Percolation in General Graphs

Fan Chung, Paul Horn, and 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. An extended abstract of this paper appeared in the WAW 2009 proceedings [Chung et al. 09]. The main theorems are strengthened with much weaker assumptions.

1. Introduction

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

We are interested in random subgraphs G_p of a graph G obtained as follows: for each edge in G 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 in which the host graph is K_n. Other examples are percolation problems that have

long been studied in theoretical physics [Grimmett 89, Kesten 82], mainly with
the host graph being the lattice graph $\mathbb{Z}^k$. In this paper, we consider a general
host graph, such as 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 a 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 the case of K_n, Erdős and Rényi answered this question in their seminal
paper [Erdős and Rényi 59]: 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 there is a giant component of size proportional to n. For general
host graphs, the answer has been more elusive. Results have been obtained either
for very dense graphs or bounded-degree graphs. It was shown in [Bollobas et
al. 10] 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. In [Frieze et al. 04], the authors consider the case in which the host
graph is d-regular, and they show that the critical probability is close to $1/d$,
strengthening earlier results on hypercubes [Ajtai et al. 82, Bollobas et al. 92] and
Cayley graphs [Malon and Pak 02]. For expander graphs with degrees bounded
by d, it is proved in [Alon et al. 04] 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, that 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 [Borgs et al. 05, Borgs et al. 06], the au-
thors find the order of this critical window for transitive graphs and cubes. In
[Nachmias 07], a similar situation to that of [Frieze et al. 04] is considered, and
random walk techniques are used to study percolation within the critical win-
dow for quasirandom transitive graphs. Percolation within the critical window
on random regular graphs is also studied in [Nachmias and Peres 07]. Our re-
sults differ from these in that we study percolation on graphs with a much more
general degree sequence. A greater precision of these results, however, would be
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 that 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 the 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.

Theorem 1.1. *Suppose G has the maximum degree Δ satisfying $\Delta = o(\tilde{d}/\sigma)$. For $p \leq (1 - c)/\tilde{d}$ for some $c > 0$, asymptotically almost surely (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 $\Delta = o(\tilde{d}/\sigma)$ in the statement of Theorem 1.1, we really mean that for a family of graphs $\{G^{(n)}\}$, $\Delta_n = o(\tilde{d}_n/\sigma_n)$. In that sense, the theorem can be restated as follows.

Theorem 1.2. *Suppose a family $\{G^{(n)}\}$ has maximum degree Δ_n satisfying $\Delta_n = o(\tilde{d}/\sigma)$. For $p \leq (1 - c)/\tilde{d}$, 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 a giant component where $p \geq (1 + c)/\tilde{d}$, we need to consider some additional conditions. We say that a (family of) graph(s) 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 $\sigma\, \mathrm{vol}(G)/\Delta = \omega(\log\log(n))$ and $\mathrm{vol}_2(G)/\Delta^2 = \omega(\log\log(n))$, we say that G is admissible. Note that to check whether G is f-admissible, it suffices to check only subsets comprising the vertices with the k highest degrees.

Theorem 1.3. *Suppose $p \geq (1 + c)/\tilde{d}$. Suppose G has maximum degree Δ satisfying $\Delta = o(\tilde{d}/\sigma)$, $\sigma = o(\log^{-1}(n))$, 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. The value $p = 1/\tilde{d}$ 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 nor the size (number of vertices) of the largest component is the right measure in this sense.

We say that a (family of) graph(s) is weakly admissible if 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 $\sigma \, \mathrm{vol}(G)/\Delta = \omega(\log\log(n))$ and $\mathrm{vol}_2(G)/\Delta^2 = \omega(\log\log(n))$. Note that this is a much weaker (easier to satisfy) condition than admissibility, 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 $\Delta = o(\tilde{d}/\sigma \log n)$.

