

On the expansion of the giant component in percolated (n, d, λ) graphs

Eran Ofek *

August 3, 2006

Abstract

Let $d \geq d_0$ be a sufficiently large constant. A $(n, d, c\sqrt{d})$ graph G is a d -regular graph over n vertices whose second largest (in absolute value) eigenvalue is at most $c\sqrt{d}$. For any $0 < p < 1$, G_p is the graph induced by retaining each edge of G with probability p . It is known that for $p > \frac{1}{d}$ the graph G_p almost surely contains a unique giant component (a connected component with linear number vertices). We show that for $p \geq \frac{5c}{\sqrt{d}}$ the giant component of G_p almost surely has an edge expansion of at least $\frac{1}{\log n}$.

1 Introduction

This paper deals with the affect of percolation on the edge expansion property of *algebraic expander* graphs. These are d -regular graphs in which the second largest eigenvalue (in absolute value) λ of their adjacency matrix is smaller than $d/5$. We call such a graph a (n, d, λ) algebraic expander. A more intuitive (combinatorial) notion of expansion for a finite graph G is the *edge expansion*, defined as:

$$c_E(G) := \inf_{\substack{S \subset V_G, \\ |S| < |G|/2}} \frac{|\partial_E S|}{|S|},$$

where $\partial_E S$ denotes the set of edges with exactly one vertex in S . It is known (due to Tanner, Alon and Milman [5], [23]) that algebraic expansion implies also a lower bound on the edge expansion: for a (n, d, λ) algebraic expander it holds that $c_E(G) \geq \frac{d-\lambda}{2}$. (There is also an inequality in the opposite direction: $c_E(G) \leq \sqrt{2d(d-\lambda)}$, see [1] for details).

Expander graphs received a considerable amount of attention in the literature in recent years, mostly because these graphs have numerous applications in theoretical computer science; see, for example, [6, 15, 22, 20]. It is well known

*Department of Computer Science and Applied Mathematics Weizmann Institute, Rehovot 76100, Israel. eran.ofek@weizmann.ac.il

that for any fixed $d \geq 3$, random d -regular graphs of size n are asymptotically almost surely expanders, as n grows. The problem of constructing infinite families of bounded degree expanders is more difficult, and there are several known constructions of this type [18, 17, 21, 9]. The result in this paper applies to the constructions of [17, 21, 9].

Various applications of expanders rely on their fault-tolerance as networks. For example, after deleting an appropriate constant fraction of the edges (arbitrarily), the remaining graph still contains some linear size connected components or some linear size paths; see [3, 24]. We show that for algebraic expanders if the deletions are random and independent then with high probability (with probability that tends to 1 as n increases) the giant component has an edge expansion proportional to $\frac{c\sqrt{d}}{\log n}$. Up to constants, this bound is tight since with probability bounded away from 0, the giant component will contain a $\frac{\log n}{2c\sqrt{d}}$ long "chain" of vertices each of them, except the first and the last ones, has degree of exactly 2 in the giant component. The edge expansion of such a "chain" is $\frac{4c\sqrt{d}}{\log n}$.

Given a graph G , we use G_p to denote the subgraph of G obtained by retaining each edge of G independently with probability p . The graph G_p is the *percolated* version of G . For any graph property of G one can ask if this property is almost surely retained in G_p . A well studied example is the existence and the uniqueness of a giant component. Roughly speaking, a giant component is a connected component of G_p that contains a linear fraction of vertices. A question of the same flavour can be asked also for an infinite graph G : for which values of p , G_p is likely to contain an infinite cluster (connected component)? is the infinite cluster likely to be unique? For several types of graphs, e.g. the d dimensional grid, the finite/infinite versions turned out to be related. For many interesting graphs the probability of containing a giant component (or infinite cluster in the infinite case) exhibits a sharp threshold around some value called the *critical probability* (this is due to 0/1 laws). The critical probability is denoted by p_c . For values of p slightly smaller than p_c the probability for giant component is close to 0 and for p slightly larger than p_c the probability for giant component is close to 1. Benjamini and Schramm [8] showed that if G is an infinite graph with a positive vertex Cheeger constant ¹ $c_V(G) > 0$, then the critical probability for the existence of an infinite cluster in G_p is $< \frac{1}{1+c_V(G)} < 1$. They also observed that their proof can be applied to the finite case. Their technique can be easily applied also to the edge Cheeger constant as shown in [19].

A family of expanders is a sequence of d -regular graphs $G(n)$, where $G(n)$ has n vertices and edge expansion of least $b > 0$ (independent of n). Alon, Benjamini and Stacey [2] studied the existence and uniqueness of a giant component when percolation is applied to families of edge expander graphs. One of their results is about expander families with increasing girth (the girth of a graph G is the

¹the vertex Cheeger constant is defined as $c_V(G) := \inf_{\substack{S \subset V_G, \\ |S| < |G|/2}} \frac{|\partial_V S|}{|S|}$, where $\partial_V S$ consists of all vertices of $V \setminus S$ that have neighbors in S .

length of minimum size cycle in it). They show that for an expander family $G(n)$, with $\text{girth}(G(n))$ that goes to infinity as n increases, the critical probability p_c for the existence (and uniqueness) of a giant component is exactly $\frac{1}{d-1}$. Specifically, for any fixed ϵ , and $p \geq \frac{1+\epsilon}{d-1}$ w.h.p. (with high probability, i.e. with probability that goes to 1 as n , the size of graph, goes to infinity) G_p contains a connected component with a linear number of vertices. The fraction of vertices in the giant component depends on ϵ . The girth, the edge expansion, d and ϵ influence the speed in which the probability for a g.c. (giant component) goes to 1. For $p \leq \frac{1-\epsilon}{d-1}$, w.h.p. G_p breaks into connected components of sub-linear size. It is further shown in [2] that if $G(n)$ is an infinite family of d -regular graphs, each one with edge expansion of at least $b > 0$, then for p sufficiently close to 1 (which depends on b) $G(n)_p$ is w.h.p. a $\frac{1}{\log n}$ expander. They leave as an open problem the values of p which are close (from above) to the critical probability p_c . Note that p_c can be as small as $\frac{1}{d-1}$ (for example, an infinite family of expanders with girth that goes to infinity).

