Percolation in General Graphs

Fan Chung [*]   Paul Horn [†]   Linyuan Lu [‡]

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 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 $p \leq \frac{1-c}{\tilde{d}}$, 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. Also 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 $g(n) = o(f(n))$.

Necessarily, our results are asymptotic and in that sense we work with infinite families of graphs $\{G^{(n)}\}$ with $|G| = n$; however our results require no continuity in the family save that some parameters remain under control. When we say, for instance, that $\Delta = o(\frac{\tilde{d}}{\sigma})$ in the statement of Theorem 1, we really mean that for 'for a family of graphs $\{G^{(n)}\}$, $\Delta_n = o(\frac{\tilde{d}_n}{\sigma_n})$. In that sense, Theorem 1 can be restated as

Theorem 2. *Suppose for a family $\{G^{(n)}\}$, has maximum degree Δ_n satisfying $\Delta_n = o(\frac{\tilde{d}}{\sigma})$. For $p \leq \frac{1-c}{\tilde{d}}$, a.a.s. every connected component in $G_p^{(n)}$ has volume at most $O(\sqrt{\mathrm{vol}_2(G^{(n)})}g(n))$ where $g(n)$ is any slowly growing function as $n \to \infty$.*

For simplicity of exposition we try to suppress dependence on the family and on n as much as possible.

In order to prove the emergence of giant component where $p \geq (1 + c)/\tilde{d}$, we need to consider some additional conditions.

We say that a (family of) graphs is f-admissible if for all sets S, $\mathrm{vol}_2(S) \geq \epsilon\mathrm{vol}_2(G)$ implies that $\mathrm{vol}(S) \geq f(\epsilon)\mathrm{vol}(G)$ with f a positive function. If such an f exists, and $\frac{\sigma\mathrm{vol}(G)}{\Delta} = \omega(\log\log(n))$ and $\frac{\mathrm{vol}_2(G)}{\Delta^2} = \omega(\log\log(n))$ we say that G is admissible. Note that to check whether G is f admissible, it suffices to check only subsets comprised of the vertices with the k highest degrees.

Theorem 3. *Suppose $p \geq \frac{1+c}{\tilde{d}}$. Suppose G has maximum degree Δ satisfying $\Delta = o(\frac{\tilde{d}}{\sigma})$, $\sigma = o(\log^{-1}(n))$, and G is admissible. Then G_p contains a unique component of volume $\Theta(\mathrm{vol}(G))$ a.a.s.*

Admissibility may seem a strong condition: it implies that if a (family of) subset(s) has $\mathrm{vol}_2(S) = \Theta(\mathrm{vol}_2(G))$, then $\mathrm{vol}(S) = \Theta(\mathrm{vol}(G))$. This suggests that finding a giant component of size $\Theta(\mathrm{vol}(G))$ may not quite be the 'right' definition of a giant component for this type of graph. $p = \frac{1}{\tilde{d}}$ is a threshold in terms of the size of the largest component for a much wider class of graphs than admissible graphs; however neither the volume or the size (number of vertices) of the largest component is the right measure in this sense.

We say that a (family of) graphs is weakly admissible if for any set S with $\mathrm{vol}(S) = O(\sigma\mathrm{vol}(G)\log(n))$ has $\mathrm{vol}_2(S) = o(\mathrm{vol}_2(G))$ and also $\frac{\sigma\mathrm{vol}(G)}{\Delta} = \omega(\log\log(n))$ and $\frac{\mathrm{vol}_2(G)}{\Delta^2} = \omega(\log\log(n))$. Note that this is a much weaker (easier to satisfy) condition than admissible, which is equivalent to saying that if $\mathrm{vol}(S) = o(\mathrm{vol}(G))$ then $\mathrm{vol}_2(S) = o(\mathrm{vol}_2(G))$. Note that weak admissibility is implied by, for instance, the condition that $\Delta = o(\frac{\tilde{d}}{\sigma\log n})$.

We prove the following theorem showing that for admissible graphs, $\frac{1}{\tilde{d}}$ is a threshold for having a giant component in the sense of second order volume.