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

Theorem 1.4. *Suppose $p > (1 + c)/\tilde{d}$. Suppose G has maximum degree Δ satisfying $\Delta = o(\tilde{d}/\sigma)$, $\sigma = o(\log^{-1}(n))$, 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 1.5. *There exist weakly admissible graphs satisfying the conditions of Theorem 1.3 such that even if $p = (1 + \epsilon)/\tilde{d}$, G_p contains no giant component in the volume sense for ϵ sufficiently small.*

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

Theorem 1.6. *Suppose $p > (1 + c)/d$. Suppose G has maximum degree Δ satisfying $\Delta = o(\tilde{d}/\sigma)$ and that $\sigma = o(\log^{-1}(n))$ and $\sigma \, \mathrm{vol}(G)/\Delta = \omega(\log\log(n))$. 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 satisfies $\Delta = o(\tilde{d}/\sigma)$, the spectral norm of the adjacency matrix satisfies

$\|A\| = \rho = (1 + o(1))\tilde{d}$. Thus for graphs satisfying the conditions of Theorem 1.3 or 1.5, the threshold for having a giant component in the sense of first- or second-order volume is $1/\tilde{d}$.

To examine when the conditions of Theorems 1.3 and 1.5 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 greater than d, there cannot be too many. Weak admissibility removes this requirement. In [Chung et al. 04], it is shown that for random graphs with a given expected degree sequence $\sigma = O(1/\sqrt{d})$, 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 et al. 04] with general degree sequences and is also a strengthening of the original results of Erdős and Réyni to general host graphs.

The paper is organized as follows: In Section 2 we introduce notation and some basic facts. In Section 3, we examine several spectral lemmas that allow us to control the expansion and establish Theorems 1.3 and 1.5, and in Section 5, we complete the proof of Theorem 1.6.

2.   Preliminaries

Suppose G is a connected graph on the 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 $\delta = \min_v d_v$ the minimum degree.

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

$$\mathcal{L} = I - D^{-1/2}AD^{-1/2}.$$

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

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

Many properties of the λ_i can be found in [Chung 97]. 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 nonbipartite. For random graphs with a given expected degree sequence [Chung et al. 04], $\sigma = O(1/\sqrt{d})$, 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, for example, [Chung 97].

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

Lemma 2.1. *The largest eigenvalue, ρ, of the adjacency matrix of G satisfies*

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

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

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

Then

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

An important tool that we use is the following standard lemma in the vein of the expander mixing lemma (see [Chung 97]).

Lemma 2.2. *For any two sets X and Y, the number of edges between X and Y, denoted by $E(X, Y)$, satisfies*

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

An immediate corollary is the following.

Lemma 2.3. *Let S be a set and fix $\epsilon > 0$. Let $\Gamma(v)$ denote the neighborhood of v. Define*

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

Then

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

Proof. Let

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

and

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

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

By Lemma 2.2 and the definition of X^+,

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

Thus

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

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

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

Lemma 2.4. *For $\epsilon > 0$ and $x > 0$,*

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

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 $e^{-x} = 1 - x + \frac{x^2}{2} - \frac{x^3}{3!} + \cdots$. The result clearly holds for $\epsilon > 1$. $\qquad\square$

3.   The Range of p with No Giant Component

In this section, we will prove Theorem 1.1.

Proof of Theorem 1.1. It suffices to prove the following claim.

Claim 3.1. *If $p\rho < 1$, where ρ is the largest eigenvalue of the adjacency matrix, then with probability at least $1 - \frac{1}{C^2(1-p\rho)}$, all components have volume at most $C\sqrt{\operatorname{vol}_2(G)}$.*

Proof of Claim 3.1. 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 $C\sqrt{\mathrm{vol}_2(G)}/\mathrm{vol}(G)$. Therefore, the probability that the random pair of vertices are in the same component is at least

$$x\left(\frac{C\sqrt{\mathrm{vol}_2(G)}}{\mathrm{vol}(G)}\right)^2 = \frac{C^2 x\tilde{d}}{\mathrm{vol}(G)}. \tag{3.1}$$

On the other hand, for any fixed pair of vertices u and v and any path P of length k in G, the probability that u and v are 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 $d_u/\mathrm{vol}(G)$ and $d_v/\mathrm{vol}(G)$ 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 (3.1), 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 3.1 is proved. $\square$

Then letting C be an arbitrarily slowly growing function completes the proof of the theorem. $\square$

4.   Growing a Giant Component

In this section we begin by establishing the two lemmas that are the key to our analysis. Both lemmas concern the neighborhood of a set S that is fairly

large (for our purposes $\operatorname{vol}(S) > \sigma \operatorname{vol}(G)$) intersecting 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 1.4 and 1.6.