Instead of analyzing the expansion of the giant component of G_p , one may ask whether the giant component contains a "large" subgraph that roughly retains the expansion of G . A question of this flavour was studied in [7], where they used G to represent a network that have faulty nodes (in this context G_p denotes the graph derived from G by removing each node with probability $1-p$). One of the problems studied in [7] is: for which values of p is G_p likely to contain a linear sized subgraph that retains (up to a constant factor) the vertex expansion of G ? Note that the new question allows us to remove from the giant component the bad parts that have poor expansion. For the d -dimensional mesh they show that when $p \geq 1 - \frac{1}{16ed^{16}}$ the graph G_p almost surely contains a subgraph of size $\geq n/2$ whose expansion is at least $\frac{1}{4d}$ times the expansion of the d -dimensional mesh.

Percolation of (n, d, λ) graphs has been previously studied by Frieze, Krivelevich and Martin [13]. They gave tight results about the existence and the uniqueness of the giant component when $\lambda = o(d)$. Specifically, for $p < \frac{1}{d}$ the graph G_p almost surely contains only connected components of size $O(\log n)$. For $p > \frac{1}{d}$ the graph G_p has almost surely a unique giant component and all other components are of size at most $O(\log n)$.

1.1 Our result

Theorem 1.1. *Let $d \geq d_0$ be a fixed constant, let G be a $(n, d, c\sqrt{d})$ algebraic-expander and let $p \geq \frac{5c}{\sqrt{d}}$ (assuming $c < \frac{\sqrt{d}}{5}$). W.h.p. the edge expansion of the giant component in G_p is at least $\frac{c\sqrt{d}}{122 \log n}$.*

Theorem 1.1 implies that in the case of algebraic expanders even when $p \ll 1$ the giant component has edge expansion $\geq \frac{1}{\log n}$. In contrast, the result in [2] is based on a weaker assumption (edge expansion greater than ϵ) but it implies that the giant component has edge expansion $\geq \frac{1}{\log n}$ only for values of p close to 1. While Theorem 1.1 requires a somewhat stronger assumption (spectral

gap) from G , it implies that the giant component in G_p has expansion $\geq \frac{1}{\log n}$ also for values of p close to 0 (depending on the degree d and $\lambda = c\sqrt{d}$).

The main idea in the proof of Theorem 1.1 is to iteratively remove from G_p vertices of low degree until we are left with an induced subgraph G_p^k that has minimal degree $\geq \frac{3pd}{5}$. Using known techniques it can be shown that for large enough d this process removes only a small fraction of the vertices. Moreover, the obtained subgraph G_p^k has edge expansion bounded away from 0. To show that the giant component of G_p has expansion $\geq \frac{1}{\log n}$ (which is best possible up to constants) we need to handle sets of the giant component that contain vertices from $OUT := V \setminus G_p^k$. To do this it is enough to show that in the graph induced by G_p on OUT , the connected components are smaller than $\log n$. Following the work of Alon and Kahale [4] on coloring random 3-colorable graphs, several papers [10, 12, 14, 11] dealt with similar versions of this problem: proving that a set OUT which is the outcome of some procedure applied to a random graph has no large connected components. Yet, in all the above cases the graph model was a simple variant of the $G_{n,p}$ model. Our result can be thought of as a "derandomization" of the previous results as we deal with predetermined constant degree "pseudo-random" graphs for which there is less randomness in the induced model.

Remarks:

1. Possibly, Theorem 1.1 can be extended also for values of $p > \frac{c}{\sqrt{d}}$, using the same proof technique. To keep the proof simple, we did not try to optimize this constant.
2. The requirement in Theorem 1.1 that d is a fixed constant is too strong. For d which is a function of n , it can be replaced by the condition $\lambda = c\sqrt{d} = \omega(\log d)$ (inducing a stronger version of Theorem 1.1. We omit the details). Note that for $d = o(n)$ it holds that $c\sqrt{d} = \omega(\log d)$ (it is known that for any graph $\lambda \geq \sqrt{d - \frac{d^2}{n-1}}$).

1.2 Notation

For a set $U \subset V$, $G[U]$ denotes the subgraph induced by the edges of G on the vertices of U . We use $e(S, S)$ to denote twice the number of edges having only vertices in S . The graph induced by retaining each edge of G independently with probability of p is denoted by G_p . The degree of a vertex v inside a graph G is denoted by \deg_v^G . The second largest eigenvalue (in absolute value) of G is denoted by λ . We use the term with high probability (w.h.p) to denote a sequence of probabilities that tends to 1 as n , the size of G , goes to infinity.

1.3 Spectral gap and pseudo-randomness

In the following proofs we will use the fact that a graph G with a noticeable spectral gap is pseudo-random. This is formulated by the following Lemma also

known as the expander mixing lemma (see [6] for proof).

