF grant ITR 0426858 and ONR MURI 2008-2013
⋆⋆ This author was supported in part by NSF grant DMS 0701111

graphs, the answer has been more elusive. Results have been obtained either for very dense graphs or bounded degree graphs. Bollobas, Borgs, Chayes and Riordan [4] showed that for dense graphs (where the degrees are of order $\Theta(n)$), the giant component threshold is $1/\rho$ where ρ is the largest eigenvalue of the adjacency matrix. Frieze, Krivelevich and Martin [12] consider the case where the host graph is d-regular with adjacency eigenvalue λ and they show that the critical probability is close to $1/d$, strengthening earlier results on hypercubes [2, 3] and Cayley graphs [15]. For expander graphs with degrees bounded by d, Alon, Benjamini and Stacey [1] proved that the percolation threshold is greater than or equal to $1/(2d)$.

There are several recent papers, mainly in studying percolation on special classes of graphs, which have gone further. Their results nail down the precise critical window during which component sizes grow from $\log(n)$ vertices to a positive proportion of the graph. In [5, 6], Borgs, et. al. find the order of this critical window for transitive graphs, and cubes. Nachmias [16] looks at a similar situation to that of Frieze, Krivelevich and Martin [12] and uses random walk techniques to study percolation within the critical window for quasi-random transitive graphs. Percolation within the critical window on random regular graphs is also studied by Nachmias and Peres in [17]. Our results differ from these in that we study percolation on graphs with a much more general degree sequence. The greater preciseness of these results, however, is quite desirable. It is an interesting open question to describe the precise scaling window for percolation for the more general graphs studied here.

Here, we are interested in percolation on graphs which are not necessarily regular, and can be relatively sparse (i.e., $o(n^2)$ edges.) Compared with earlier results, the main advantage of our results is the ability to handle general degree sequences. To state our results, we give a few definitions here. For a subset S of vertices, the volume of S, denoted by $\mathrm{vol}(S)$ is the sum of degrees of vertices in S. The kth order volume of S is the kth moment of the degree sequence, i.e. $\mathrm{vol}_k(S) = \sum_{v \in S} d_v^k$. We write $\mathrm{vol}_1(S) = \mathrm{vol}(S)$ and $\mathrm{vol}_k(G) = \mathrm{vol}_k(V(G))$, where $V(G)$ is the vertex set of G. We denote by $\tilde{d} = \mathrm{vol}_2(G)/\mathrm{vol}(G)$ the second order degree of G, and by σ the spectral gap of the normalized Laplacian, which we fully define in Section 2. Further, recall that $f(n)$ is $O(g(n))$ if $\limsup_{n\to\infty} |f(n)|/|g(n)| < \infty$, and $f(n)$ is $o(g(n))$ if $\lim_{n\to\infty} |f(n)|/|g(n)| = 0$.

We will prove the following

Theorem 1. *Suppose G has the maximum degree Δ satisfying $\Delta = o(\frac{\tilde{d}}{\sigma})$. For $p \le \frac{1-c}{\tilde{d}}$, a.a.s. every connected component in G_p has volume at most $O(\sqrt{\mathrm{vol}_2(G)}g(n))$, where $g(n)$ is any slowly growing function as $n \to \infty$.*

Here, an event occurring a.a.s. indicates that it occurs with probability tending to one as n tends to infinity. In order to prove the emergence of giant component where $p \ge (1+c)/\tilde{d}$, we need to consider some additional conditions. Suppose there is a set U satisfying

(i) $\mathrm{vol}_2(U) \ge (1-\epsilon)\mathrm{vol}_2(G)$.
(ii) $\mathrm{vol}_3(U) \le M d \mathrm{vol}_2(G)$

where ϵ and M are constant independent of n. In this case, we say G is (ϵ, M)-admissible and U is an (ϵ, M)-admissible set.

We note that the admissibility measures of skewness of the degree sequence. For example, all regular graphs are $(\epsilon, 1)$-admissible for any ϵ, but a graph needs not be regular to be admissible. We also note that in the case that $\text{vol}_3(G) \leq M d \text{vol}_2(G)$, G is (ϵ, M)-admissible for any ϵ.

Theorem 2. *Suppose $p \geq \frac{1+c}{\tilde{d}}$ for some $c \leq \frac{1}{20}$. Suppose G satisfies $\Delta = o(\frac{\tilde{d}}{\sigma})$, $\Delta = o(\frac{d\sqrt{n}}{\log n})$ and $\sigma = o(n^{-\kappa})$ for some $\kappa > 0$, and G is $(\frac{c\kappa}{10}, M)$-admissible. Then a.a.s. there is a unique giant connected component in G_p with volume $\Theta(\text{vol}(G))$, and no other component has volume more than $\max(2d \log n, \omega(\sigma\sqrt{\text{vol}(G)}))$.*

Here, recall that $f(n) = \Theta(g(n))$ if $f(n) = O(g(n))$ and $g(n) = O(f(n))$. In this case, we say that f and g are of the same order. Also, $f(n) = \omega(g(n))$ if $g(n) = o(f(n))$.

We note that under the assumption that the maximum degree Δ of G satisfying $\Delta = o(\frac{\tilde{d}}{\sigma})$, it can be shown that the spectral norm of the adjacency matrix satisfies $\|A\| = \rho = (1 + o(1))\tilde{d}$. Under the assumption in Theorem 2, we observe that the percolation threshold of G is $\frac{1}{\tilde{d}}$.