The following lemma provides the crux of the proof of Theorem 1.4 (and hence also of Theorem 1.3).

Lemma 4.1. *Suppose G is a weakly admissible graph and $p > (1 + 1000\epsilon)/\tilde{d}$, for some $\epsilon < 1/100$. Further, suppose S and T are two sets satisfying $\operatorname{vol}_2(T) > (1 - 5\epsilon)\operatorname{vol}_2(G)$ and $\operatorname{vol}(S) > \sigma \operatorname{vol}(G)$. Then the neighborhood $\Gamma_p(S)$ of S in G_p defined by $\Gamma_p(S) = \{u \in V : \{u, v\} \in E(G_p)\}$ for some $v\}$ satisfies either*

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

with probability at least

$$1 - \max\left\{\exp\left(-\frac{\alpha \operatorname{vol}(S)}{\Delta}\right), \exp\left(-\frac{(\epsilon^2 - \epsilon^3)\operatorname{vol}_2(G)}{8\Delta^2}\right)\right\}$$

for some constant α.

Proof. Let

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

Then by Lemma 2.3, $\operatorname{vol}(X) \leq \frac{2\sigma}{\epsilon}\operatorname{vol}(G)$. Note that since G is weakly admissible, this implies that $\operatorname{vol}_2(X) = o(\operatorname{vol}_2(G))$, and in particular, we may assume that $\operatorname{vol}_2(X) < \epsilon \operatorname{vol}_2(G)$ for sufficiently large n.

Then

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

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

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

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

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

We use the following Chernoff bounds; see, e.g., [Chung and Lu 06]: if $X = \sum d_v X_i$, where X_i are independent indicator random variables and $|d_v| < \Delta$, then

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

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

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

In the second case,

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

In the case that we wish to show the emergence of a giant component in the volume sense when $p > 1/d$, we need the following lemma.

Lemma 4.2. *Suppose $p \geq (1 + 1000\epsilon)/d$ for some $\epsilon < 1/100$, and G is a weakly admissible graph. Then if S and T are sets with $\operatorname{vol}(S) > \sigma \operatorname{vol}(G)$ and $\operatorname{vol}(T) > (1 - 5\epsilon) \operatorname{vol}(G)$, then either*

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

or

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

with probability at least

$$1 - \max\left\{\exp\left(-\frac{\alpha \operatorname{vol}(S)}{\Delta}\right), \exp\left(-\frac{(\epsilon^2 - \epsilon^3)\operatorname{vol}(G)}{8\Delta}\right)\right\},$$

for some constant α.

Note that the proof of Lemma 4.2 is essentially analogous to the proof of Lemma 4.1. However, we will highlight the key differences; essentially, we use $\operatorname{vol}(\cdot)$ instead of $\operatorname{vol}_2(\cdot)$ in a few places and apply Cauchy–Schwarz.

Proof of Lemma 4.2. We define X, T', and T'' as before. We have that $\operatorname{vol}(X) = o(\operatorname{vol}(T))$, since $\sigma = o(1)$, so that $\operatorname{vol}(X) \le \epsilon \operatorname{vol}(T)$ for n sufficiently large. At this point either $\operatorname{vol}(T') > (1 - 7\epsilon)\operatorname{vol}(G)$ or $\operatorname{vol}(T'') > \epsilon \operatorname{vol}(G)$. (Note that in the proof of Lemma 4.1, we needed a statement about $\operatorname{vol}_2(T)$ here and used admissibility.) In the case that $\operatorname{vol}(T') > (1 - 7\epsilon)\operatorname{vol}(G)$, we observe that

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

where here we use Cauchy–Schwarz and the fact that $p \ge ((1 + 1000\epsilon))/d$.