Lemma 1.2. *Let G be a d -regular graph with second largest (in absolute value) eigenvalue λ . Then, for any $S, T \subseteq V$:*

$$|e(S, T) - \frac{d}{n}|S||T|| < \lambda\sqrt{|S||T|},$$

where $e(S, T)$ is the number of directed edges from S to T in the adjacency matrix of G .

In terms of undirected edges (when G is undirected), $e(S, T)$ equals the number of edges between $S \setminus T$ to T plus twice the number of edges that contain only vertices of $S \cap T$.

Corollary 1.3. *Let G be a $(n, d, c\sqrt{d})$ algebraic expander. For any set U of size $\leq \delta n$ the average degree in $G[U]$ is at most $c\sqrt{d}(1 + \frac{\delta\sqrt{d}}{c})$.*

Proof. The number of edges inside $G[U]$ is $e(U, U)/2$ since every edge whose both endpoints are in U is in fact two directed edges from U to U . It follows that the average degree in $G[U]$ is $\frac{e(U, U)}{|U|}$. By the expander mixing lemma:

$$e(U, U) \leq \frac{d|U|^2}{n} + c\sqrt{d}|U| \leq c\sqrt{d}|U| \left(1 + \frac{\sqrt{d}|U|}{cn} \right) \leq c\sqrt{d}|U|(1 + \frac{\delta\sqrt{d}}{c}).$$

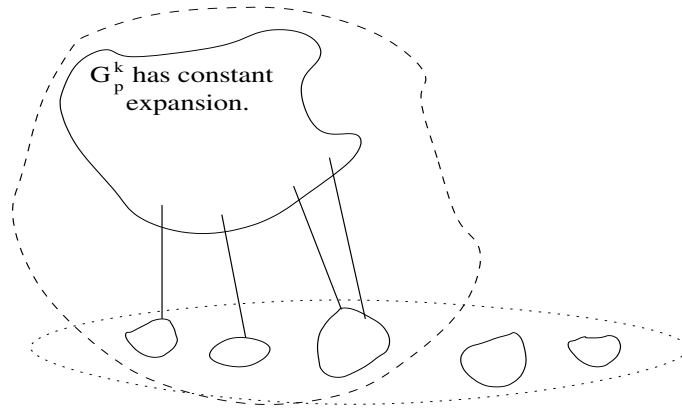
□

We will frequently use the fact that for a set U of size δn , where $\frac{\delta\sqrt{d}}{c} \ll 1$, the average degree in $G[U]$ is bounded by roughly $c\sqrt{d}$ (this follows from Corollary 1.3). When c is close to its smallest possible value for constant degree graphs (i.e. $\lambda = c\sqrt{d} \approx 2\sqrt{d-1}$, see [1] for details) there is a slightly stronger bound on the density of small sets given by [16]. We do not use this stronger bound as it gives asymptotically the same result for large values of c .

2 Proof of Theorem 1.1

We use a process similar to [24] which is aimed to reveal a large edge expanding subgraph of G_p . Let $S_0 = \{v \in G_p : \deg_v^{G_p} \notin [\frac{4pd}{5}, \frac{6pd}{5}]\}$. We begin by removing from G_p the vertices of S_0 . The new induced graph, which we denote by G_p^0 , may contain vertices whose degree in G_p^0 is $< \frac{4}{5}pd$, because removing S_0 affects the degrees of the remaining vertices. We then obtain a sequence of subgraphs G_p^i by iteratively removing from G_p^{i-1} any vertex whose degree in G_p^{i-1} is $< \frac{3}{5}pd$. The process stops once the minimal degree in the remaining graph G_p^k is at least $\frac{3}{5}pd$. The above process induces a sequence $S_0 \subset S_1 \subset \dots \subset S_k$, where $S_i = V \setminus V(G_p^i)$. Note that S_i differs from S_{i-1} by a single vertex whose degree in G_p^{i-1} is $< \frac{3}{5}pd$ (and thus it is removed). We denote the set S_k by OUT .

The giant component (inside the dashed line).



OUT (inside the dotted line),
contains only small connected components.

Figure 1: The structure of G_p .

Remark: Since d is fixed and $n \rightarrow \infty$ it holds that the constant c (from $\lambda = c\sqrt{d}$) is at least 1 (in fact $c \geq 1 - o(1)$ even if d grows with n but it is $o(n)$). We will use this fact occasionally.

2.1 Proof overview

The main idea in the proof of Theorem 1.1 is as follows. We remove from G_p low degree vertices until the induced graph G_p^k has a large enough minimal degree. We first show that G_p^k itself has edge expansion of at least $\frac{pd}{13}$ and contains almost all the vertices of G (thus it must be contained in a giant component); this part of the proof uses standard techniques. We then show that in $G_p[OUT]$ the largest connected component is of size at most $\log n$ (this implies the uniqueness of the giant component). The above two facts imply that any set that belongs to the giant component and is entirely in $V(G_p^k)$ or entirely in OUT , has expansion $\Omega(\frac{1}{\log n})$. Using also the property that any vertex of $V(G_p^k)$ has at most $\frac{3pd}{5}$ neighbors in OUT we then prove that any set of the giant component with size $\leq n/2$ has edge expansion $\Omega(\frac{1}{\log n})$.