To examine when the conditions of Theorems 1 and 2 are satisfied, we note that admissibility implies that $\tilde{d} = \Theta(d)$, which essentially says that while there can be some vertices with degree much higher than d, there cannot be too many. Chung, Lu and Vu [8] show that for random graphs with a given expected degree sequence $\sigma = O(\frac{1}{\sqrt{d}})$, and hence for graphs with average degree n^ϵ the spectral condition of Theorem 2 easily holds for random graphs. The results here can be viewed as a generalization of the result of Frieze, Krivelevich and Martin [12] with general degree sequences and is also a strengthening of the original results of Erdős and Reyni to general host graphs.

The paper is organized as follows: In Section 2 we introduce the notation and some basic facts. In Section 3, we examine several spectral lemmas which allow us to control the expansion. In Section 4, we prove Theorem 1, and in Section 5, we complete the proof of Theorem 2.

2 Preliminaries

Suppose G is a connected graph on vertex set V. Throughout the paper, G_p denotes a random subgraph of G obtained by retaining each edge of G independently with probability p.

Let $A = (a_{uv})$ denote the adjacency matrix of G, defined by

$$a_{uv} = \begin{cases} 1 & \text{if } \{u, v\} \text{ is an edge;} \\ 0 & \text{otherwise.} \end{cases}$$

We let $d_v = \sum_u a_{uv}$ denote the degree of vertex v. Let $\Delta = \max_v d_v$ denote the maximum degree of G and $\delta = \min_v d_v$ denote the minimum degree. For each vertex set S and a positive integer k, we define the k-th volume of G to be

$$\mathrm{vol}_k(S) = \sum_{v \in S} d_v^k.$$

The volume $\mathrm{vol}(G)$ is simply the sum of all degrees, i.e. $\mathrm{vol}(G) = \mathrm{vol}_1(G)$. We define the average degree $d = \frac{1}{n}\mathrm{vol}(G) = \frac{\mathrm{vol}_1(G)}{\mathrm{vol}_0(G)}$ and the second order average degree $\tilde{d} = \frac{\mathrm{vol}_2(G)}{\mathrm{vol}_1(G)}$.

Let $D = diag(d_{v_1}, d_{v_2}, \ldots, d_{v_n})$ denote the diagonal degree matrix. Let $\mathbf{1}$ denote the column vector with all entries 1 and $\mathbf{d} = D\mathbf{1}$ be column vector of degrees. The normalized Laplacian of G is defined as

$$\mathcal{L} = I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}}.$$

The spectrum of the Laplacian is the eigenvalues of $\mathcal{L}$ sorted in increasing order.

$$0 = \lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_{n-1}.$$

Many properties of λ_i's can be found in [7]. For example, the least eigenvalue λ_0 is always equal to 0. We have $\lambda_1 > 0$ if G is connected and $\lambda_{n-1} \leq 2$ with equality holding only if G has a bipartite component. Let $\sigma = \max\{1 - \lambda_1, \lambda_{n-1} - 1\}$. Then $\sigma < 1$ if G is connected and non-bipartite. For random graphs with a given expected degree sequence [8], $\sigma = O(\frac{1}{\sqrt{d}})$, and in general for regular graphs it is easy to write σ in terms of the second largest eigenvalue of the adjacency matrix. Furthermore, σ is closely related to the mixing rate of random walks on G, see e.g. [7].

The following lemma measures the difference of adjacency eigenvalue and $\tilde{d}$ using σ.

Lemma 1. *The largest eigenvalue of the adjacency matrix of G, ρ, satisfies*

$$|\rho - \tilde{d}| \leq \sigma \Delta.$$

Proof: Recall that $\varphi = \frac{1}{\sqrt{\mathrm{vol}(G)}} D^{1/2}\mathbf{1}$ is the unit eigenvector of $\mathcal{L}$ corresponding to eigenvalue 0. We have

$$\|I - \mathcal{L} - \varphi\varphi^*\| \leq \sigma.$$

Then,

$$\begin{aligned} |\rho - \tilde{d}| = \left| \|A\| - \|\frac{\mathbf{d}\mathbf{d}^*}{\mathrm{vol}(G)}\| \right| &\leq \|A - \frac{\mathbf{d}\mathbf{d}^*}{\mathrm{vol}(G)}\| \\ &= \|D^{1/2}(I - \mathcal{L} - \varphi\varphi^*)D^{1/2}\| \\ &\leq \|D^{1/2}\| \cdot \|I - \mathcal{L} - \varphi\varphi^*\| \cdot \|D^{1/2}\| \leq \sigma\Delta. \end{aligned}$$

□

For any subset of the vertices, S, we let let $\bar{S}$ denote the complement set of S. The vertex boundary of S in G, denoted by $\Gamma^G(S)$ is defined as follows:

$$\Gamma^G(S) = \{u \notin S \mid \exists v \in S \text{ such that } \{u,v\} \in E(G)\}.$$

When S consists of one vertex v, we simply write $\Gamma^G(v)$ for $\Gamma^G(\{v\})$. We also write $\Gamma(S) = \Gamma^G(S)$ if there is no confusion.

Similarly, we define $\Gamma^{G_p}(S)$ to be the set of neighbors of S in our percolated subgraph G_p.