Theorem 4. *Suppose $p > \frac{1+c}{\tilde{d}}$. Suppose G has maximum degree Δ satisfying $\Delta = o(\frac{\tilde{d}}{\sigma})$, $\sigma = o(\log^{-1}(n))$ and G is weakly admissible. Then G_p contains a unique component with second order volume $\Theta(\mathrm{vol}_2(G))$ a.a.s.*

One may ask whether weak admissibility is sufficient to guarantee a giant component in the volume sense as well. This, however is not the case. In particular, we show the following:

Theorem 5. *There exist weakly admissible graphs, satisfying the conditions of Theorem 3, such that even if $p = \frac{1+\epsilon}{\tilde{d}}$, G_p contains no giant component in the volume sense for ϵ sufficiently small.*

Finally, we show that by the time $p > \frac{1}{d}$, then G_p actually contains a giant component in the volume sense (as opposed to simply in the $\mathrm{vol}_2(G)$ sense.)

Theorem 6. *Suppose $p > \frac{1+1000c}{d}$, for some $1000c \leq \frac{1}{20}$. Suppose G has maximum degree Δ satisfying $\Delta = o(\frac{\tilde{d}}{\sigma})$ and $\sigma = o(\log^{-1}(n))$ and $\frac{\sigma \mathrm{vol}(G)}{\Delta} = \omega(\log\log(n))$, then G_p contains a component of size $\Theta(\mathrm{vol}(G))$.*

We show below that under the assumption that the maximum degree Δ of G satisfying $\Delta = o(\frac{\tilde{d}}{\sigma})$, the spectral norm of the adjacency matrix satisfies $\|A\| = \rho = (1+o(1))\tilde{d}$. Thus for graphs satisfying the conditions of Theorems 3 or 5 the threshold for having a giant component in the sense of the first or second order volume is $\frac{1}{d}$.

To examine when the conditions of Theorems 3 and 5 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. Weak admissibility removes this requirement. 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 $\gg \log^2(n)$ the spectral condition 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, and establish Theorems 3 and 5, and in Section 5, we complete the proof of Theorem 6.

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

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

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

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.

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

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 $\lambda_{n-1} \le 2$ 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*

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

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

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

Then,

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

$\square$

A chief tool we use is the following standard lemma in the vein of the expander mixing lemma (see [?])

Lemma 2. *For any two sets X and Y,*

$$\left| e(X,Y) - \frac{\mathrm{vol}(X)\mathrm{vol}(Y)}{\mathrm{vol}(G)} \right| \leq \sigma\sqrt{\mathrm{vol}(X)\mathrm{vol}(Y)}.$$

An immediate corollary is the following:

Lemma 3. *Let S be a set and fix $\epsilon > 0$. Let*

$$X = \left\{ v \in G \setminus S : \left| |\Gamma(v) \cap S| - \frac{d_v\mathrm{vol}(S)}{\mathrm{vol}(G)} \right| \geq \epsilon\frac{d_v\mathrm{vol}S}{\mathrm{vol}(G)} \right\}$$

Then

$$\mathrm{vol}(X) \leq \frac{2\sigma^2}{\epsilon}\frac{\mathrm{vol}(G)^2}{\mathrm{vol}(S)}.$$

Proof. Let

$$X^+ = \left\{ v \in G \setminus S : |\Gamma(v) \cap S| \geq (1+\epsilon)\frac{d_v\mathrm{vol}(S)}{\mathrm{vol}(G)} \right\}$$

and

$$X^- = \left\{ v \in G \setminus S : |\Gamma(v) \cap S| \leq (1-\epsilon)\frac{d_v\mathrm{vol}(S)}{\mathrm{vol}(G)} \right\}$$

Then $X = X^+ \cup X^-$.

