

Expansion in Supercritical Random Subgraphs of Expanders and its Consequences

Sahar Diskin ^{*} Michael Krivelevich [†]

December 21, 2023

Abstract

In 2004, Frieze, Krivelevich and Martin [17] established the emergence of a giant component in random subgraphs of pseudo-random graphs. We study several typical properties of the giant component, most notably its expansion characteristics. We establish an asymptotic vertex expansion of connected sets in the giant by a factor of $\tilde{O}(\epsilon^2)$. From these expansion properties, we derive that the diameter of the giant is **whp** $O_\epsilon(\log n)$, and that the mixing time of a lazy random walk on the giant is asymptotically $O_\epsilon(\log^2 n)$. We also show similar asymptotic expansion properties of (not necessarily connected) linear sized subsets in the giant, and the typical existence of a large expander as a subgraph.

1 Introduction

Percolation theory, the study of which was initiated by Broadbent and Hammersley [9] in 1957, is a mathematical discipline researching the following probabilistic model: given a graph G , the percolated subgraph G_p is the random subgraph of G obtained by retaining each edge of G independently with probability p . We refer the reader to [21], [18] and [8] for systematic coverage of percolation theory.

Percolation on several concrete base graphs has been studied extensively. One notable example is percolation on the complete graph K_n , which is equivalent to the classical model of the random graph $G(n, p)$. In their groundbreaking paper [12] from 1960, Erdős and Rényi discovered that $G(n, p)$ (working on its close analogue $G(n, m)$) undergoes a phase transition around probability $p = \frac{1}{n}$: for any constant $\epsilon > 0$, if $p = \frac{1-\epsilon}{n}$ then all the connected components of $G(n, p)$ are typically of size $O(\log n)$, while for $p = \frac{1+\epsilon}{n}$, **whp**¹ there exists a unique component of linear size in $G(n, p)$, usually called the *giant component*. In the same paper [12], Erdős and Rényi also obtained the asymptotic order of this giant component. We refer the reader to [20], [6] and [16] for a systematic coverage of random graphs. Similar phenomenon of an emergence of a unique linear sized component has been discovered in percolation on other concrete base graphs, with one well studied example being the d -dimensional hypercube [1, 7].

When trying to show the emergence of a unique giant component in percolation on a large class of graphs \mathcal{G} , certainly some assumptions on $G \in \mathcal{G}$ have to be made. For example, letting \mathcal{G} be the family of d -regular graphs, $G \in \mathcal{G}$ could be a collection of vertex disjoint cliques of size $d + 1$, having no large connected components to begin with. It is thus necessary to impose

^{*}School of Mathematical Sciences, Tel Aviv University, Tel Aviv 6997801, Israel. Email: sahardiskin@mail.tau.ac.il.

[†]School of Mathematical Sciences, Tel Aviv University, Tel Aviv 6997801, Israel. Email: krivelev@tauex.tau.ac.il. Research supported in part by USA–Israel BSF grant 2018267.

¹With high probability, that is, with probability tending to 1 as n tends to infinity.

some restrictions on the edge-distribution of the graphs. Hence, it may be natural to consider *pseudo-random graphs*.

In this paper, we are concerned with a specific model of pseudo-random graphs called (n, d, λ) -graphs (also called spectral or algebraic expanders). Informally, pseudo-random graphs are graphs whose edge distribution resembles that of random graphs, with the same number of vertices and edge density. An (n, d, λ) -graph is a d -regular graph on n vertices, where its eigenvalues (that is, the eigenvalues of its adjacency matrix) $d = \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ satisfy $\lambda = \max\{|\lambda_2|, |\lambda_n|\}$. The greater the ratio between d and λ , also called the spectral ratio, the more tightly the distribution of the edges of the graph approaches that of a truly random graph $G(n, \frac{d}{n})$. This follows from the expander mixing lemma, due to Alon and Chung [2] (stated as Lemma 2.1 in this paper). Thus, (n, d, λ) -graphs serve frequently as a standard model for pseudo-random graphs. Crucially, we note that, for $d \leq 0.99n$, $\lambda = \Omega(\sqrt{d})$ (see, for example, [23]). Thus, a large spectral ratio implies that d is large as well. We refer to [23] for a comprehensive survey on the subject of pseudo-random graphs. Let us note that throughout the paper, we will be interested in asymptotic statements which hold with high probability as n tends to infinity. Furthermore, d (and thus λ) may depend on n . Moreover, ϵ always stands for a small enough constant which does not depend on n (that is, there exists some absolute constant $\epsilon_0 > 0$ such that for all $0 < \epsilon < \epsilon_0$, the statement holds). In particular, with a slight abuse of notation, when considering an (n, d, λ) -graph, one can formally view it as a sequence of (n, d_n, λ_n) graphs $(G_n)_n$, where for each graph in the sequence the required assumptions on d_n, λ_n hold.