3 Several spectral lemmas

We begin by proving two lemmas, first relating expansion in G to the spectrum of G, then giving a probabilistic bound on the expansion in G_p

Lemma 2. *For two disjoint sets S and T, we have*

$$\left|\sum_{v\in T} d_v|\Gamma(v)\cap S| - \frac{\mathrm{vol}(S)\mathrm{vol}_2(T)}{\mathrm{vol}(G)}\right| \leq \sigma\sqrt{\mathrm{vol}(S)\mathrm{vol}_3(T)}.$$

$$\left|\sum_{v\in T} d_v|\Gamma(v)\cap S|^2 - \frac{\mathrm{vol}(S)^2\mathrm{vol}_3(T)}{\mathrm{vol}(G)^2}\right| \leq \sigma^2\mathrm{vol}(S)\max_{v\in T}\{d_v^2\} + 2\sigma\frac{\sqrt{\mathrm{vol}(S)^3\mathrm{vol}_5(T)}}{\mathrm{vol}(G)}.$$

Proof: Let $\mathbf{1}_S$ (or $\mathbf{1}_T$) be the indicative column vector of the set S (or T) respectively. Note

$$\sum_{v\in T} d_v|\Gamma(v)\cap S| = \mathbf{1}_S^* AD\mathbf{1}_T.$$

$$\mathrm{vol}(S) = \mathbf{1}_S^*\mathbf{d}.$$

$$\mathrm{vol}_2(T) = \mathbf{d}^* D\mathbf{1}_T.$$

Here $\mathbf{1}_S^*$ denotes the transpose of $\mathbf{1}_S$ as a row vector. We have

$$\begin{aligned}
&\left|\sum_{v\in T} d_v|\Gamma(v)\cap S| - \frac{\mathrm{vol}(S)\mathrm{vol}_2(T)}{\mathrm{vol}(G)}\right| \\
&= |\mathbf{1}_S^* AD\mathbf{1}_T - \frac{1}{\mathrm{vol}(G)}\mathbf{1}_S^*\mathbf{d}\mathbf{d}^* D\mathbf{1}_T| \\
&= |\mathbf{1}_S^* D^{\frac{1}{2}}(D^{-\frac{1}{2}}AD^{-\frac{1}{2}} - \frac{1}{\mathrm{vol}(G)}D^{\frac{1}{2}}\mathbf{1}\mathbf{1}^* D^{\frac{1}{2}})D^{\frac{3}{2}}\mathbf{1}_T|
\end{aligned}$$

Let $\varphi = \frac{1}{\sqrt{\mathrm{vol}(G)}}D^{1/2}\mathbf{1}$ denote the eigenvector of $I-\mathcal{L}$ for the eigenvalue 1. The matrix $I-\mathcal{L}-\varphi\varphi^*$, which is the projection of $I-\mathcal{L}$ to the hyperspace $\varphi^\perp$, has L_2-norm σ.

We have

$$\left|\sum_{v\in T} d_v|\Gamma(v)\cap S| - \frac{\mathrm{vol}(S)\mathrm{vol}_2(T)}{\mathrm{vol}(G)}\right| = |\mathbf{1}_S^* D^{\frac{1}{2}}(I-\mathcal{L}-\varphi\varphi^*)D^{\frac{3}{2}}\mathbf{1}_T|$$
$$\leq \sigma\|D^{\frac{1}{2}}\mathbf{1}_S\|\cdot\|D^{\frac{3}{2}}\mathbf{1}_T\|$$
$$\leq \sigma\sqrt{\mathrm{vol}(S)\mathrm{vol}_3(T)}.$$

Let e_v be the column vector with v-th coordinate 1 and 0 else where. Then $|\Gamma^G(v)\cap S| = \mathbf{1}_S^* A e_v$. We have

$$\sum_{v\in T} d_v|\Gamma^G(v)\cap S|^2 = \sum_{v\in T} d_v \mathbf{1}_S^* A e_v e_v^* A\mathbf{1}_S = \mathbf{1}_S^* A D_T A \mathbf{1}_S.$$

Here $D_T = \sum_{v\in T} d_v e_v e_v^*$ is the diagonal matrix with degree entry at vertex in T and 0 else where. We have