By Lemma 2 and the definition of X^+

$$\frac{\mathrm{vol}(X^+)\mathrm{vol}(S)}{\mathrm{vol}(G)} + \sigma\sqrt{\mathrm{vol}(X^+)\mathrm{vol}(S)} \geq e(X^+, S) \geq (1+\epsilon)\frac{\mathrm{vol}(X^+)\mathrm{vol}(S)}{\mathrm{vol}(G)}.$$

Thus:

$$\mathrm{vol}(X^+) \leq \frac{\sigma^2}{\epsilon}\frac{\mathrm{vol}(G)^2}{\mathrm{vol}(S)}.$$

That $\mathrm{vol}(X^-) \leq \frac{\sigma^2}{\epsilon}\frac{\mathrm{vol}(G)^2}{\mathrm{vol}(S)}$ follows analogously, completing the proof. $\square$

In order to obtain our main result, we also need the following inequality:

Lemma 4. *For $\epsilon > 0$,*

$$(1 - e^{-x}) \geq \min\{(1-\epsilon)x, \epsilon - \epsilon^2\}.$$

Proof. By concavity of $(1 - e^{-x})$ and the fact that $1 - e^{-x}$ is increasing, it suffices to check that for $x < 1$:

$$e^{-x} < 1 - x + x^2,$$

which follows from the Taylor expansion $e^{-x} = 1 - x + \frac{x^2}{2} - \frac{x^3}{3!} + ...s$. The result clearly holds for $\epsilon > 1$. $\square$

3   The range of p with no giant component

In this section, we will prove Theorem 1.

Proof. 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 $C\sqrt{\text{vol}_2(G)}$.

Proof of Claim A: Let x be the probability that there is a component of G_p having volume greater than $C\sqrt{\text{vol}_2(G)}$. 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 $C\sqrt{\text{vol}_2(G)}$, the probability of each vertex in this component is at least $\frac{C\sqrt{\text{vol}_2(G)}}{\text{vol}(G)}$. Therefore, the probability that the random pair of vertices are in the same component is at least

$$x\left(\frac{C\sqrt{\text{vol}_2(G)}}{\text{vol}(G)}\right)^2 = \frac{C^2 x\tilde{d}}{\text{vol}(G)}. \tag{1}$$

On the other hand, for any fixed pair of vertices u and v and any path P of length k in G. The probability of u and v is connected by this path in G_p is exactly p^k. The number of k-path 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 $\frac{d_u}{\text{vol}(G)}$ and $\frac{d_v}{\text{vol}(G)}$ respectively, the probability that the random pair of vertices are in the same connected component is at most

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

We have

$$\sum_{k=0}^{n} \frac{1}{\text{vol}(G)^2} p^k \mathbf{d}^* A^k \mathbf{d} \leq \sum_{k=0}^{\infty} \frac{p^k \rho^k \text{vol}_2(G)}{\text{vol}(G)^2}$$

$$\leq \frac{\tilde{d}}{(1-p\rho)\text{vol}(G)}.$$

Combining with (1), we have

$$\frac{C^2 x\tilde{d}}{\text{vol}(G)} \leq \frac{\tilde{d}}{(1-p\rho)\text{vol}(G)}.$$

which implies

$$x \leq \frac{1}{C^2(1-p\rho)}.$$

Claim A is proved, and letting C be an arbitrarily slowly growing function completes the proof. $\quad\square$

4 Growing a Giant Component

In this section we begin by establishing the two lemmas which are the key to our analysis. Both look at the neighborhood of a set S which is fairly large (for our purposes $\text{vol}(S) > \sigma\text{vol}(G)$) into a very large set T in the percolated graph. For our purposes, T represents the unexplored area of the graph; we will stop once we have explored a positive fraction of T.

The differences between these lemmas is fairly minor; but the proof changes slightly for the purposes of proving Theorems 4 and 6.

The following lemma provides the crux of the proof of Theorem 4 (and, hence also Theorem 3).

Lemma 5. *Suppose G is a weakly admissible graph and $p > \frac{1+1000\epsilon}{\tilde{d}}$, for some $\epsilon < \frac{1}{100}$. Further suppose S and T are two sets satisfying $\text{vol}_2(T) > (1 - 5\epsilon)\text{vol}_2(\tilde{G})$ and $\text{vol}(S) > \sigma\text{vol}(G)$. Then either*

$$\text{vol}(\Gamma_p(S) \cap T) > (1 + \epsilon)\text{vol}(S)$$

or

$$\text{vol}_2(\Gamma_p(S) \cap T) > \frac{1}{2}(\epsilon^2 - \epsilon^3)\text{vol}_2(G)$$

with probability at least $1 - \max\{\exp\left(-\frac{\alpha\text{vol}(S)}{\Delta}\right), \exp\left(-\frac{(\epsilon^2-\epsilon^3)\text{vol}_2(G)}{8\Delta^2}\right)\}$. for some constant α.