Returning to the subject at hand, the emergence of a unique giant component in percolation on (n, d, λ) -graphs has been shown in 2004 by Frieze, Krivelevich, and Martin [17]. For such graphs, it was shown in [17] that around $p = \frac{1}{d}$ a phase transition occurs, with respect to the sizes of the components, having many properties similar to the phase transition of $G(n, p)$ around $p = \frac{1}{n}$.

To be more concrete, let $y = y(\epsilon)$ be the unique solution in $(0, 1)$ of:

$$y \exp(-y) = (1 + \epsilon) \exp(-(1 + \epsilon)). \quad (1)$$

Theorem 1 of [17] then states the following:

Theorem 1.1 ([17]). *Let $\epsilon > 0$ be a small enough constant and let $\delta > 0$ be such that $\delta \leq \epsilon^4$. Let $G = (V, E)$ be an (n, d, λ) -graph with $\frac{\lambda}{d} \leq \delta$. Let $p = \frac{1+\epsilon}{d}$. Form G_p by retaining every edge of G independently with probability p . Then, **whp**, there is a unique giant component L_1 of asymptotic size $(1 - \frac{y}{1+\epsilon})n$, where y is defined as in (1). **Whp**, all the other components are of size $O(\log n)$.*

Some comments are in order here. In the same paper [17], the authors showed that when $p = \frac{1-\epsilon}{d}$ (for a small constant $\epsilon > 0$), then typically all the connected components are of order $O_\epsilon(\log n)$. Furthermore, this phase transition and the asymptotic order of the giant component are analogous, both quantitatively and numerically, to the classical phase transition in $G(n, p)$ around $p = \frac{1}{n}$. We thus call the regime where $p = \frac{1-\epsilon}{d}$ the *subcritical regime* and the regime where $p = \frac{1+\epsilon}{d}$ the *supercritical regime*. Observe that from (1) and Theorem 1.1, it can be seen that **whp**, $|L_1| = (1 + o_\epsilon(1))2\epsilon n$. In [17], Theorem 1.1 is stated for $d \rightarrow \infty$ and $\frac{\lambda}{d} < \delta$. It can routinely be verified that this pair of conditions can be replaced with $\frac{\lambda}{d} \leq \epsilon^4$ for a small constant $\epsilon > 0$. Indeed, note that $\frac{\lambda}{d} \leq \epsilon^4$ implies $d = \Omega(\epsilon^{-8})$.

The giant components in the aforementioned percolated concrete graphs (such as $G(n, p)$ and the percolated hypercube) are known to be typically far from tree-like, and in fact are fairly rich combinatorial structures, with many interesting properties to study. In the case of $G(n, p)$, there has been extensive research into several properties of the giant, such as the typical diameter of ([13, 31]), and the asymptotic mixing time of a random walk ([15, 4]) on the giant

component. One approach taken in order to study such properties is to examine the expansion properties of the giant component. For example, this has been done in supercritical $G(n, p)$ by Benjamini, Kozma, and Wormald [4] in order to study the mixing time of a random walk on the giant, and very recently by Erde, Kang and Krivelevich [11] in supercritical percolation on the hypercube in order to obtain asymptotic bounds on the diameter and the mixing time of a random walk on the giant. Krivelevich, Reichman, and Samotij [25] studied the expansion properties of perturbed connected graphs, and obtained from them asymptotic bounds on the diameter and the mixing time. In [23], Krivelevich showed how to argue about expansion and to obtain several interesting properties of graphs from their expansion properties.

The aim of this paper is to study the typical expansion properties of the giant component of its random subgraph, and to derive from them several properties of the giant component, such as its asymptotic diameter and mixing time.

With this in mind, we are now able to state and discuss our results.

We start with some notation. Given a graph G , we let $N_G(S)$ be the external neighbourhood of $S \subseteq V(G)$ in G , and let $\partial_G(S)$ be the set of all the edges in G with exactly one endpoint in S (the edge boundary of S in G).

Our first main result regards the vertex-expansion properties:

Theorem 1. *Let $\epsilon > 0$ be a small enough constant. Let $G = (V, E)$ be an (n, d, λ) -graph with $\frac{\lambda}{d} \leq \epsilon^4$. Let $p = \frac{1+\epsilon}{d}$. Form G_p by retaining every edge of G independently with probability p . Let L_1 be the largest component of G_p . Then, there exists an absolute constant $c > 0$ such that **whp** for any $S \subseteq L_1$:*

1. if $\frac{16 \ln n}{\epsilon^2} \leq |S| \leq \frac{\epsilon^2 n}{15}$ and S spans a connected subgraph in G_p , then:

$$\left| N_{G_p}(S) \right| \geq \frac{c\epsilon^2 |S|}{\ln\left(\frac{1}{\epsilon}\right)};$$

and,

2. if $\frac{\epsilon^2 n}{15} \leq |S| \leq \frac{12\epsilon n}{11}$, then:

$$\left| N_{G_p}(S) \right| \geq \frac{c\epsilon^2 |S|}{\ln^2\left(\frac{1}{\epsilon}\right)}.$$

Let us first note that it is necessary to require in the first part of the statement above that the sets of up to a small linear size are connected. Indeed, for simplicity, let us consider the classical model of $G(n, p)$. There it can be seen (for example, from the structural result of Ding, Lubetzky, and Peres [10]) that typically one can find a non-connected subset of the giant component which is of a linear size, but expands by less than a $\tilde{O}(\epsilon^2)$ factor.