$$\left|\sum_{v\in T} d_v|\Gamma^G(v)\cap S|^2 - \frac{\mathrm{vol}(S)^2\mathrm{vol}_3(T)}{\mathrm{vol}(G)^2}\right|$$
$$= |\mathbf{1}_S^* A D_T A\mathbf{1}_S - \frac{1}{\mathrm{vol}(G)^2}\mathbf{1}_S^*\mathbf{d}\mathbf{d}^* D_T \mathbf{d}\mathbf{d}^*\mathbf{1}_S|$$
$$\leq |\mathbf{1}_S^* A D_T A\mathbf{1}_S - \frac{1}{\mathrm{vol}(G)}\mathbf{1}_S^*\mathbf{d}\mathbf{d}^* D_T A\mathbf{1}_S|$$
$$+|\frac{1}{\mathrm{vol}(G)}\mathbf{1}_S^*\mathbf{d}\mathbf{d}^* D_T A\mathbf{1}_S - \frac{1}{\mathrm{vol}(G)^2}\mathbf{1}_S^*\mathbf{d}\mathbf{d}^* D_T d\mathbf{d}^*\mathbf{1}_S|$$
$$= |\mathbf{1}_S D^{\frac{1}{2}}(I-\mathcal{L}-\varphi\varphi^*)D^{\frac{1}{2}} D_T A\mathbf{1}_S|$$
$$+|\frac{1}{\mathrm{vol}(G)}\mathbf{1}_S^*\mathbf{d}\mathbf{d}^* D_T D^{\frac{1}{2}}(I-\mathcal{L}-\varphi\varphi^*)D^{\frac{1}{2}}\mathbf{1}_S|$$
$$\leq |\mathbf{1}_S D^{\frac{1}{2}}(I-\mathcal{L}-\varphi^*\varphi)D^{\frac{1}{2}} D_T D^{\frac{1}{2}}(I-\mathcal{L}-\varphi\varphi^*)D^{\frac{1}{2}}\mathbf{1}_S|$$
$$+2|\frac{1}{\mathrm{vol}(G)}\mathbf{1}_S^*\mathbf{d}\mathbf{d}^* D_T D^{\frac{1}{2}}(I-\mathcal{L}-\varphi\varphi^*)D^{\frac{1}{2}}\mathbf{1}_S|$$
$$\leq \sigma^2\mathrm{vol}(S)\max_{v\in T}\{d_v^2\} + 2\sigma\frac{\sqrt{\mathrm{vol}(S)^3\mathrm{vol}_5(T)}}{\mathrm{vol}(G)}.$$

□

Lemma 3. *Suppose that two disjoint sets S and T satisfy*

$$\mathrm{vol}_2(T) \geq \frac{5p}{2\delta}\sigma^2\max_{v\in T}\{d_v^2\}\mathrm{vol}(G) \quad (1)$$

$$\frac{25\sigma^2\mathrm{vol}_3(T)\mathrm{vol}(G)^2}{\delta^2\mathrm{vol}_2(T)^2} \leq \mathrm{vol}(S) \leq \frac{2\delta\mathrm{vol}_2(T)\mathrm{vol}(G)}{5p\mathrm{vol}_3(T)} \quad (2)$$

$$\mathrm{vol}(S) \le \frac{\delta^2 \mathrm{vol}_2(T)^2}{25p^2\sigma^2\mathrm{vol}_5(T)}. \tag{3}$$

Then we have that

$$\mathrm{vol}(\Gamma^{G_p}(S) \cap T) > (1-\delta)p\frac{\mathrm{vol}_2(T)}{\mathrm{vol}(G)}\mathrm{vol}(S).$$

with probability at least $1 - \exp\left(-\frac{\delta(1-\delta)p\mathrm{vol}_2(T)\mathrm{vol}(S)}{10\Delta\mathrm{vol}(G)}\right)$.

Proof: For any $v \in T$, let X_v be the indicative random variable for $v \in \Gamma^{G_p}(S)$. We have

$$\mathbb{P}(X_v = 1) = 1 - (1-p)^{|\Gamma^G(v)\cap S|}.$$

Let $X = |\Gamma^{G_p}(S) \cap T|$. Then X is the sum of independent random variables X_v.

$$X = \sum_{v\in T} d_v X_v.$$

Note that

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{v\in T} d_v \mathbb{E}(X_v) \\
&= \sum_{v\in T} d_v (1 - (1-p)^{|\Gamma^G(v)\cap S|}) \\
&\ge \sum_{v\in T} d_v (p|\Gamma^G(v)\cap S| - \frac{p^2}{2}|\Gamma^G(v)\cap S|^2) \\
&\ge p(\frac{\mathrm{vol}(S)\mathrm{vol}_2(T)}{\mathrm{vol}(G)} - \sigma\sqrt{\mathrm{vol}(S)\mathrm{vol}_3(T)}) \\
&\quad - \frac{p^2}{2}(\frac{\mathrm{vol}(S)^2\mathrm{vol}_3(T)}{\mathrm{vol}(G)^2} + \sigma^2\mathrm{vol}(S)\max_{v\in T}\{d_v^2\} + 2\sigma\frac{\sqrt{\mathrm{vol}(S)^3\mathrm{vol}_5(T)}}{\mathrm{vol}(G)}) \\
&> (1 - \frac{4}{5}\delta)p\frac{\mathrm{vol}_2(T)}{\mathrm{vol}(G)}\mathrm{vol}(S)
\end{aligned}$$

by using Lemma 2 and the assumptions on S and T.

We apply the following Chernoff inequality, see e.g. [10]

$$\mathbb{P}(X \le \mathbb{E}(X) - a) \le e^{-\frac{a^2}{2\sum d_v^2\mathbb{E}[\mathrm{X}_v^2]}} \le e^{-\frac{a^2}{2\Delta\mathbb{E}(X)}}.$$

We set $a = \alpha\mathbb{E}(X)$, with α chosen so that $(1-\alpha)(1-\frac{4}{5}\delta) = (1-\delta)$. Then

$$\begin{aligned}
\mathbb{P}(X \le (1-\delta)p\frac{\mathrm{vol}_2(T)}{\mathrm{vol}(G)}\mathrm{vol}(S)) &< \mathbb{P}(X \le (1-\alpha)\mathbb{E}(X)) \\
&\le \exp\left(-\frac{\alpha^2\mathbb{E}(\mathrm{X})}{2\Delta}\right) < \exp\left(-\frac{\alpha(1-\delta)p\mathrm{vol}_2(T)\mathrm{vol}(S)}{2\Delta\mathrm{vol}(G)}\right).
\end{aligned}$$

To complete the proof, note $\alpha > \delta/5$. □

4 The range of p with no giant component

In this section, we will prove Theorem 1.