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

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

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

Before we complete the proof of Theorems 1.4 and 1.6, let us describe our strategy. Essentially, we want to run a branching-process-type argument, but since 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 $\sigma \operatorname{vol}(G) < S_0 < \sigma \operatorname{vol}(G) + \Delta$, 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 $2\sigma \operatorname{vol}(G) + \Delta$. If S_{t+1} is larger than $2\sigma \operatorname{vol}(G) + \Delta$, we (arbitrarily) order the vertices of T_t and add them in order until $2\sigma \operatorname{vol}(G) < \operatorname{vol}(S_{t+1}) < 2\sigma \operatorname{vol}(G) + \Delta$. 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} with $\sigma \operatorname{vol}(G) < \operatorname{vol}(S_{t+2}) < \sigma \operatorname{vol}(G) + \Delta$. (Note that to do this, we will once again use our order on the vertices to add vertices in order so that $\operatorname{vol}(S_{t+2})$ 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 $S_t = S'_0$ that lies in a single component and a set $T_t = T'_0$ that contains all vertices that 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 $\operatorname{vol}(S_{t+1}) < (1+\epsilon) \operatorname{vol}(S_t)$ or $\operatorname{vol}_2(T_t) < (1-2\epsilon) \operatorname{vol}_2(G)$ (or in the case of the proof of Theorem 1.6, when $\operatorname{vol}_2(T'_t) < (1-2\epsilon) \operatorname{vol}(G)$). At this point we stop and output the component containing S'_t.

Note that we need to 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 4.1 or 4.2 in Phase 1.

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

Lemma 4.3. *Suppose G is a weakly admissible graph with $\sigma = o(\log^{-1}(n))$ and $p \geq (1 + 1000\epsilon)/\tilde{d}$ for some $\epsilon < 1/100$. Then a.a.s., Phase 1 terminates in $O(\log(n))$ steps with all vertices in S_t in a single component, and $\operatorname{vol}_2(T_t) > (1-\epsilon) \operatorname{vol}_2(G)$.*

Lemma 4.4. *Suppose G is a weakly admissible graph with $\sigma = o(\log^{-1}(n))$ and $p \geq (1 + 1000\epsilon)/\tilde{d}$ for some $\epsilon < 1/100$. Then a.a.s., Phase 2 terminates in $O(\log(n))$ steps with a single component, where $\operatorname{vol}_2(S'_t) = \Theta(\operatorname{vol}_2(G))$.*

Theorem 1.4 follows directly from the proofs of Lemmas 4.3 and 4.4. Theorem 1.3 follows immediately from Theorem 1.4 and the stronger condition of admissibility.

Theorem 1.6 follows from the following slight variants of Lemmas 4.3 and 4.4, whose proofs are essentially identical using Lemma 4.2 instead of Lemma 4.1.

We remark that the requirement that we have $\epsilon < 1/100$ in the statements of these lemmas does not affect the existence of the giant component. If we were trying to carefully determine the size of the giant component, however, this would limit us. This assumption is an artifact of the proof of Lemma 4.1, where we need $(1 - \epsilon^2)(1 - 7\epsilon)(1 + 1000\epsilon) > (1 + 100\epsilon)$, for instance.

Lemma 4.5. *Suppose G is a weakly admissible graph with $\sigma = o(\log^{-1}(n))$ and $p \geq (1 + 1000\epsilon)/d$ for some $\epsilon < 1/100$. 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 4.6. *Suppose G is a weakly admissible graph with $\sigma = o(\log^{-1}(n))$ and $p \geq (1 + 1000\epsilon)/d$ for some $\epsilon < 1/100$. Then a.a.s., Phase 2 terminates in $O(\log(n))$ steps with a single component, where $\mathrm{vol}(S'_t) = \Theta(\mathrm{vol}(G))$.*

Since the proofs of Lemmas 4.5 and 4.6 are essentially identical to those of Lemmas 4.3 and 4.4, we will suppress them.

Proof of Lemma 4.3. Let S_0 be an arbitrary starting set satisfying the conditions of Phase 1. By Lemma 4.1, $\mathrm{vol}(\Gamma(S_t) \cap T_t) > (1 + \epsilon)\,\mathrm{vol}(S_t)$ with failure probability bounded above as in the statement of the lemma. Note that the set T' in Lemma 4.1 will be empty (at least for n sufficiently large), since in Phase 1, $\mathrm{vol}(S_t) < 2\sigma\,\mathrm{vol}(G) + \Delta < 3\sigma\,\mathrm{vol}(G)$ and so

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

by the condition $\Delta = o(\tilde{d}/\sigma)$. 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 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