Clearly, the vertex-expansion properties of Theorem 1 imply that the same holds for edge-expansion as well. However, for linear sized (not necessarily connected) subsets, we can obtain a slightly better edge-expansion factor. Indeed, we have the following asymptotic edge-expansion property:

Theorem 2. *Let $\epsilon > 0$ be a small enough constant. Let $G = (V, E)$ be an (n, d, λ) -graph with $\frac{\lambda}{d} \leq \epsilon^4$. Let $p = \frac{1+\epsilon}{d}$. Form G_p by retaining every edge of G independently with probability p . Let L_1 be the giant component of G_p . Then, there exists an absolute constant $c > 0$ such that **whp** for any $S \subseteq L_1$ such that $\frac{\epsilon^2 n}{15} \leq |S| \leq \frac{12\epsilon n}{11}$, we have*

$$\left| \partial_{G_p}(S) \right| \geq \frac{c\epsilon^2 |S|}{\ln\left(\frac{1}{\epsilon}\right)}.$$

Let us mention a few related results. In 2007, Ofek [30] studied the edge-expansion property of G_p when the spectral ratio is nearly optimal and $p \geq \frac{C}{\sqrt{d}}$, obtaining a typical edge-expansion by an $\Omega\left(\frac{1}{\ln n}\right)$ factor. In supercritical $G(n, p)$, Fountoulakis and Reed [15], and independently Benjamini, Kozma, and Wormald [4] showed that asymptotically the edge-expansion of connected subsets of the giant is by an $\Omega(\epsilon)$ factor. Krivelevich [22, 23] showed a method for finding large expanders as subgraphs of general graphs, and in particular the typical existence of a linearly large constant factor expander in supercritical $G(n, p)$. Theorems 1 and 2 here establish **whp** an $\tilde{\Omega}(\epsilon^2)$ factor.

From Theorem 1, we are able to derive that the largest connected component of G_p **whp** contains a large expander as a subgraph. Indeed, a simple corollary of Property 2 of Theorem 1 is the following:

Theorem 3. *Let $\epsilon > 0$ be a small enough constant. Let $G = (V, E)$ be an (n, d, λ) -graph with $\frac{\lambda}{d} \leq \epsilon^4$. Let $p = \frac{1+\epsilon}{d}$. Form G_p by retaining every edge of G independently with probability p . Let L_1 be the giant component of G_p . Then, **whp** there exists $L'_1 \subseteq L_1$ with $|L'_1| \geq \frac{11\epsilon n}{6}$ such that for every $S \subseteq L'_1$ with $|S| \leq \frac{|L'_1|}{2}$,*

$$\left|N_{L'_1}(S)\right| \geq \frac{c\epsilon^2|S|}{\ln^2\left(\frac{1}{\epsilon}\right)}.$$

Remark 1.2. *Recalling that **whp** $|L_1| = (2 + o_\epsilon(1))\epsilon n$, we note that the fraction $\frac{7}{4}$ in Theorem 3 can be replaced by any constant strictly smaller than 2, while adjusting the statements and proofs of the other theorems as well. It is left as such for ease of presentation.*

From the typical existence of a large, relatively good expander as a subgraph, one can obtain several other interesting properties of the giant component, such as the existence **whp** of a large complete minor, a long cycle and cycles of different lengths. We refer the reader to [23] for a survey including many results of this type, and to [19] for an extensive survey on expanders and their applications.

From the expansion properties of the giant component, we will be able to derive an asymptotic bound on its diameter:

Theorem 4. *Let $\epsilon > 0$ be a small enough constant. Let $G = (V, E)$ be an (n, d, λ) -graph with $\frac{\lambda}{d} \leq \epsilon^4$. Let $p = \frac{1+\epsilon}{d}$. Form G_p by retaining every edge of G independently with probability p . Let L_1 be the largest component of G_p . Then, there exists an absolute constant C such that **whp** the diameter of L_1 is at most*

$$C \frac{\ln\left(\frac{1}{\epsilon}\right) \ln n}{\epsilon^2}.$$

For $G(n, p)$ with $p = \frac{1+\epsilon}{n}$, the diameter of the giant component is known to be typically $\Theta\left(\frac{\ln n}{\epsilon}\right)$ (see the results of Fernholz and Ramachandran [13] and of Riordan and Wormald [31]). Theorem 4 is within an $\tilde{\Omega}\left(\frac{1}{\epsilon}\right)$ factor from that.