Proof of Theorem 1: It suffices to prove the following claim.

Claim A: If $p\rho < 1$, where ρ is the largest eigenvalue of the adjacency matrix, with probability at least $1 - \frac{1}{C^2(1-p\rho)}$, all components have volume at most $C\sqrt{\mathrm{vol}_2(G)}$.

Proof of Claim A: Let x be the probability that there is a component of G_p having volume greater than $C\sqrt{\mathrm{vol}_2(G)}$. Now we choose two random vertices with the probability of being chosen proportional to their degrees in G. Under the condition that there is a component with volume greater than $C\sqrt{\mathrm{vol}_2(G)}$, the probability of each vertex in this component is at least $\frac{C\sqrt{\mathrm{vol}_2(G)}}{\mathrm{vol}(G)}$. Therefore, the probability that the random pair of vertices are in the same component is at least

$$x\left(\frac{C\sqrt{\mathrm{vol}_2(G)}}{\mathrm{vol}(G)}\right)^2 = \frac{C^2 x\tilde{d}}{\mathrm{vol}(G)}. \tag{4}$$

On the other hand, for any fixed pair of vertices u and v and any fixed path P of length k in G, the probability that u and v is connected by this path in G_p is exactly p^k. The number of k-paths from u to v is at most $\mathbf{1}_u^* A^k \mathbf{1}_v$. Since the probabilities of u and v being selected are $\frac{d_u}{\mathrm{vol}(G)}$ and $\frac{d_v}{\mathrm{vol}(G)}$ respectively, the probability that the random pair of vertices are in the same connected component is at most

$$\sum_{u,v} \frac{d_u}{\mathrm{vol}(G)} \frac{d_v}{\mathrm{vol}(G)} \sum_{k=0}^{n} p^k \mathbf{1}_u^* A^k \mathbf{1}_v = \sum_{k=0}^{n} \frac{1}{\mathrm{vol}(G)^2} p^k \mathbf{d}^* A^k \mathbf{d}.$$

We have

$$\sum_{k=0}^{n} \frac{1}{\mathrm{vol}(G)^2} p^k \mathbf{d}^* A^k \mathbf{d} \le \sum_{k=0}^{\infty} \frac{p^k \rho^k \mathrm{vol}_2(G)}{\mathrm{vol}(G)^2} \le \frac{\tilde{d}}{(1-p\rho)\mathrm{vol}(G)}.$$

Combining with (4), we have $\frac{C^2 x\tilde{d}}{\mathrm{vol}(G)} \le \frac{\tilde{d}}{(1-p\rho)\mathrm{vol}(G)}$. which implies $x \le \frac{1}{C^2(1-p\rho)}$.

Claim A is proved, and the theorem follows taking C to be $g(n)$. □

5 The emergence of the giant component

Lemma 4. *Suppose G contains an (ϵ, M)-admissible set U. Then we have*

1. $\tilde{d} \le \frac{M}{(1-\epsilon)^2} d$.
2. *For any $U' \subset U$ with $\mathrm{vol}_2(U') > \eta \mathrm{vol}_2(U)$, we have*

$$\mathrm{vol}(U') \ge \frac{\eta^2(1-\epsilon)\tilde{d}}{Md} \mathrm{vol}(G).$$

Proof: Since G is (ϵ, M)-admissible, we have a set U satisfying

(i) $\mathrm{vol}_2(U) \geq (1-\epsilon)\mathrm{vol}_2(G)$
(ii) $\mathrm{vol}_3(U) \leq Md\mathrm{vol}_2(G)$.

We have

$$\tilde{d} = \frac{\mathrm{vol}_2(G)}{\mathrm{vol}(G)} \leq \frac{\mathrm{vol}_2(G)}{\mathrm{vol}(U)} \leq \frac{1}{1-\epsilon}\frac{\mathrm{vol}_2(U)}{\mathrm{vol}(U)} \leq \frac{1}{1-\epsilon}\frac{\mathrm{vol}_3(U)}{\mathrm{vol}_2(U)} \leq \frac{M}{(1-\epsilon)^2}d.$$

For any $U' \subset U$ with $\mathrm{vol}_2(U') > \eta\mathrm{vol}_2(U)$, we have

$$\mathrm{vol}(U') \geq \frac{\mathrm{vol}_2(U')^2}{\mathrm{vol}_3(U')} \geq \frac{\eta^2\mathrm{vol}_2(U)^2}{\mathrm{vol}_3(U)} \geq \frac{\eta^2(1-\epsilon)\mathrm{vol}_2(G)}{Md} \geq \frac{\eta^2(1-\epsilon)\tilde{d}}{Md}\mathrm{vol}(G).$$

□

Proof of Theorem 2: It suffices to assume $p = \frac{1+c}{\tilde{d}}$ for some $c < \frac{1}{20}$.