2.2 Proof details

The expected degree in G_p , which is pd , is large enough so that only a few vertices are removed in the process of extracting G_p^k (namely OUT is small). To show this we use the following idea from [24, 4]. Initially, the set S_0 is small (roughly $e^{-\Omega(c\sqrt{d})n}$). Every vertex which is removed in the iteration process has at least $\frac{pd}{5}$ edges to vertices that were previously removed. Thus, if the

iteration process is too long, namely $i = 2|S_0|$, the set S_i , whose cardinality is $3|S_0| \leq 3e^{-\Omega(c\sqrt{d})}n$ has average degree of at least $2\frac{2|S_0|pd/5}{3|S_0|} > \frac{4}{3}c\sqrt{d}$. On the other hand, Corollary 1.3 implies that the average degree of such set is at most $c\sqrt{d}(1 + 3e^{-\Omega(c\sqrt{d})}\frac{\sqrt{d}}{c}) < \frac{4}{3}c\sqrt{d}$, hence the number of iterations k must be smaller than $2|S_0|$ (note that we used $3e^{-\Omega(c\sqrt{d})}\frac{\sqrt{d}}{c} \ll 1$, which holds for large enough d). Lemma 2.1 sums up the bound on $|OUT|$.

Lemma 2.1. *W.h.p. the number of vertices in G_p^k is at least $(1 - e^{-\frac{1}{12}c\sqrt{d}})n$.*

The full proof of Lemma 2.1 is deferred to Section 2.3.

The minimal degree in G_p^k is $\frac{3pd}{5} \geq 3c\sqrt{d}$. A set S smaller than $\frac{cn}{\sqrt{d}}$ contains at most $c|S|\sqrt{d}$ internal edges (by Corollary 1.3), thus for such set the expansion is at least $\frac{3pd}{5} - 2c\sqrt{d} \geq \frac{pd}{5}$. To establish the edge expansion of larger sets, a standard argument using the Chernoff and union bounds suffices. In Section 2.3 we give the full proof of the following Lemma.

Lemma 2.2. *W.h.p. the graph G_p^k has an edge expansion of at least $\frac{pd}{13}$.*

We next show that w.h.p. the connected components in $G_p[OUT]$ are of size at most $\frac{122\log n}{c\sqrt{d}}$. A direct ‘brute force’ approach using the union bound over all possible trees of size $\frac{122\log n}{c\sqrt{d}}$ does not seem to work here because we don’t have a good enough upper bound on the probability that a fixed tree T is in OUT . Note that we can not simply claim that every vertex in OUT has a low degree in G_p (if this were true then probably a simple argument would have sufficed). It may be the case that a vertex in OUT has high degree in G_p but it is connected (directly or via other vertices) to vertices of low degree in G_p .

Several papers [10, 12, 14, 11] dealt with similar versions of this problem (proving that a set OUT which is the outcome of some procedure applied to a random graph has no large connected components) and all of them use the technique from [4]. We slightly simplify this proof technique by introducing *balanced* connected components.

Definition 2.3. *A set of vertices is called balanced if at least 1/3 fraction of its vertices belong to S_0 .*

It turns out that (w.h.p.) every connected component of $G_p[OUT]$ is balanced (Lemma 2.4). Thus, in order to show that $G_p[OUT]$ has no large connected components, it is enough to show that $G_p[OUT]$ has no large balanced connected components. This turns out to be easier.

Lemma 2.4. *With high probability any connected component of $G_p[OUT]$ is balanced.*

Proof. Let $U \subseteq OUT$ be a connected component of $G_p[OUT]$. Every vertex removed in iterations $1, 2, \dots, k$ has at least $\frac{pd}{5} \geq c\sqrt{d}$ edges to previously removed vertices. Thus, if at least 2/3 of the vertices of U are added during the iterations then the average degree in $G_p[U]$ is at least $\frac{4}{3}c\sqrt{d}$. On the other hand, since

$|U| < e^{-\frac{1}{12}c\sqrt{d}}n$ (this follows from Lemma 2.1 and the fact that $U \subset OUT$) by Corollary 1.3 the average degree in $G[U]$ is at most

$$c\sqrt{d}(1 + \frac{\sqrt{d}}{c}e^{-c\sqrt{d}/12}) < 1.1c\sqrt{d},$$

which is a contradiction (the last inequality holds for sufficiently large d). \square

Having established that (w.h.p.) every connected component in $G_p[OUT]$ is balanced, our problem (showing that w.h.p. there are no large connected components in $G_p[OUT]$) is reduced to showing that w.h.p. any balanced connected component of $G_p[OUT]$ is of size $\leq \frac{122 \log n}{c\sqrt{d}}$. We further simplify our problem by noting that every balanced connected component of $G_p[OUT]$ of size $> \frac{122 \log n}{c\sqrt{d}}$ contains some balanced tree whose size is in $[\frac{61 \log n}{c\sqrt{d}}, \frac{122 \log n}{c\sqrt{d}}]$. This follows from Lemma 2.5 whose proof is deferred to Section 2.3. Thus, in order to show that $G_p[OUT]$ has no maximal balanced connected component of size $\frac{122 \log n}{c\sqrt{d}}$ it is enough to show that it has no balanced tree whose size is in the range $[\frac{61 \log n}{c\sqrt{d}}, \frac{122 \log n}{c\sqrt{d}}]$.