Proof. Let

$$X = \{v \in T : \left||\Gamma(v) \cap S| - \frac{d_v\text{vol}(S)}{\text{vol}(G)}\right| \geq \epsilon\frac{d_v\text{vol}(S)}{\text{vol}(G)}.$$

Then by Lemma 3 $\text{vol}(X) \leq \frac{2\sigma}{\epsilon}\text{vol}(G)$. Note that as G is weakly admissible, this implies that $\text{vol}_2(X) = o(\text{vol}_2(G))$ and in particular we may assume that $\text{vol}_2(X) < \epsilon\text{vol}_2(G)$ for sufficiently large n.

Then:

$$
\begin{aligned}
\mathbb{E}[\text{vol}(\Gamma(S) \cap (T \setminus X))] &\geq \sum_{v \in T \setminus X} d_v(1 - (1 - p)^{|\Gamma(v) \cap S|}) \\
&\geq \sum_{v \in T \setminus X} d_v(1 - \exp(-p|\Gamma(v) \cap S|)) \\
&\geq \sum_{v \in T \setminus X} d_v\left(1 - \exp\left(-(1 - \epsilon)p\frac{d_v\text{vol}(S)}{\text{vol}(G)}\right)\right) \\
&\geq \sum_{v \in T \setminus X} d_v \min\left\{(1 - \epsilon)^2 p\frac{d_v\text{vol}(S)}{\text{vol}(G)}, \epsilon - \epsilon^2\right\}.
\end{aligned}
$$

8

We may split T into two parts T' and T'': let T' denote the set of vertices in $T \setminus X$ such that $(1-\epsilon)^2 p \frac{d_v \text{vol}(S)}{\text{vol}(G)} < \epsilon - \epsilon^2$, and $T'' = T \setminus (X \cup T')$. If $\text{vol}_2(T') > (1-7\epsilon)\text{vol}_2(G)$, then

$$\begin{aligned}
\mathbb{E}[\text{vol}(\Gamma(S) \cap (T \setminus X))] &\geq \sum_{v \in T'} (1-\epsilon)^2 p \frac{d_v^2 \text{vol}(S)}{\text{vol}(G)} \\
&= (1-\epsilon)^2 p \frac{\text{vol}_2(T')\text{vol}_2(T')\text{vol}(S)}{\text{vol}(G)} \\
&\geq (1-\epsilon)^2 (1-7\epsilon) \frac{(1+1000\epsilon)}{\tilde{d}} \frac{\text{vol}_2(G)\text{vol}(S)}{\text{vol}(G)} \\
&> (1+100\epsilon)\text{vol}(S).
\end{aligned}$$

Otherwise, $\text{vol}_2(T'') > \epsilon \text{vol}_2(G)$, and

$$\begin{aligned}
\mathbb{E}[\text{vol}_2(\Gamma(S) \cap (T \setminus X))] &\geq \sum_{v \in T''} d_v^2 (\epsilon - \epsilon^2) \\
&\geq (\epsilon^2 - \epsilon^3)\text{vol}_2(G).
\end{aligned}$$

We use the following Chernoff bounds, see e.g. [10]: If $X = \sum d_v X_i$ where X_i are independent indicator random variables and $|d_v| < \Delta$,

$$\mathbb{P}(X \leq \mathbb{E}[X] - a) \leq \exp\left(-\frac{a^2}{2\sum d_v^2 \mathbb{E}[X_v^2]}\right) \leq \exp\left(-\frac{a^2}{2\Delta \mathbb{E}[X]}\right).$$