Let $\epsilon = \frac{c\kappa}{10}$ be a small constant, and U be a (ϵ, M)-admissible set in G. Define U' to be the subset of U containing all vertices with degree at least $\sqrt{\epsilon}d$. We have

$$\mathrm{vol}_2(U') \geq \mathrm{vol}_2(U) - \sum_{d_v < \sqrt{\epsilon}d} d_v^2 \geq (1-\epsilon)\mathrm{vol}_2(G) - \epsilon n d^2 \geq (1-2\epsilon)\mathrm{vol}_2(G).$$

Hence, U' is a $(2\epsilon, M)$ admissible set. We will concentrate on the neighborhood expansion within U'.

Let $\delta = \frac{c}{2}$ and $C = \frac{25M}{\delta^2(1-4\epsilon)^2}$. Take an initial set $S_0 \subset U'$ with $\max(C\sigma^2\mathrm{vol}(G), \Delta\ln n) \leq \mathrm{vol}(S_0) \leq \max(C\sigma^2\mathrm{vol}(G), \Delta\ln n) + \Delta$.

Let $T_0 = U' \setminus S_0$. For $i \geq 1$, we will recursively define $S_i = \Gamma^{G_p}(S_{i-1}) \cap U'$ and $T_i = U' \setminus \cup_{j=0}^{i} S_j$ until $\mathrm{vol}_2(T_i) \leq (1-3\epsilon)\mathrm{vol}_2(G)$ or $\mathrm{vol}(S_i) \geq \frac{2\delta\mathrm{vol}_2(T_i)\mathrm{vol}(G)}{5p\mathrm{vol}_3(T_i)}$.

Condition 1 in Lemma 3 is always satisfied.

$$\begin{aligned}\frac{5p}{2\delta}\sigma^2 \max_{v\in T_i} d_v^2\mathrm{vol}(G) \leq \frac{5(1+c)}{2\tilde{d}\delta}\sigma^2\Delta^2\mathrm{vol}(G) &= (\frac{\sigma\Delta}{\tilde{d}})^2\frac{5(1+c)}{2\delta}\mathrm{vol}_2(G)\\ &= o(\mathrm{vol}_2(G)) \leq \mathrm{vol}_2(T_i).\end{aligned}$$

Condition 3 in Lemma 3 is also trivial because

$$\begin{aligned}\frac{\delta^2\mathrm{vol}_2(T_i)^2}{25p^2\sigma^2\mathrm{vol}_5(T_i)} \geq \frac{\delta^2\mathrm{vol}_2(T_i)^2}{25p^2\sigma^2\Delta^2\mathrm{vol}_3(T_i)} &\geq \frac{\delta^2(1-3\epsilon)\mathrm{vol}_2(G)}{25p^2\sigma^2\Delta^2 Md}\\ &\geq (\frac{\tilde{d}}{\sigma\Delta})^2\frac{\delta^2(1-3\epsilon)}{25(1+c)^2M}\mathrm{vol}(G) = \omega(\mathrm{vol}(G)).\end{aligned}$$

Now we verify condition 2. We have

$$\mathrm{vol}(S_0) > C\sigma^2 \mathrm{vol}(G) = \frac{25M}{\delta^2(1-4\epsilon)^2}\sigma^2 \mathrm{vol}(G) \geq \frac{25\sigma^2 \mathrm{vol}_3(T_0)\mathrm{vol}(G)^2}{\delta^2 \mathrm{vol}_2(T_0)^2}.$$

The conditions of Lemma 3 are all satisfied. Then we have that

$$\mathrm{vol}(\Gamma^{G_p}(S_0) \cap T_0) > (1-\delta)p\frac{\mathrm{vol}_2(T_0)}{\mathrm{vol}(G)}\mathrm{vol}(S_0).$$

with probability at least $1 - \exp\left(-\frac{\delta(1-\delta)p\mathrm{vol}_2(T_0)\mathrm{vol}(S_0)}{10\Delta\mathrm{vol}(G)}\right)$.

Since $(1-\delta)p\frac{\mathrm{vol}_2(T_i)}{\mathrm{vol}(G)} \geq (1-\delta)(1-3\epsilon)(1+c) = \beta > 1$ by our assumption that c is small (noting that ϵ and δ are functions of c), the neighborhood of S_i grows exponentially, allowing condition 2 of Lemma 3 to continue to hold and us to continue the process. We stop when one of the following two events happens,

- $\mathrm{vol}(S_i) \geq \frac{2\delta\mathrm{vol}_2(T_i)\mathrm{vol}(G)}{5p\mathrm{vol}_3(T_i)}$.
- $\mathrm{vol}_2(T_i) \leq (1-3\epsilon)\mathrm{vol}_2(G)$.

Let us denote the time that this happens by t.