Lemma 2.5. *Let $G = (V, E)$ be a connected graph whose vertices are partitioned into two sets: S and $V \setminus S$. Assume $|S| \geq \frac{|V|}{k}$, where k is an integer. For any $1 \leq t \leq |V(G)|/2$ there exists a tree whose size is in $[t, 2t - 1]$ and at least $\frac{1}{k}$ fraction of its vertices are from S .*

To summarize, in order to show that w.h.p. $G_p[OUT]$ has no connected components of size $\frac{122 \log n}{c\sqrt{d}}$ we need only to prove the following Claim.

Claim 2.6. *W.h.p. $G_p[OUT]$ has no balanced trees of size in $[\frac{61 \log n}{c\sqrt{d}}, \frac{122 \log n}{c\sqrt{d}}]$.*

Proof. Fix $t \in [\frac{61 \log n}{c\sqrt{d}}, \frac{122 \log n}{c\sqrt{d}}]$. We want to bound the probability that $G_p[OUT]$ contains a balanced tree of size t

$$\sum_{\substack{T \text{ is a tree} \\ \text{of size } t}} \Pr[T \subset OUT \wedge T \text{ is balanced}]. \quad (1)$$

The number of trees of size t in a d -regular graph G is at most nd^{2t} , since each tree can be uniquely mapped into a closed path of length $2t$. For each tree of size t there are at most 2^t ways of choosing a subset of size $\geq t/3$. Any fixed set of size $\geq t/3$ is in S_0 with probability of at most $e^{-\frac{1}{2} \frac{1}{25} pdt/6}$ (the number of edges touching a set of size l is $m \in [\frac{dl}{2}, dl]$). The expected number of surviving edges is pm . It is not hard to see that if the set is contained in S_0 , then the number of surviving edges that touch it is at most $\frac{4pm}{5}$. Thus, the probability

that there is a balanced tree of size t in OUT is at most:

$$\begin{aligned}
tnd^{2t}2^t e^{-\frac{1}{300}pdt} &\leq \exp(\log t + \log n + 2t \log d + t - \frac{1}{300}pdt) \\
&\leq \exp(\log n + 3t \log d - \frac{c\sqrt{dt}}{60}) \stackrel{\substack{c \geq 1, \\ d \text{ sufficiently} \\ \text{large}}}{\leq} \exp(\log n - 0.99\frac{c\sqrt{dt}}{60}) \\
&\stackrel{t \geq \frac{61 \log n}{c\sqrt{d}}}{\leq} \exp(-0.065 \log n).
\end{aligned}$$

Applying the union bound over all values of t (the range is at most $\log n$) completes the proof. \square

Corollary 2.7. *W.h.p. all the connected components of $G_p[OUT]$ are of size at most $\frac{122 \log n}{c\sqrt{d}}$.*

Proof. By Claim 2.6 w.h.p. $G_p[OUT]$ has no balanced trees of size in $[\frac{61 \log n}{c\sqrt{d}}, \frac{122 \log n}{c\sqrt{d}}]$. It then follows, using Lemma 2.5, that w.h.p. $G_p[OUT]$ has no balanced connected components of size $\geq \frac{122 \log n}{c\sqrt{d}}$. By Lemma 2.4 w.h.p. every connected component in $G_p[OUT]$ is balanced. Combining these two facts we derive that w.h.p. $G_p[OUT]$ has no connected component of size $\geq \frac{122 \log n}{c\sqrt{d}}$. \square

Since (w.h.p.) all connected components of $G_p[OUT]$ are of size at most $\frac{122 \log n}{c\sqrt{d}}$, sets from OUT that belong to the giant component have expansion of at least $\frac{c\sqrt{d}}{122 \log n}$. To complete the proof of Theorem 1.1 it remains to handle sets of the giant component that intersects both $V(G_p^k)$ and OUT .

Lemma 2.8. *W.h.p. any set S that belongs to the giant component, whose size is at most $n/2$, has edge expansion of at least $\frac{c\sqrt{d}}{122 \log n}$.*

Proof. We will use the following two properties (which hold w.h.p.):

- (1) subsets of the g.c. (giant component) which are entirely in $V(G_p^k)$ have expansion of at least $\frac{pd}{13} \geq \frac{c\sqrt{d}}{122 \log n}$ (see Lemma 2.2).
- (ii) connected components of the g.c. which are entirely in OUT have size of at most $\frac{c\sqrt{d}}{122 \log n}$ (see Corollary 2.7).

The above two properties guarantee that a subset of the g.c. which is entirely in $V(G_p^k)$ or entirely in OUT has the desired expansion. We remind the reader that $OUT = V \setminus V(G_p^k)$ (see the definition of OUT at the beginning of Section 2). Now let S be a set of the g.c. that intersects both OUT and $V(G_p^k)$. Denote by \bar{S} the complement of S in the giant component. Denote by S_1, S_2 the intersection of S with $OUT, V(G_p^k)$ respectively. We further partition S_1 into S_{11}, S_{12} by classifying the connected components of $G[S_1]$ as follows. All the connected components of $G_p[S_1]$ that have at least one edge into \bar{S} belong to

S_{11} and all the connected components of $G[S_1]$ that have no edges to \bar{S} belong to S_{12} (each of the connected component in S_{12} must have at least one edge to S_2 as otherwise it does not belong to the giant component).