In [15], Fountoulakis and Reed obtained a good control over the typical edge-expansion of connected subsets of the giant component in $G(n, p)$ and utilised it in order to bound the mixing time of a random walk on the giant component. Adjusting their method to our setting, we obtain:

Theorem 5. *Let $\epsilon > 0$ be a small enough constant. Let $G = (V, E)$ be an (n, d, λ) -graph with $\frac{\lambda}{d} \leq \epsilon^4$. Let $p = \frac{1+\epsilon}{d}$. Form G_p by retaining every edge of G independently with probability p . Let L_1 be the largest component of G_p . Then, there exists an absolute constant C such that **whp** the mixing time of the lazy random walk on L_1 is at most*

$$C \frac{\ln^2 n}{\epsilon^4}.$$

For the supercritical $G(n, p)$, Fountoulakis and Reed [15], and independently Benjamini, Kozma and Wormald [4] established the asymptotic order of magnitude of the mixing time of a random walk on the giant component is asymptotically $\Theta(\log^2 n)$. Thus, the order of magnitude (in n) of our estimate in Theorem 5 is optimal.

Let us briefly remark that the proofs for $G(n, p)$ usually rely on the symmetry of the complete graph, on its density (that is, its degree is roughly its number of vertices), and on delicate arguments regarding its 2-core and structure (see [15], [4], [31]). Hence, these techniques do not carry on to expanders, which could be much sparser, and lack the symmetry of the complete graph. We thus require several different tools, with the key ones appearing in Lemma 2.6 and in the proof of Property 1 of Theorem 1, allowing us to establish the typical expansion properties and derive from them the asymptotic diameter and mixing time of the giant component.

In Section 2 we state and establish several lemmas which we will use throughout the proofs, as well as the key lemma for proving Theorems 1 and 2 (Lemma 2.6), utilising a random variant of the Breadth First Search algorithm. In Section 3 we prove Theorems 1, 2, and 3. Theorem 4 is proven in Section 4, and Theorem 5 is proven in Section 5.

2 Auxiliary Lemmas

We denote by $e_G(A, B)$ the number of edges in G with one endpoint in A and the other endpoint in B . When $A = B$, we write for short $e_G(A) := \frac{e_G(A, A)}{2}$. Note that $e_G(A)$ is the number of edges inside A . When the graph we refer to is obvious, we sometimes omit the subscript.

We begin by stating a couple of lemmas which we will use throughout the proofs. The first one is the famed *expander mixing lemma* due to Alon and Chung [2] (which we state in the same form as Theorem 2.11 of [26]), which gives us control over the edge-distribution of the graph through its spectral ratio:

Lemma 2.1 ([2]). *Let $G = (V, E)$ be an (n, d, λ) -graph. Then, for every two subsets $A, B \subseteq V$,*

$$\left| e_G(A, B) - \frac{d}{n}|A||B| \right| \leq \lambda \sqrt{|A||B| \left(1 - \frac{|A|}{n}\right) \left(1 - \frac{|B|}{n}\right)}.$$

We also require the following claim, which follows from the expander mixing lemma.

Claim 2.2. *Let G be an (n, d, λ) -graph, and let $\delta = \frac{\lambda}{d}$. Let $\alpha, \beta > 0$ and $A \subseteq V$ be such that $|A| \geq \beta n$. Define:*

$$B = \left\{ v \in V \setminus A : d_G(v, A) \leq (1 - \alpha) \frac{|A|d}{n} \right\}.$$

Then $|B| \leq \frac{\delta^2 n}{\alpha^2 \beta}$.

Proof. By the definition of B , we have that $e_G(A, B) \leq (1 - \alpha) \frac{|A||B|d}{n}$. On the other hand, by the expander mixing lemma, we have that:

$$e_G(A, B) \geq \frac{|A||B|d}{n} - \lambda \sqrt{|A||B|}.$$

Combining these together, we obtain:

$$\frac{|A||B|d}{n} - \lambda \sqrt{|A||B|} \leq (1 - \alpha) \frac{|A||B|d}{n}.$$

Since $|A| \geq \beta n$, we thus have that $|B| \leq \frac{\delta^2 n}{\alpha^2 \beta}$, as required. \square

The next lemma bounds the number of subtrees of a given order in a d -regular graph:

Lemma 2.3. *Let $t_k(G)$ denote the number of k -vertex trees contained in d -regular graph G on n vertices. Then,*

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

This follows directly from Lemma 2 of [5].