If the first, but not the second, case occurs we have

$$\mathrm{vol}(S_t) \geq \frac{2\delta\mathrm{vol}_2(T_t)\mathrm{vol}(G)}{5p\mathrm{vol}_3(T_t)} \geq \frac{2\delta(1-3\epsilon)}{5M(1+c)}\mathrm{vol}(G).$$

In the second case, we have

$$\mathrm{vol}_2(\bigcup_{j=0}^{t} S_j) = \mathrm{vol}_2(U') - \mathrm{vol}_2(T_t) \geq \epsilon\mathrm{vol}_2(G) \geq \epsilon\mathrm{vol}(U').$$

By Lemma 4 with $\eta = \epsilon$, we have $\mathrm{vol}(\bigcup_{j=0}^{t} S_j) \geq \frac{\epsilon^2(1-2\epsilon)\tilde{d}}{Md}\mathrm{vol}(G)$. On the other hand, note that since $\mathrm{vol}(S_i) \geq \beta\mathrm{vol}(S_{i-1})$, we have that $\mathrm{vol}(S_i) \leq \beta^{i-t}\mathrm{vol}(S_t)$, and hence we have $\mathrm{vol}(\bigcup_{j=0}^{t} S_j) \leq \sum_{j=0}^{t}\beta^{-j}\mathrm{vol}(S_t)$ so

$$\mathrm{vol}(S_t) \geq \frac{\epsilon^2(1-2\epsilon)\tilde{d}(\beta-1)}{Md\beta}\mathrm{vol}(G).$$

In either case we have $\mathrm{vol}(S_t) = \Theta(\mathrm{vol}(G))$. For the moment, we restrict ourselves to the case where $C\sigma^2 n > \Delta \ln n$.

Each vertex in S_t is in the same component as some vertex in S_0, which has size at most $\frac{\mathrm{vol}(S_0)}{\sqrt{\epsilon d}} \leq C'\sigma^2 n$. We now combine the k_1 largest components to form a set $W^{(1)}$ with $\mathrm{vol}(W^{(1)}) > C\sigma^2\mathrm{vol}(G)$, such that k_2 is minimal. If $k_1 \geq 2$, $\mathrm{vol}(W^{(1)}) \leq 2C\sigma^2\mathrm{vol}(G)$. Note that since the average size of a component is $\frac{\mathrm{vol}(S_t)}{|S_0|} \geq C_1\frac{\mathrm{vol}(G)}{\sigma^2 n}$, $k_1 \leq C_1'\sigma^4 n$.

We grow as before: Let $W_0^{(1)} = W^{(1)}$, $Q_0^{(1)} = T_{t-1} \setminus W_0^{(1)}$. Note that the conditions for Lemma 3 are satisfied by $W_0^{(1)}$ and $T_0^{(1)}$. We run the process as before, setting $W_t^{(1)} = \Gamma(W_t^{(1)}) \cap Q_{t-1}^{(1)}$ and $Q_t^{(1)} = Q_{t-1}^{(1)} \setminus W_t^{(1)}$ stopping when either $\mathrm{vol}(Q_t^{(1)}) < (1-4\epsilon)\mathrm{vol}_2(G)$ or $\mathrm{vol}(W_t^{(1)}) > \frac{2\delta\mathrm{vol}_2(Q_t^{(1)})\mathrm{vol}(G)}{5p\mathrm{vol}_3(Q_t^{(1)})} \geq \frac{2\delta(1-4\epsilon)}{5M(1+c)}\mathrm{vol}(G)$. As before, in either case $\mathrm{vol}(W_t^{(1)}) = \Theta(\mathrm{vol}(G))$. Note that if $k_1 = 1$, we are now done as all vertices in $W_t^{(1)}$ lie in the same component of G_p.

Now we iterate. Each of the vertices in $W_t^{(1)}$ lies in one of the k_1 components of $W_0^{(1)}$. We combine the largest k_2 components to form a set $W^{(2)}$ of size $> C\sigma^2\mathrm{vol}(G)$. If $k_2 = 1$, then one more growth finishes us, otherwise $\mathrm{vol}(W^{(2)}) < 2C\sigma^2\mathrm{vol}(G)$, the average size of components is at least $C_2\frac{\mathrm{vol}(G)}{\sigma^4 n}$ and hence $k_2 \leq C_2'\sigma^6 n$.

We iterate, growing $W^{(m)}$ until either $\mathrm{vol}(Q_t^{(m)}) < (1-(m+3)\epsilon)\mathrm{vol}_2(G)$ or $\mathrm{vol}(W_t^{(m)}) > \frac{2\delta\mathrm{vol}_2(Q_t^{(m)})\mathrm{vol}(G)}{5p\mathrm{vol}_3(Q_t^{(m)})}$, so that $W_t^{(m)}$ has volume $\theta(\mathrm{vol}(G))$ and then creating $W^{(m+1)}$ by combining the largest k_{m+1} components to form a $W^{(m+1)}$ with volume at least $C\sigma^2 n$. Once $k_m = 1$ for some m all vertices in $W^{(m)}$ are in the same component and one more growth round finishes the process, resulting in a giant component in G. Note that the average size of a component in $W_n^{(m)}$ has size at least $C_m\frac{\mathrm{vol}(G)}{\sigma^{2(m+1)}n}$ (that is, components must grow by a factor of at least $\frac{1}{\sigma^2}$ each iteration) and if $k_m > 1$, we must have $k_m \leq C_m'\sigma^{2(m+1)}n$. If $m = \lceil\frac{1}{2\kappa}\rceil - 1$, this would imply that $k_m = o(1)$ by our condition $\sigma = o(n^{-\kappa})$, so after at most $\lceil\frac{1}{2\kappa}\rceil - 1$ rounds, we must have $k_m = 1$ and the process will halt with a giant connected component.