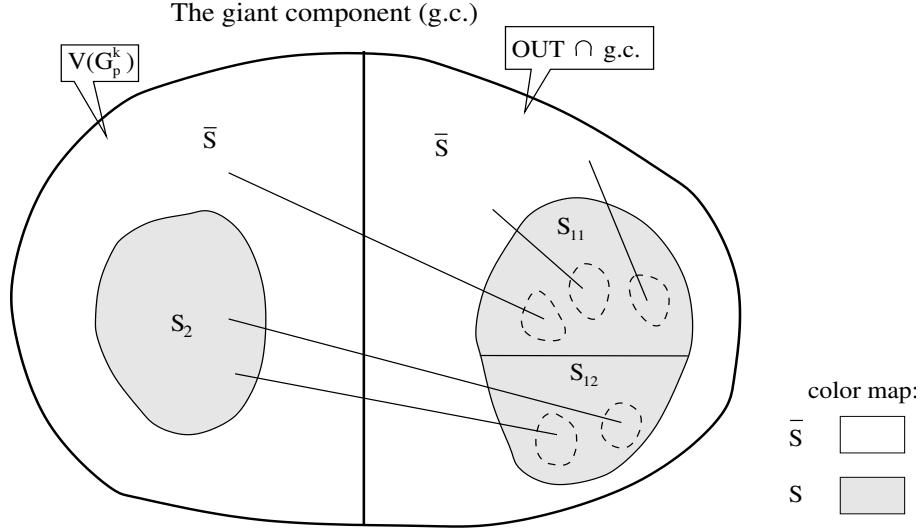


Figure 2: The structure of $S = S_{11} \uplus S_{12} \uplus S_2$.

To show that S has an expansion of at least $\frac{c\sqrt{d}}{122\log n}$, it is enough to show that:

$$|E(S_{11}, \bar{S})| \geq |S_{11}| \frac{c\sqrt{d}}{122\log n}, \quad (2)$$

$$|E(S_{12} \cup S_2, \bar{S})| \geq |S_{12} \cup S_2| \frac{1}{18}. \quad (3)$$

The first inequality holds since every connected component of $G[OUT]$ has size of at most $\frac{122\log n}{c\sqrt{d}}$. The second inequality is derived as follows: $|E(S_{12} \cup S_2, \bar{S})| \geq |E(S_2, \bar{S})| \geq |S_2|pd/13$. Thus

$$\frac{|E(S_{12} \cup S_2, \bar{S})|}{|S_{12}| + |S_2|} \geq \frac{|S_2|pd/13}{|S_{12}| + |S_2|} \geq \frac{1}{\frac{13}{pd}(|S_{12}|/|S_2| + 1)} \geq \frac{1}{18},$$

where the last inequality holds because every vertex of G_p^k has at most $\frac{6pd}{5}$ neighbors in OUT and every connected component of $G[S_{12}]$ must have at least one edge to $V(G_p^k)$ (as otherwise it does not belong to the giant component). \square

2.3 Proofs of lemmas 2.1, 2.2, 2.5

The proofs of Lemmas 2.1 and 2.2 are rather standard and are based on the fact that every small enough set S (of size $\ll \frac{cn}{\sqrt{d}}$) contains at most $|S|c\sqrt{d}(1+0.1)/2$ internal edges (Corollary 1.3).

Proof of Lemma 2.1. A fixed vertex v belongs to S_0 with probability of at most $e^{-\frac{1}{2}(\frac{1}{5})^2 pd} \leq e^{-\frac{c\sqrt{d}}{10}}$. Thus, the expected size of S_0 is at most $e^{-\frac{1}{10}c\sqrt{d}}n$. We claim that with probability of $1 - o(1)$ the cardinality of S_0 is at most $e^{-\frac{1}{12}c\sqrt{d}}n$.

We use the edge exposure martingale to prove that $|S_0|$ is concentrated around its expectation. Fix some order on the $m = nd/2$ edges of G . Denote by B_i the set containing first i edge so that

$$\emptyset = B_0 \subset B_1 \subset B_2 \subset \dots \subset B_m = E(G).$$

Our probability measure is a product measure induced by retaining each edge of $E(G)$ independently with probability p ; recall that G_p denotes a random instance. The support of our probability measure is $\{0, 1\}^{E(G)}$ (for every edge in $E(G)$ there is a value in $\{0, 1\}$ indicating if the edge retained or not). The value $|S_0|$ can be thought as a function $f : \{0, 1\}^{E(G)} \rightarrow \mathbb{Z}$. Let X_i be the expectation of $f(G_p)$ after exposing the values of the edges in B_i . The sequence X_0, X_1, \dots, X_m is a Doob Martingale. Note that $X_0 = \mathbb{E}[f(G_p)]$ and X_m is the random variable $f(G_p)$. To use the Azuma inequality we need to upper bound the martingale difference $|X_{i+1} - X_i|$ (for $i = 0, \dots, m-1$). The function f (i.e. $|S_0|$) is 2-Lipschitz with respect to the sequence B_i : for any graph taken from $\{0, 1\}^{E(G)}$ changing its value on $B_i \setminus B_{i-1}$ (i.e removing/adding edge number i) can change the value of f by at most 2. Since the probability space G_p is a product measure, the Martingale difference is bounded by 2 (see [6] theorem 4.7.1). By Azuma's inequality, if the Martingale difference is bounded by Δ , then for any $\zeta > 0$ it holds:

$$\Pr[X_m > X_0 + \zeta] \leq e^{-\zeta^2/(2m\Delta^2)}. \quad (4)$$

Substituting $\zeta = e^{-\frac{1}{10}c\sqrt{d}}n$, $\Delta = 2$, $m = nd/2$ we derive that w.h.p. $|S_0|$ is at most $2e^{-\frac{1}{10}c\sqrt{d}}n$ (we also use $X_0 \leq e^{-\frac{1}{10}c\sqrt{d}}n$). From here we will assume $|S_0| \leq 2e^{-\frac{1}{10}c\sqrt{d}}n$

We next show that the number of vertices removed after removing S_0 (that is k) is at most $2|S_0|$. Every vertex that is removed in the iterative process has at least $\frac{pd}{5} \geq c\sqrt{d}$ edges connecting it to previously removed vertices, because its degree drops from at least $\frac{4pd}{5}$ (as it does not belong to S_0) down to some value below $\frac{3pd}{5}$ (at the point it was removed).