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

By admissibility, $\mathrm{vol}_2(\bigcup_t S_t) = o(\mathrm{vol}_2(G))$, so $\mathrm{vol}_2(T_t) > (1 - \epsilon)\,\mathrm{vol}_2(G)$ 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 that here we use the condition that

$$\frac{\sigma\,\mathrm{vol}(G)}{\Delta} = \omega(\log\log(n)) \quad \text{and} \quad \frac{\mathrm{vol}_2(G)}{\Delta^2} = \omega(\log\log(n))$$

to observe that the failure probability in Lemma 4.1 is $o(\log^{-1}(n))$. $\qquad\square$

Proof of Lemma 4.4. By Lemma 4.3, 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 $\sigma \operatorname{vol}(G)$ with all vertices lying in the same component, and furthermore,

$$\operatorname{vol}_2(T_0') > (1 - \epsilon) \operatorname{vol}_2(G).$$

Since we stop when $\operatorname{vol}_2(T_t) < (1 - 2\epsilon) \operatorname{vol}_2(G)$ and we continue only as long as $\operatorname{vol}(S_{t+1}) > (1 + \epsilon) \operatorname{vol}(S_{t+1})$, the hypotheses of Lemma 4.1 are always satisfied. As long as the a.a.s. conclusions of Lemma 4.1 hold, we will continue either until some step s where $\operatorname{vol}_2(T_s) < (1 - 2\epsilon) \operatorname{vol}_2(G)$, or until the second condition of Lemma 4.1 holds, that is, $\operatorname{vol}_2(S_{s+1}) > (\epsilon^3 - \epsilon^2) \operatorname{vol}_2(G)$. In the second case, we are done since $\operatorname{vol}_2(S_{s+1}) = \Theta(\operatorname{vol}_2(G))$ and S_s is clearly part of a giant component in G. If we stop because $\operatorname{vol}_2(T_s) < (1 - 2\epsilon) \operatorname{vol}_2(G)$, then note that

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

and thus there exists a giant component, as desired.

Since while we continue, $\operatorname{vol}(S_t)$ 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 1.5; that is, we wish to give an example of a weakly admissible graph such that even when $p = (1 + \epsilon)/\tilde{d}$, 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 [Chung and Lu 06] 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 vertices v_i and v_j with probability $w_i w_j / \sum w_i$. We denote by $\operatorname{vol}(G) = \sum w_i$ the expected volume of a graph in $G(\mathbf{w})$ and by $\operatorname{vol}_2(G) = \sum w_i^2$ the expected second-order degree. As long as the w_i are sufficiently large (in our example they are polynomial in the degrees), it is easy to see that the actual volume and second-order volume are tightly concentrated on their expectations.

Claim 5.1. *Consider a graph $G \in G(\mathbf{w})$, where $\mathbf{w}$ is a vector with $n - n^{0.25}$ w_i's with $w_i = n^{0.2}$ and $n^{0.25}$ w_i's with $w_i = n^{0.9}$. 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 $p = (1 + \epsilon)/\tilde{d}$.

First observe that

$$\text{vol}(G) = (n - n^{0.25})n^{0.2} + n^{0.9}n^{0.25} = (1 + o(1))n^{1.2},$$
$$\text{vol}_2(G) = (n - n^{0.25})n^{0.4} + n^{1.8}n^{0.25} = (1 + o(1))n^{2.05},$$
$$\tilde{d} = \frac{\text{vol}(G)}{\text{vol}_2(G)} = (1 + o(1))n^{0.85}.$$

The graph G_p restricted to the vertices of weight $n^{0.2}$ 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 $n^{0.9}$ has more than

$$2\frac{n^{0.9}}{\tilde{d}} = (2 + o(1))n^{0.05}$$

neighbors in G_p. Thus no vertex in that subgraph can be adjacent to more than $(2 + o(1))n^{0.05}$ 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

$$n^{0.9}n^{0.25} + (2 + o(1))n^{.05}n^{0.2}\log(n) = o(n^{1.2}) = o(\text{vol}(G)).$$

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, $\sigma = O(1/\sqrt{w})$, where w is the expected average degree. Thus for G, we have that $\sigma = O(n^{-0.1})$. Note that

$$\Delta = (1 + o(1))n^{0.9} = o(n^{9.95}) = o(\tilde{d}/\sigma).$$

Furthermore, note that the volume of the set consisting of all vertices of weight $n^{0.9}$ has volume $n^{1.15} = \omega(n^{0.1}\log(n)) = \omega(\sigma\,\text{vol}(G)\log(n))$. In particular, if S has $\text{vol}(S) = O(\sigma\,\text{vol}(G)\log(n))$, then S contains at most $O(n^{0.2}\log(n))$ vertices of weight $n^{0.9}$, and hence

$$\text{vol}_2(S) = O(n^{1.9}n^{0.2}\log(n)) = o(\text{vol}_2(G)).$$

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 $p = (1 + \epsilon)/\tilde{d}$, 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 1.5.