In the case where $\Delta\ln n > C\sigma^2 n$, we note that $|S_0| \leq \frac{\mathrm{vol}(S_0)}{\sqrt{\epsilon}d} \leq C'\frac{\Delta\ln n}{\sqrt{\epsilon}d}$, and the average volume of components in S_t is at least $\frac{C''\mathrm{vol}(G)d}{\Delta\ln n} = \omega(\Delta\ln n)$, so we can form $W^{(1)}$ by taking just one component for n large enough, and the proof goes as above.

We note that throughout, if we try to expand we have that

$$\mathrm{vol}(Q_t^{(m)}) > (1-(m+3)\epsilon)\mathrm{vol}_2(G) > \left(1-\left(\frac{1}{2\kappa}+4\right)\epsilon\right)\mathrm{vol}_2(G) > \left(1-\frac{9c}{20}\right)\mathrm{vol}_2(G).$$

By our choice of c being sufficiently small, $(1-(m+3)\epsilon)(1-\delta)(1+c) > 1$ at all times, so throughout, noting that $\mathrm{vol}(S_i)$ and $\mathrm{vol}(W_i^{(m)})$ are at least $\Delta\ln n$, we are guaranteed our exponential growth by Lemma 3 with an error probability bounded by

$$\exp\left(-\frac{\delta(1-\delta)p\mathrm{vol}_2(T_i)\mathrm{vol}(S_i)}{10\Delta\mathrm{vol}(G)}\right) \leq \exp\left(-\frac{\delta(1-\delta)(1-\frac{9c}{20})\mathrm{vol}(S_i)}{2\Delta}\right) \leq n^{-K}.$$

We run for a constant number of phases, and run for at most a logarithmic number of steps in each growth phase as the sets grow exponentially. Thus, the probability of failure is at most

$C'' \log(n)n^{-K} = o(1)$ for some constant C'', thus completing our argument that G_p contains a giant component with high probability.

Finally, we prove the uniqueness assertion. With probability $1 - C'' \log(n)n^{-K}$ there is a giant component X. Let u be chosen at random; we estimate the probability that u is in a component of volume at least $\max(2d \log n, \omega(\sigma\sqrt{\mathrm{vol}(G)}))$. Let Y be the component of u. Theorem 5.1 of [7] asserts that if $\mathrm{vol}(Y) \geq \max(2d \log n, \omega(\sigma\sqrt{\mathrm{vol}(G)}))$:

$$e(X, Y) \geq \frac{\mathrm{vol}(X)\mathrm{vol}(Y)}{\mathrm{vol}(G)} - \sigma\sqrt{\mathrm{vol}(X)\mathrm{vol}(Y)} \geq 1.5d \log n$$

Note that the probability that Y is not connected to X given that $\mathrm{vol}(Y) = \omega(\sigma\sqrt{\mathrm{vol}(G)})$ is $(1-p)^{e(X,Y)} = o(n^{-1})$, so with probability $1 - o(1)$ no vertices are in such a component - proving the uniqueness of large components. □

References

1. N. Alon, I. Benjamini, and A. Stacey, Percolation on finite graphs and isoperimetric inequalities, *Annals of Probability*, **32**, no. 3 (2004), 1727–1745.
2. M. Ajtai, J. Komlós and E. Szemerédi, Largest random component of a k-cube, *Combinatorica* **2** (1982), 1–7.
3. B. Bollobas, Y. Kohayakawa and T. Luczak, The evolution of random subgraphs of the cube, *Random Structures and Algorithms* 3(**1**), 55–90, (1992)
4. B. Bollobas, C. Borgs, J. Chayes, O. Riordan, Percolation on dense graph sequences, preprint.
5. C. Borgs, J. Chayes, R. van der Hofstad, G. Slade, and J. Spencer. Random subgraphs of finite graphs. I. The scaling window under the triangle condition, *Random Structures and Algorithms*, 27 (**2**), 137–184 (2005).
6. C. Borgs, J. Chayes, R. van der Hofstad, G. Slade, and J. Spencer. Random subgraphs of finite graphs. III. The scaling window under the triangle condition, *Combinatorica*, 26 (**4**), 395–410 (2006).
7. F. Chung, *Spectral Graph Theory*, AMS Publications, 1997.
8. F. Chung, L. Lu, and V. Vu, The spectra of random graphs with given expected degrees, *Internet Mathematics* **1** (2004) 257–275.
9. F. Chung, and L. Lu, Connected components in random graphs with given expected degree sequences, *Annals of Combinatorics* **6**, (2002), 125–145,
10. F. Chung, and L. Lu, *Complex Graphs and Networks*, AMS Publications, 2006.
11. P. Erdős and A. Rényi, On Random Graphs I, *Publ. Math Debrecen* **6**, (1959), 290-297.
12. A. Frieze, M. Krivelevich, R. Martin, The emergence of a giant component of pseudo-random graphs, *Random Structures and Algorithms* **24**, (2004), 42–50.
13. G. Grimmett, *Percolation*, Springer, New York, 1989.
14. H. Kesten, *Percolation theory for mathematicians*, volume 2 of Progress in Probability and Statistics. Birkhäuser Boston, Mass., (1982)
15. C. Malon and I. Pak, Percolation on finite Cayley graphs, *Lecture Notes in Comput. Sci.* **2483**, Springer, Berlin, (2002), 91–104.
16. A. Nachmias, Mean-field conditions for percolation in finite graphs, preprint (2007)
17. A. Nachmias and Y. Peres, Critical percolation on random regular graphs, preprint (2007)