Let $D_k(G) := \{v \in V(G) : d_G(v) > k\}$. We will use the following lemma, which bounds the typical number of edges incident to high degree vertices in G_p :

Lemma 2.4. *Let $G = (V, E)$ be a d -regular graph on n vertices. Let $\epsilon > 0$ be a small enough constant and let $p \leq \frac{2}{d}$. Form G_p by retaining each edge of G with probability p . Let X be the random variable defined as*

$$X := \sum_{v \in D_{5 \ln(\frac{1}{\epsilon})}(G_p)} d_{G_p}(v).$$

Then, *whp*,

$$X \leq 2n\epsilon^{\ln \ln(\frac{1}{\epsilon})}.$$

Proof. Let $k = n\epsilon^{\ln \ln(\frac{1}{\epsilon})}$. Assume that $X > 2k$. Then,

$$\sum_{v \in D_{5 \ln(\frac{1}{\epsilon})}(G_p)} d_{G_p}(v) = 2e_{G_p}(D_{5 \ln(\frac{1}{\epsilon})}(G_p)) + e_{G_p}(D_{5 \ln(\frac{1}{\epsilon})}(G_p), D_{5 \ln(\frac{1}{\epsilon})}(G_p)^C) > 2k,$$

and in particular,

$$e_{G_p}(D_{5 \ln(\frac{1}{\epsilon})}(G_p)) + e_{G_p}(D_{5 \ln(\frac{1}{\epsilon})}(G_p), D_{5 \ln(\frac{1}{\epsilon})}(G_p)^C) > k.$$

Let S be a minimal subset of $D_{5 \ln(\frac{1}{\epsilon})}(G_p)$, satisfying that $e_{G_p}(S) + e_{G_p}(S, S^C) > k$. This implies that for every vertex $u \in S$, we have that $e_{G_p}(S \setminus \{u\}) + e_{G_p}(S \setminus \{u\}, S^C \cup \{u\}) \leq k$. Now,

$$\sum_{v \in S \setminus \{u\}} d_{G_p}(v) = 2e_{G_p}(S \setminus \{u\}) + e_{G_p}(S \setminus \{u\}, S^C \cup \{u\}) \leq 2k.$$

Since every vertex contributes at least $5 \ln(\frac{1}{\epsilon})$ edges to $\sum_{v \in S \setminus \{u\}} d_{G_p}(v)$, we have that

$$|S| = 1 + |S \setminus \{u\}| \leq 1 + \frac{2k}{5 \ln(\frac{1}{\epsilon})} \leq \frac{k}{2 \ln(\frac{1}{\epsilon})}.$$

We have at most $\binom{n}{|S|} \leq \binom{n}{k/2 \ln(\frac{1}{\epsilon})}$ ways to choose a set of size $|S|$. We then have at most $\binom{|S|d}{k} \leq \binom{kd/2 \ln(\frac{1}{\epsilon})}{k}$ ways to choose k of the edges touching S , and include them in G_p with probability p^k . Therefore, the probability that $X > k$ is at most:

$$\begin{aligned} \binom{n}{\frac{k}{2 \ln(\frac{1}{\epsilon})}} \binom{\frac{kd}{2 \ln(\frac{1}{\epsilon})}}{k} p^k &\leq \left(\left(\frac{2en \ln(\frac{1}{\epsilon})}{n\epsilon^{\ln \ln(\frac{1}{\epsilon})}} \right)^{\frac{1}{2 \ln(\frac{1}{\epsilon})}} \frac{ekd}{2k \ln(\frac{1}{\epsilon})} \cdot \frac{2}{d} \right)^k \\ &\leq \left(\left(\frac{6 \ln(\frac{1}{\epsilon})}{\epsilon^{\ln \ln(\frac{1}{\epsilon})}} \right)^{\frac{1}{2 \ln(\frac{1}{\epsilon})}} \frac{e}{\ln(\frac{1}{\epsilon})} \right)^k \\ &= \left(\left(6 \ln\left(\frac{1}{\epsilon}\right) \right)^{\frac{1}{2 \ln(\frac{1}{\epsilon})}} \ln^{1/2}\left(\frac{1}{\epsilon}\right) \cdot \frac{e}{\ln(\frac{1}{\epsilon})} \right)^k \\ &\leq \left(\frac{e}{\ln^{1/3}\left(\frac{1}{\epsilon}\right)} \right)^k = o(1), \end{aligned}$$

where, since we assume that ϵ is small enough, we have that $\ln^{1/3}\left(\frac{1}{\epsilon}\right) > e$. □

The next lemma bounds the typical number of edges incident to connected subsets:

Lemma 2.5. *Let $G = (V, E)$ be a d -regular graph. Let $p \leq \frac{2}{d}$. Form G_p by retaining each edge of G with probability p . Then, **whp** for all $S \subseteq V(G)$ such that $|S| = k \geq \ln n$ and $G_p[S]$ is connected,*

$$e_{G_p}(S) + e_{G_p}(S, S^C) < 10k.$$

Proof. Since any connected set in G_p has a spanning tree, it is sufficient to show the statement is true for the edges incident to trees of order k in G_p . By Lemma 2.3, there are at most $n(ed)^{k-1}$ trees on k vertices in G . The probability such a tree falls into G_p is p^{k-1} . A set of k vertices has at most kd edges touching it in G . We have at most $\binom{kd}{9k}$ ways to choose the additional $9k$ edges incident to the set, and they are in G_p with probability p^{9k} . As such, by the union bound, the probability of an event violating the statement of the lemma is at most:

$$\begin{aligned} n(ed)^{k-1}p^{k-1}\binom{kd}{9k}p^{9k} &\leq n \cdot (2e)^{k-1} \left(\frac{2e}{9}\right)^{9k} \\ &\leq n \exp\left(-\frac{5k}{2}\right) = o(1/n), \end{aligned}$$

for $k \geq \ln n$. The union bound over the at most n possible values of k completes the proof. \square

The final lemma of this section will be a key tool in proving Theorem 2. In order to prove it, we will make use of the Breadth First Search (BFS) algorithm.

This variation of the BFS constructs G_p while exploring the spanning trees of the connected components. The algorithm maintains three sets of vertices: S , the vertices whose exploration is complete; Q , the vertices currently being explored, kept in a queue; and T , the vertices that have not been explored yet. The algorithm starts with $S = Q = \emptyset$ and $T = V(G)$, and ends when $Q \cup T = \emptyset$. At each step, if Q is non-empty, the algorithm queries T for neighbours of the first vertex in Q (edge by edge), according to the edges of G and an order σ on V (that is, at each step we query a single edge). The algorithm is fed X_i , $0 \leq i \leq \frac{nd}{2}$, i.i.d Bernoulli(p) random variables, each corresponding to a positive (with probability p) or negative (with probability $1 - p$) answer to such a query. Thus, each X_i corresponds to some edge of G (noting that this edge depends on the execution of the algorithm), which is included in G_p if $X_i = 1$ and is excluded otherwise. If Q is non-empty and we discovered a neighbour of $v \in Q$ in T , we place the neighbour as the last vertex in Q and continue. If Q is non-empty and the first vertex in Q has no more queries to ask, then we move the first vertex of Q to S . If $Q = \emptyset$, we move the next vertex from T (according to σ) into Q . Observe that the BFS uncovers a spanning forest of G_p . Note that, formally, in order to obtain a graph distributed according to G_p , after the run of the BFS we query all edges in G left unqueried and include them in G_p according to the remaining random bits X_i — this will not be of any importance to us in our analysis, and thus we restrict our attention to the BFS which uncovers a spanning forest of G_p .

We will make use of the following properties of the BFS:

1. A vertex moves from Q to S only if it has no more neighbours in T . Thus, at any moment t all the edges between S and T have been queried;
2. While Q is non-empty, we are exploring the same connected component;
3. At any moment t , we have that $|S \cup Q| \geq 1 + \sum_{i=1}^t X_i$, since the first vertex moves into Q without a query, and every positive answer to a query corresponds to a vertex that moves from T into Q (and perhaps later on into S); and,
4. At any moment t , we have asked no queries between the vertices of Q and T , except perhaps for the first vertex of Q .

We note that the initial analysis of the BFS will be similar to that of the Depth First Search (DFS) in [27].

Lemma 2.6. *Let $\epsilon > 0$ be a small enough constant and let $\delta \leq \epsilon^4$. Let $G = (V, E)$ be an (n, d, λ) -graph, with $\frac{\lambda}{d} \leq \delta$. Let $p = \frac{1+\epsilon}{d}$. Form G_p by retaining each edge of G independently with probability p . Then, the probability there is a connected component B , $|B| = k$, in G_p with $\frac{\epsilon^2 n}{15} \leq k \leq \frac{11\epsilon n}{10}$ is at most $\exp\left(-\frac{\epsilon^2 k}{220^2}\right)$.*

Proof. We require the following claim, whose proof we defer to the end of this proof:

Claim 2.7. *With probability at least $1 - \exp\left(-\frac{\epsilon^3 n}{201^2}\right)$, there is an interval starting at least as early as $\frac{\epsilon^2 nd}{20}$, where the BFS uncovers a connected component of size larger than $\frac{11\epsilon n}{10}$.*

We now consider the three possible cases: we finished discovering B prior to $\frac{\epsilon^2 nd}{20}$; B is the connected component asserted by Claim 2.7; we began discovering B subsequent to discovery of the connected component described in Claim 2.7.