Setting $a = \alpha \mathbb{E}[X]$ we have that in the first case:

$$\mathbb{P}(\text{vol}(\Gamma(S) \cap (T_t \setminus X)) < (1+\epsilon)\text{vol}(G)) \leq \exp\left(\frac{-\alpha^2 \mathbb{E}[X]}{2\Delta}\right) \leq \exp\left(-\frac{\alpha^2 (1+100\epsilon)\text{vol}(S)}{\Delta}\right).$$

In the second case:

$$\mathbb{P}\left(\text{vol}_2(\Gamma(S) \cap (T_t \setminus X)) < \frac{1}{2}(\epsilon^2 - \epsilon^3)\text{vol}_2(G)\right) < \exp\left(-\frac{(\epsilon^2 - \epsilon^3)\text{vol}_2(G)}{8\Delta^2}\right).$$

$\square$

In the case where we wish to show the emergence of a giant component in the volume sense when $p > \frac{1}{d}$, we need the following:

Lemma 6. *Suppose $p \geq \frac{1+1000\epsilon}{d}$ and G is a minimally admissible graph. Then if S and T are sets with $\text{vol}(S) > \sigma\text{vol}(G)$ and $\text{vol}(T) > (1-5\epsilon)\text{vol}(G)$ then either Then either*

$$\text{vol}(\Gamma_p(S) \cap T) > (1+\epsilon)\text{vol}(S)$$

or

$$\text{vol}(\Gamma_p(S) \cap T) > \frac{1}{2}(\epsilon^2 - \epsilon^3)\text{vol}_2(G)$$

with probability at least $1 - \max\{\exp\left(-\frac{\alpha\text{vol}(S)}{\Delta}\right), \exp\left(-\frac{(\epsilon^2-\epsilon^3)\text{vol}(G)}{8\Delta}\right)\}$, for some constant α.

9

Note that the proof of Lemma 6 is essentially analogous to the proof of Lemma 5, however we will highlight the key differences; essentially we use $\text{vol}(\cdot)$ instead of $\text{vol}_2(\cdot)$ in a few places and apply Cauchy-Schwarz.

Proof of Lemma 6. We define X, T' and T'' as before. We have that $\text{vol}(X) = o(\text{vol}(T))$, as $sigma = o(1)$ so that $\text{vol}(X) \leq \epsilon \text{vol}(T)$ for n sufficiently large. At this point either $\text{vol}(T') > (1 - 7\epsilon)\text{vol}(G)$ or $\text{vol}(T'') > \epsilon \text{vol}(G)$. (Note, in the proof of Lemma 5, we needed a statement about $\text{vol}_2(T)$ here and used the admissibility.) In the case where $\text{vol}(T') > (1 - 7\epsilon)\text{vol}(G)$, we observe that

$$\mathbb{E}[\text{vol}(\Gamma(S) \cap (T \setminus X))] \geq (1 - \epsilon)^2 p \frac{\text{vol}_2(T')\text{vol}S}{\text{vol}(G)} \geq (1 + 100\epsilon)\text{vol}(S);$$

where here we use Cauchy-Schwarz and the fact that $p \geq \frac{(1 + 1000\epsilon)}{d}$.

If $vol(T'') > \epsilon \text{vol}(G)$, then

$$\mathbb{E}[\text{vol}(\Gamma(S) \cap (T \setminus X))] \geq (\epsilon^2 - \epsilon^3)\text{vol}(G).$$

Concentration, as before, follows from the Chernoff bounds. $\square$

Before we complete the proof of Theorems 4 and 6, let us describe our strategy. Essentially, we want to run a branching-process type argument, but as the underlying graph may be rather inhomogeneous in terms of its degrees directly running such an argument can be difficult. To overcome this difficulty, we analyze a two phase process.