By contradiction, assume that $k \geq 2|S_0|$. Consider the situation immediately after iteration $i = 2|S_0|$. Recall that S_i is the set of vertices removed so far. The average degree in $G_p[S_i]$ is at least $2\frac{2|S_0|pd/5}{3|S_0|} \stackrel{pd \geq \frac{5c}{\sqrt{d}}}{>} \frac{4}{3}c\sqrt{d}$ and thus the average degree in $G[S_i]$ is at least $\frac{4}{3}c\sqrt{d}$. At this point $|S_i| = 3|S_0| \leq 6e^{-\frac{1}{10}c\sqrt{d}}n$ and by Corollary 1.3 the average degree in $G[S_i]$ is at most $c\sqrt{d}(1 + \frac{6e^{-\frac{c\sqrt{d}}{10}}\sqrt{d}}{c}) < 1.1c\sqrt{d}$ (for sufficiently large d), which leads to contradiction.

It thus follows that $|S_k| < 3|S_0| \leq 6e^{-\frac{1}{10}c\sqrt{d}}n \leq e^{-\frac{1}{12}c\sqrt{d}}n$ (where the last inequality holds for sufficiently large d). \square

Proof of Lemma 2.2. The proof is divided into two parts. First consider sets of cardinality $\leq \frac{cn}{\sqrt{d}}$. Fix a set $S \subset V$. The edge expansion of S (in G_p^k) is at least:

$$\sum_{v \in S} \deg(v) - e(S, S),$$

(remember that $e(S, S)$ is twice the number of edges inside S). Every vertex v of G_p^k has degree of at least $\frac{3pd}{5} \geq 3c\sqrt{d}$ in G_p^k . By Corollary 1.3 $e(S, S)$ in G is at most $|S|c\sqrt{d}(1+1) = 2|S|c\sqrt{d}$. It follows that the edge expansion of S (in G_p) is at least $\frac{3pd}{5} - 2c\sqrt{d} \geq \frac{pd}{5}$.

Consider now a set $S \subset V$ of size αn such that $\frac{c}{\sqrt{d}} < \alpha \leq \frac{1}{2}$. By the expander mixing lemma, the number of edges between S and $V \setminus S$ in G is at least:

$$\begin{aligned} \alpha(1-\alpha)dn - c\sqrt{d}\sqrt{\alpha(1-\alpha)}n &= \alpha(1-\alpha)dn \left(1 - \frac{c}{\sqrt{d\alpha(1-\alpha)}}\right) \\ &\geq \alpha(1-\alpha)dn/3. \end{aligned}$$

The last inequality follows from $\frac{c}{\sqrt{d}} \leq \alpha \leq \frac{1}{2}$ and $c \leq \frac{\sqrt{d}}{5}$ (we use here the assumption $p \geq \frac{5c}{\sqrt{d}}$). The number of edges between S and $V \setminus S$ in G_p is smaller than $p\alpha(1-\alpha)dn/6$ with probability of at most $e^{-\frac{1}{8}p\alpha(1-\alpha)dn/3}$ (follows from the Chernoff bound). The number of subsets of size αn is at most $\left(\frac{ne}{\alpha n}\right)^{\alpha n} = e^{\alpha n(1+\log \frac{1}{\alpha})}$. Thus the probability for a ‘bad’ set of size αn is at most:

$$\begin{aligned} \exp\left(\alpha n(1+\log \frac{1}{\alpha}) - \frac{1}{8}p\alpha(1-\alpha)dn/3\right) &\leq \exp\left(\alpha n(1+\log \frac{1}{\alpha}) - \frac{5c\sqrt{d}(1-\alpha)}{24}\right) \\ &\leq \exp\left(-\alpha cn\sqrt{d}/10\right), \end{aligned}$$

where the last inequality holds for large enough d (and using also the fact $\frac{1}{\alpha} < \sqrt{d}$). Summing over all values of αn gives that w.h.p. there is no bad set. In other words, every set of size in $[\frac{cn}{\sqrt{d}}, \frac{n}{2}]$ has an edge expansion of at least $\frac{pd}{12}$ in G_p . Since the number of edges that contain at least one vertex from OUT is bounded by $de^{-\frac{c\sqrt{d}}{12}}n$, we conclude that any subset U of $V(G_p^k)$ with size $\geq \frac{cn}{\sqrt{d}}$ has edge expansion of at least

$$\left(\frac{pd}{12}|U| - de^{-c\sqrt{d}/12}n\right)/|U| \geq \frac{pd}{13}.$$

□

Proof of lemma 2.5. We use the following well known fact: any tree T contains a center vertex v such that each subtree hanged on v contains strictly less than half of the vertices of T .

Let T be an arbitrary spanning tree of G . Observe that at least a fraction of $\frac{1}{k}$ of its vertices belongs to S . We show that any such tree T has a subtree T' of size $|T|/2 < |T'| < |T|$ with at least a fraction of $\frac{1}{k}$ of its vertices in S . Repeating this argument until the first time that $|T'| < 2t$ proves the lemma.