References

[Ajtai et al. 82] M. Ajtai, J. Komlós, and E. Szemerédi. "Largest Random Component of a k-Cube." *Combinatorica* 2 (1982), 1–7.

[Alon et al. 04] N. Alon, I. Benjamini, and A. Stacey. "Percolation on Finite Graphs and Isoperimetric Inequalities." *Annals of Probability* 32:3 (2004), 1727–1745.

[Bollobas et al. 92] B. Bollobas, Y. Kohayakawa, and T. Luczak. "The Evolution of Random Subgraphs of the Cube." *Random Structures and Algorithms* 3:1 (1992), 55–90.

[Bollobas et al. 10] B. Bollobas, C. Borgs, J. Chayes, and O. Riordan. "Percolation on Dense Graph Sequences." Preprint, 2010.

[Borgs et al. 05] 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 (2005), 137–184.

[Borgs et al. 06] 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 (2006), 395–410.

[Chung 97] F. Chung. *Spectral Graph Theory*, CBMS Regional Conference Series in Mathematics 92. Providence: American Mathematical Society, 1997.

[Chung and Lu 02] F. Chung and L. Lu. "Connected Components in Random Graphs with Given Expected Degree Sequences." *Annals of Combinatorics* 6 (2002), 125–145.

[Chung and Lu 06] F. Chung and L. Lu. *Complex Graphs and Networks*, CBMS Regional Conference Series in Mathematics 107. Providence: American Mathematical Society, 2006.

[Chung et al. 04] F. Chung, L. Lu, and V. Vu. "The Spectra of Random Graphs with Given Expected Degrees." *Internet Mathematics* 1:3 (2004), 257–275.

[Chung et al. 09] Fan Chung, Paul Horn, and Linyuan Lu. "The Giant Component in a Random Subgraph of a Given Graph." In *Algorithms and Models for the Web-Graph: 6th International Workshop, WAW 2009, Barcelona, Spain, February 12–13, 2009, Proceedings*, Lecture Notes in Computer Science 5427, pp. 38–49. Berlin: Springer, 2009.

[Erdős and Rényi 59] P. Erdős and A. Rényi. "On Random Graphs I." *Publ. Math Debrecen* 6 (1959), 290–297.

[Frieze et al. 04] A. Frieze, M. Krivelevich, and R. Martin. "The Emergence of a Giant Component of Pseudo-random Graphs." *Random Structures and Algorithms* 24 (2004), 42–50.

[Grimmett 89] G. Grimmett. *Percolation*. New York: Springer, 1989.

[Kesten 82] H. Kesten. *Percolation Theory for Mathematicians*, Progress Probability and Statistics 2. Boston: Birkhäuser, 1982.

[Malon and Pak 02] C. Malon and I. Pak. "Percolation on Finite Cayley Graphs." In Randomization and Approximation Techniques in Computer Science: 6th International Workshop, RANDOM 2002 Cambridge, MA, USA, September 13-15, 2002, Proceedings, Lecture Notes in Computer Science 2483, pp. 91–104. Berlin: Springer, 2002.

[Nachmias 07] A. Nachmias. "Mean-Field Conditions for Percolation in Finite Graphs." Preprint, 2007.

[Nachmias and Peres 07] A. Nachmias and Y. Peres. "Critical Percolation on Random Regular Graphs." Preprint, 2007.

Fan Chung, University of California, San Diego, Department of Mathematics, 9500 Gilman Drive, La Jolla, CA 92093-0112 (fan@math.ucsd.edu)

Paul Horn, Emory University, Department of Mathematics and Computer Science, 400 Dowman Dr., W401, Atlanta, GA 30322 (phorn@mathcs.emory.edu)

Linyuan Lu, Department of Mathematics, University of South Carolina, Columbia, SC 29208 (lu@math.sc.edu)