Phase 1: We start with an initial set S_0 with $\sigma \text{vol}(G) < S_0 < \sigma \text{vol}(G) + \Delta$, and take T_0 to be $V(G) \setminus S_0$. At each step, we take $S_{t+1} = \Gamma_p(S_t) \cap T_t$, and $T_{t+1} = T_t \setminus S_{t+1}$, with the following caveat. We never want the size of S_{t+1} to be larger than $2\sigma \text{vol}(G) + \Delta$; if S_{t+1} would be larger than $2\sigma \text{vol}(G) + \Delta$, we (arbitrarily) order the vertices of T_t and add them in order until $2\sigma \text{vol}(G) < \text{vol}(S_{t+1}) < 2\sigma \text{vol}(G) + \Delta$. Once this occurs, we perform a special round. Each of the vertices in S_0 is adjacent to some set of vertices in S_{t+1}; we will combine the largest k components to get a set S_{t+2} of with $\sigma \text{vol}(G) < \text{vol}(S_{t+2}) < \sigma \text{vol}(G) + \Delta$. (Note that to do this, we will once again use our order on the vertices to add vertices in order so that $\text{vol}(S_{t+2})$ is not too large.) We will then let $T_{t+2} = T_{t+1}$ and continue the process. Phase 1 ends when all vertices in some S_t lie in the same component.

Phase 2: At the beginning of Phase 2, we have a single set $S_t = S_0'$ which lie in a single component and a set $T_t = T_0'$ which contains all vertices which have previously been unexplored. We then consider repeatedly setting $S_{t+1'} = \Gamma_p(S_t') cap T_t'$ and $T_{t+1}' = T_t' \setminus S_{t+1}'$ until the point where either $\text{vol}(S_{t+1}) < (1 + \epsilon)\text{vol}(S_t)$ or $\text{vol}_2(T_t) < (1 - 2\epsilon)\text{vol}_2(G)$ (or, in the case of the proof of Theorem 6, when $\text{vol}_2(T_t') < (1 - 2\epsilon)\text{vol}(G)$. At this point we stop and output the component containing S_t'.

Note that we need be slightly careful during the execution of Phase 1; the key here is that we do not wish to investigate too much of the graph before we know that we are actually in the

giant component. We need to ensure that T is large enough at the end of Phase 1 that we can successfully use Lemma 5 or 6 in Phase 1.

To complete the proof of the main theorems, we only need to establish the following lemmas:

Lemma 7. *Suppose G is a weakly admissible graph with $\sigma = o(\log^{-1}(n))$ and $p \geq \frac{1+1000\epsilon}{\tilde{d}}$, then a.a.s. Phase 1 terminates in $O(\log(n))$ steps with all vertices in S_t in a single component, and $\mathrm{vol}_2(T_t) > (1-\epsilon)\mathrm{vol}_2(G)$.*

Lemma 8. *Suppose G is a weakly admissible graph with $\sigma = o(\log^{-1}(n))$ and $p \geq \frac{1+1000\epsilon}{\tilde{d}}$, then a.a.s. Phase 2 terminates in $O(\log(n))$ steps with a single component where $\mathrm{vol}_2(S'_t) = \Theta(\mathrm{vol}_2(G))$.*

Theorem 4 follows directly from the proofs of Lemmas 7 and 8. Theorem 3 follows immediately from Theorem 4 and the stronger condition of admissibility.

Theorem 6 follows from the following slight variants of Lemmas refphase1 and 8, whose proofs are essentially identical using Lemma 6 instead of Lemma 5.

Lemma 9. *Suppose G is a weakly admissible graph with $\sigma = o(\log^{-1}(n))$ and $p \geq \frac{1+1000\epsilon}{d}$, then a.a.s. Phase 1 terminates in $O(\log(n))$ steps with all vertices in S_t in a single component, and $\mathrm{vol}(T_t) > (1-\epsilon)\mathrm{vol}_2(G)$.*

Lemma 10. *Suppose G is a weakly admissible graph with $\sigma = o(\log^{-1}(n))$ and $p \geq \frac{1+1000\epsilon}{d}$, then a.a.s. Phase 2 terminates in $O(\log(n))$ steps with a single component where $\mathrm{vol}(S'_t) = \Theta(\mathrm{vol}(G))$.*

As the proofs of Lemmas 9 and 10 are essentially identical to those of Lemmas 7 and 8 we will suppress their proofs.