If we finished discovering B prior to $\frac{\epsilon^2 nd}{20}$, then by properties of the BFS we have that $|B| \leq 1 + \sum_{i=1}^{\frac{\epsilon^2 nd}{20}} X_i$. As such, by a typical Chernoff-type bound (see Appendix A in [3]), we obtain that with probability at least

$$1 - \exp\left(-\frac{\frac{\epsilon^4 n^2}{16^2}}{3 \cdot (1+\epsilon)\epsilon^2 n}\right) \geq 1 - \exp\left(-\frac{\epsilon^2 n}{40}\right),$$

we have that $|B| < \frac{\epsilon^2 n}{15}$.

If B is the connected component described in Claim 2.7, then with probability at least $1 - \exp\left(-\frac{\epsilon^3 n}{201^2}\right)$, we have that $|B| > \frac{11\epsilon n}{10}$.

Suppose towards contradiction that we discovered some connected component B of size at least $\frac{\epsilon^2 n}{15}$, and that we began discovering B subsequent to the discovery of the connected component described in Claim 2.7. By Claim 2.7, with probability at least $1 - \exp\left(-\frac{\epsilon^3 n}{201^2}\right)$ we have discovered a component of size at least $\frac{11\epsilon n}{10}$, which we denote by M . Note that by properties of the BFS, by the time we have finished discovering M all the edges of G between M and T have been queried (with all queries answered in the negative). Therefore, we could have no queries between B and M , and hence we have an interval in the BFS whose length is at most $|B|d - e_G(B, M)$, where we received at least $|B| - 1$ positive answers. By the above bound on $|M|$ and by the expander mixing lemma (Lemma 2.1), with probability at least $1 - \exp\left(-\frac{\epsilon^3 n}{202^2}\right)$ we have that:

$$\begin{aligned} e_G(B, M) &\geq \frac{d}{n}|B||M| - \lambda\sqrt{|B||M|} \\ &\geq \frac{11\epsilon|B|d}{10} - \lambda\sqrt{\frac{11\epsilon n|B|}{10}} \\ &\geq \frac{12\epsilon|B|d}{11}, \end{aligned}$$

where the last inequality is due to $|B| \geq \frac{\epsilon^2 n}{15}$ and $\frac{\lambda}{d} \leq \delta \leq \epsilon^4$. Therefore, with probability at least $1 - \exp\left(-\frac{\epsilon^3 n}{202^2}\right)$, we have an interval whose length is at most

$$|B|d - e_G(B, M) \leq \left(1 - \frac{12\epsilon}{11}\right)|B|d,$$

where we received at least $|B| - 1$ positive answers. Considering the n possible starts of the interval together with a Chernoff-type bound, this happens with probability at most:

$$n \cdot P \left[\text{Bin} \left(\left(1 - \frac{12\epsilon}{11} \right) |B|d, p \right) \geq |B| - 1 \right] \leq \exp \left(-\frac{\epsilon^2 |B|}{50} \right),$$

and thus the probability of an event violating the statement of the Lemma is at most

$$\exp \left(-\frac{\epsilon^2 n}{40} \right) + \exp \left(-\frac{\epsilon^3 n}{201^2} \right) + \exp \left(-\frac{\epsilon^2 k}{50} \right) \leq \exp \left(-\frac{\epsilon^2 k}{220^2} \right),$$

completing the proof.

Proof of Claim 2.7. We begin by showing that with probability at least $1 - \exp \left(-\frac{\epsilon^3 n}{41^2} \right)$, we have that Q does not empty in the interval

$$I = \left[\frac{\epsilon^2 nd}{20}, \frac{3\epsilon nd}{4} \right].$$

Assume Q empties at some moment $t \in I$. Specifically, we let t be the first moment in that interval such that Q is empty.

Note, we may assume that at t we have $|S| \leq \frac{n}{3}$. Otherwise, as the vertices move between the sets one by one, there is some moment $t' \leq t \leq \frac{3\epsilon nd}{4}$ where $|S| = \frac{n}{3}$. By the expander mixing lemma, we then had by the moment t' at least

$$\begin{aligned} e_G(S, S^C) &\geq \frac{d}{n} \frac{2n}{3} \frac{n}{3} - \lambda \sqrt{\frac{2n^2}{9}} \\ &\geq \frac{nd}{9} > \frac{3\epsilon nd}{4} \end{aligned}$$

queries, a contradiction.