Let v be the center of T and let T_1, \dots, T_k be the subtrees hanging on v . Let T_j be the subtree with the smallest fraction of S vertices, and take T' to be the tree that remains by removing T_j from T . By properties of the center vertex, $|T|/2 < |T'| < |T|$, as desired. Moreover, if the fraction of S vertices in T_j is $\frac{1}{k}$ or less, then clearly the fraction of S vertices in T' is at least $\frac{1}{k}$. If the fraction of S vertices in T_j is strictly more than $\frac{1}{k}$, then this holds also for all other subtrees of T . By integrality of k , this implies that the fraction of S vertices in T' is at least $\frac{1}{k}$ (because then $k|S \cap T'| > |T'| - 1$ implies $k|S \cap T'| \geq |T'|$). \square

3 Open problems

We were able to show that a percolation applied to a family of d -regular expander graphs with eigenvalue gap retains some expansion properties of the original graphs, even when p is close to 0 (depending on d). There are still many open problems, we list here two of them:

1. Find other classes of expander families that retain expansion properties when percolated with values of p close to 0. For example, a family of expanders with girth that goes to infinity (for such a family some result is given at [2] for p close to 1).
2. Assuming d is a fixed constant, there is a gap between $\frac{5c}{\sqrt{d}}$, above which the edge expansion is likely to be $> \frac{1}{\log n}$ (Theorem 1.1) and the critical probability $\frac{1}{d}$ for the existence of a giant component (given in [13]).

Acknowledgments

I would like to thank Itai Benjamini for suggesting the problem and for useful discussions. I would like also to thank the anonymous referee for his helpful comments.

References

- [1] N. Alon. Eigenvalues and expanders. *Combinatorica*, 6(2):83–96, 1986.
- [2] N. Alon, I. Benjamini and A. Stacey. Percolation on finite graphs and isoperimetric inequalities. *Annals of Probability*, 32(3):1727–1745, 2004.
- [3] N. Alon and F. R. K. Chung. Explicit construction of linear sized tolerant networks. *Discrete Mathematics*, 72(1-3):15–19, 1988.
- [4] N. Alon and N. Kahale. A spectral technique for coloring random 3-colorable graphs. *SIAM Journal on Computing*, 26(6):1733–1748, 1997.
- [5] N. Alon and V.D. Milman. λ_1 , isoperimetric inequalities for graphs and superconcentrators. *Journal of Combinatorial Theory*, Ser. B, 38:73–88, 1985.

- [6] N. Alon and J. Spencer. *The Probabilistic Method*. John Wiley and Sons, 2000.
- [7] A. Bagchi, A. Bhargava, A. Chaudhary, D. Eppstein and C. Scheideler. The effect of faults on network expansion. *SPAA 2004*, 286–293.
- [8] I. Benjamini and O. Schramm. Percolation beyond \mathbb{Z}^d , many questions and a few answers. *Electronic Communications in Probability*, 1:71-82, 1996.
- [9] Y. Bilu and N. Linial. Lifts, discrepancy and nearly optimal spectral gaps. *FOCS 2004*, 404–412.
- [10] H. Chen and A. Frieze. Coloring bipartite hypergraphs. *IPCO 1996*, 345–358.
- [11] A. Coja-Oghlan. A spectral heuristic for bisecting random graphs. *SODA 2005*, 850–859.
- [12] A. Flaxman. A spectral technique for random satisfiable 3cnf formulas. *SODA 2003*, 357–363.
- [13] A. Frieze, M. Krivelevich and R. Martin. Emergence of a giant component in random subgraphs of pseudo-random graphs. *Random Structures and Algorithms*, 24(1):42–50, 2004.
- [14] A. Goerdt and A. Lanka. On the hardness and easiness of random 4-sat formulas. *ISAAC 2004*, 470–483.
- [15] R. Impagliazzo, N. Nisan and A. Wigderson. Pseudorandomness for network algorithms. *STOC 1994*, 356–364.
- [16] N. Kahale. Eigenvalues and expansion of regular graphs. *Journal of the ACM*, 42(5):1091–1106, 1995.
- [17] A. Lubotzky, R. Phillips and P. Sarnak. Ramanujan Graphs. *Combinatorica*, 8:261–277, 1988.
- [18] G.A. Margulis. Explicit constructions of concentrators. *Problemy Peredači Informacii*, 9(4):71–80, 1973; English transl. *In Problems of Information Transmission*, 325 - 332, 1975.
- [19] I. Pak and T. Smirnova-Nagnibeda. Uniqueness of percolation on nonamenable Cayley graphs. *Comptes Rendus Acad. Sci. Paris, Ser. I Math*, vol. 330(6): 495–500, 2000.
- [20] O. Reingold. Undirected ST-Connectivity in Logspace. *STOC 2005*, 376–385.
- [21] O. Reingold, S. Vadhan and A. Wigderson. Entropy Waves, the Zig-Zag graph product and new constant-degree expanders. *Annals of Mathematics*, 155:157–187, 2002.

- [22] M. Sipser and D.A. Spielman. Expander codes. *IEEE Transactions on Information Theory*, 42(6): 1710-1722, 1996.
- [23] R.M. Tanner. Explicit construction of concentrators from generalized n-gons. *SIAM Journal on Algebraic Discrete Methods*, 5:287-294, 1984.
- [24] E. Upfal. Tolerating linear number of faults in networks of bounded degree. *PODC 1992*, pages 83-89.