Proof of Lemma 7. Let S_0 be an arbitrary starting set satisfying the conditions of Phase 1. By Lemma 5, $\mathrm{vol}(\Gamma(S_t) \cap T_t) > (1+\epsilon)\mathrm{vol}(S_t)$ with failure probability at most. Note that the set T' will be empty (at least for n sufficiently large) as in Phase 1, $\mathrm{vol}(S_t) < 2\sigma\mathrm{vol}(G) + \Delta < 3\sigma\mathrm{vol}(G)$.

$$(1-\epsilon)^2 p \frac{d_v \mathrm{vol}(S)}{\mathrm{vol}(G)} \leq (1-\epsilon)^2 \frac{\Delta\sigma}{\tilde{d}} = o(1),$$

by the condition $\Delta = o(\frac{\tilde{d}}{\sigma})$; and hence this is less than $\epsilon^2 - \epsilon$ for large enough n.

Assuming $\mathrm{vol}(\Gamma(S_t) \cap T_t) > (1+\epsilon)\mathrm{vol}(S_t)$ at each step, the number of steps between having an S_t of volume $\sigma\mathrm{vol}(G)(+\Delta)$ and a set $S_{t'}$ of size $2\sigma\mathrm{vol}(G)(+\Delta)$ is bounded by a constant. Furthermore, collecting the largest components to find a new set of size $\sigma\mathrm{vol}(G)$ by collecting the largest components shrinks the number of components by a constant factor. Thus after a logarithmic number of steps, all vertices in S_t will be in the same component.

Note that $\mathrm{vol}(S_t)$ never exceeds $3\sigma\mathrm{vol}(G)$, so

$$\mathrm{vol}(\bigcup_t S_t) = O(\sigma\mathrm{vol}(G)\log(n)).$$

11

By admissibility, $\text{vol}_2(\bigcup_t S_t) = o(\text{vol}_2(G))$ so $\text{vol}_2(T_t) > (1-\epsilon)\text{vol}_2(G)$ for n sufficiently large.

In total, the probability of failure is bounded by $O(\log(n)) \times o(\log^{-1}(n)) = o(1)$, completing the proof of the theorem. Note, here we use the condition that $\frac{\sigma \text{vol}(G)}{\Delta} = \omega(\log\log(n))$ and $\frac{\text{vol}_2(G)}{\Delta^2} = \omega(\log\log(n))$ to observe the failure probability in Lemma 5 is $o(\log^{-1}(n))$. $\qquad\square$

Proof of Lemma 8. By Lemma 7, Phase 1 succeeds a.a.s., and hence at the beginning of Phase 2, S_0' is a set of vertices of volume at least $\sigma\text{vol}(G)$ with all vertices lying in the same component, and furthermore $\text{vol}_2(T_0') > (1-\epsilon)\text{vol}_2(G)$.

As we stop when $\text{vol}_2(T_t) < (1-2\epsilon)\text{vol}_2(G)$ and we only continue so long as $\text{vol}(S_{t+1}) > (1+\epsilon)\text{vol}(S_{t+1})$ the hypothesis of Lemma 5 are always satisfied. So long as the a.a.s. conclusions of Lemma 5 hold, we will either continue until some step s where $\text{vol}_2(T_s) < (1-2\epsilon)\text{vol}_2(G)$, or until the second end condition of Lemma 5 holds; that is $\text{vol}_2(S_{s+1}) > (\epsilon^3 - \epsilon^2)\text{vol}_2(G)$. In the second case, we are done; $\text{vol}_2(S_{s+1}) = \Theta(\text{vol}_2(G))$ and is clearly part of a giant component in G. If we stop because $\text{vol}_2(T_s) < (1-2\epsilon)\text{vol}_2(G)$, then note that

$$\text{vol}_2(\bigcup_t S_t) \geq \text{vol}_2(T_0) - \text{vol}_2(T_s) > \epsilon\text{vol}_2(G),$$

and thus there exists a giant component as desired.

Since, while we continue, $\text{vol}(S_t)$ is growing exponentially this can continue for at most $O(\log(n))$ steps; and as before the failure probability after so many steps is $o(1)$.

$\qquad\square$

5 Percolated graphs without a giant component.

In this section we wish to prove Theorem 5; that is we wish to give an example of a weakly admissible graph which even when $p = \frac{1+\epsilon}{\tilde{d}}$ the percolated random subgraph has no giant component in the volume sense; even though it has one in the sense of second order volume.

To construct our graph, we use the $G(\mathbf{w})$ random graph model; the monograph of the first and third authors [10] contains a thorough analysis of this model.

For a vector of weights $\mathbf{w} = (w_1, \ldots, w_n)$, the $G(\mathbf{w})$ model independently places an edge between vertex v_i and v_j with probability $\frac{w_i w_j}{\sum w_i}$. We denote $\text{vol}(G) = \sum w_i$ the expected volume of a graph in $G(\mathbf{w})$ and $\text{vol}_2(G) = \sum w_i^2$ the expected second order degree. So long as the w_i are sufficiently large (in our example they are polynomial in the degrees), it is easy to see the actual volume and second order volume are tightly concentrated on their expectations.

Claim: Consider a graph $G \in G(\mathbf{w})$ where $\mathbf{w}$ is a vector with $n - n^{.25}$ w_i's with $w_i = n^{.2}$ and $n^{.25}$ w_i's with $w_i = n^{.9}$. Then G is a.a.s. weakly admissible, but G_p does not contain a giant component in the volume sense a.a.s. if $p = \frac{1+\epsilon}{\tilde{d}}$.

First observe

$$\begin{aligned}
\mathrm{vol}(G) &= (n - n^{.25})n^{.2} + n^{.9}n^{.25} = (1 + o(1))n^{1.2} \\
\mathrm{vol}_2(G) &= (n - n^{.25})n^{.4} + n^{1.8}n^{.25} = (1 + o(1))n^{2.05} \\
\tilde{d} &= \frac{\mathrm{vol}(G)}{\mathrm{vol}_2(G)} = (1 + o(1))n^{.85}.
\end{aligned}$$

In the graph G_p restricted to the vertices of weight $n^{.2}$ is an Erdős-Rényi random graph, where two vertices are adjacent with probability

$$(1 + o(1))\frac{1 + \epsilon}{\tilde{d}} \times \frac{n^{.2} \cdot n^{.2}}{n^{1.2}} = o(1/n),$$

and hence a.a.s. the largest component in that subgraph has size $O(\log(n))$. It is easy to see that no vertex in the subgraph of vertices with weight $n^{.9}$ has more than $2\frac{n^{.9}}{\tilde{d}} = (2 + o(1))n^{.05}$ neighbors in G_p. Thus no vertex in that subgraph can be adjacent to more than $(2 + o(1))n^{.05}$ of the components of size $O(\log(n))$ in the rest of the graph so a bound on the volume of the largest component is

$$n^{.9}n^{.25} + (2 + o(1))n^{.05}n^{.2}\log(n) = o(n^{1.2}) = o(\mathrm{vol}(G)).$$

Thus G a.a.s. contains no giant component in the volume sense.

It is known that for a graph in $G(\mathbf{w})$ with $w_{\min}$ sufficiently large that $\sigma = O(\frac{1}{\sqrt{w}})$ where w is the expected average degree. Thus for G, we have that $\sigma = O(n^{-.1})$. Note that

$$\Delta = (1 + o(1))n^{.9} = o(n^{.95}) = o(\tilde{d}/\sigma).$$

Furthermore, note that the volume of the set consisting of all vertices of weight $n^{.9}$ has volume $n^{1.15} = \omega(n^{.1}\log(n)) = \omega(\sigma\mathrm{vol}(G)\log(n))$. In particular, if S has $\mathrm{vol}(S) = O(\sigma\mathrm{vol}(G)\log(n))$, then the S contains at most $O(n^{.2}\log(n))$ vertices of weight $n^{.9}$, and hence

$$\mathrm{vol}_2(S) = O(n^{1.9}n^{.2}\log(n)) = o(\mathrm{vol}_2(G)).$$

Since the degrees of all vertices in G are concentrated on their expectation this, in particular, implies that G is weakly admissible a.a.s., and thus that weak admissibility is not sufficient to imply a giant component in the volume sense when $p = \frac{1+\epsilon}{\tilde{d}}$ even though it does imply the existence of a giant component in the second order volume sense.

This completes the proof of the claim, and hence of Theorem 5.

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)