Furthermore, at t , we have by properties of the BFS that $|S| \geq \sum_{i=1}^t X_i$. By a Chernoff-type bound, we obtain that with probability at least $1 - \exp \left(-\frac{\epsilon t}{40^2 d} \right) \geq 1 - \exp \left(-\frac{\epsilon^3 n}{41^2} \right)$, we have $|S| \geq \frac{(1+\epsilon)t}{d} - \frac{\epsilon t}{39d}$. As such, by the expander mixing lemma, with the same probability

$$\begin{aligned} e_G(S, S^C) &\geq \frac{d}{n} (n - |S|) |S| - \lambda \sqrt{(n - |S|) |S|} \\ &\geq \left(1 - 2\delta - \frac{(1 + \epsilon - \frac{\epsilon}{39}) t}{dn} \right) \left(1 + \epsilon - \frac{\epsilon}{39} \right) t, \end{aligned}$$

where the second inequality uses the fact that since we assumed that $|S| \leq \frac{n}{3}$, the expression is minimised by the smallest possible value of $|S|$. We note that this accounts for the queries between S and T by t . For any $t \leq \frac{3\epsilon nd}{4}$, we then have that with the same probability:

$$\begin{aligned} t &\geq \left(1 - 2\delta - \frac{(1 + \frac{38\epsilon}{39}) t}{dn} \right) \left(1 + \frac{38\epsilon}{39} \right) t \\ &\geq \left(1 - \frac{4\epsilon}{5} \right) \left(1 + \frac{38\epsilon}{39} \right) t > t, \end{aligned}$$

a contradiction.

We now continue the analysis of the algorithm, albeit in a slightly different manner. By the properties of the BFS, at the moment $t = \frac{3\epsilon nd}{4}$, there have been no queries between the vertices of Q and T (except perhaps for the first vertex Q). We set Q_0 , T_0 and S_0 to be the sets Q

without its first vertex, T and S , respectively, at the moment $\frac{3\epsilon nd}{4}$. For any $i \geq 1$, as long as Q_{i-1} is nonempty, we define Q_i to be the neighbourhood of Q_{i-1} in T , $S_i = S_{i-1} \cup Q_{i-1}$ and $T_i = T_{i-1} \setminus Q_i$. Furthermore, we call each such step an *iteration*, and Q_i, S_i and T_i are the relevant sets at iteration i . We claim that, with probability at least $1 - \exp\left(-\frac{\epsilon^3 n}{50^2}\right)$, there is an iteration where we uncover a connected component whose size is larger than $\frac{11\epsilon n}{10}$, completing the proof.

We begin by considering the set Q_0 . Assume that $|Q_0| \leq \frac{\epsilon^2 n}{9}$. Then, as before, by the properties of the BFS and a Chernoff-type bound, we have with probability at least $1 - \exp\left(-\frac{\epsilon^3 n}{41^2}\right)$ that:

$$|S_0| \geq \frac{3(1+\epsilon)\epsilon n}{4} - \frac{3\epsilon^2 n}{130} - \frac{\epsilon^2 n}{9} \geq \frac{3\epsilon n}{4} + \frac{3\epsilon^2 n}{5}. \quad (2)$$

By the same arguments as before, we may further assume that $|S_0| \leq \frac{n}{3}$. Therefore (with the same probability), by the expander mixing lemma

$$\begin{aligned} e_G(S_0, T_0) &\geq \frac{d}{n}|S_0||T_0| - \lambda\sqrt{|S_0||T_0|} \\ &\geq (1-5\delta) \left(\frac{3\epsilon}{4} + \frac{3\epsilon^2}{5}\right) \left(1 - \frac{3\epsilon}{4} - \epsilon^2\right) nd \\ &\geq \frac{3\epsilon nd}{4} + \frac{\epsilon^2 nd}{80} > \frac{3\epsilon nd}{4}, \end{aligned}$$

where in the second inequality we used the fact that since we assume that $|S_0| \leq \frac{n}{3}$, the expression is minimised by the smallest possible value of $|S_0|$. Thus, with probability at least $1 - \exp\left(-\frac{\epsilon^3 n}{41^2}\right)$, we have that $|Q_0| \geq \frac{\epsilon^2 n}{9}$.

Now, using Claim 2.2 with $A = Q_i$ and $\alpha = \epsilon^2$, we have that the number of vertices which have less than $\frac{(1-\epsilon^2)|Q_i|d}{n}$ neighbours in Q_i is at most $\frac{n}{|Q_i|} \cdot \frac{\delta^2 n}{\epsilon^4} \leq \frac{\epsilon^4 n^2}{|Q_i|}$. Denote the set of these vertices by B_i . As such, every vertex in $T_i \setminus B_i$ has probability at least $1 - (1-p)^{\frac{(1-\epsilon^2)|Q_i|d}{n}} \geq \frac{(1+\epsilon-2\epsilon^2)|Q_i|}{n}$ to be in Q_{i+1} , and these events are independent for each vertex. Hence, at iteration $i+1$, the number of vertices in Q_{i+1} stochastically dominates

$$\text{Bin}\left(|T_i| - |B_i|, \frac{(1+\epsilon-2\epsilon^2)|Q_i|}{n}\right).$$

Since we prefer to express the above in terms of Q_i , observe that:

$$|T_i| - |B_i| \geq n - |S_0| - \sum_{j=1}^i |Q_j| - \frac{\epsilon^4 n^2}{|Q_i|}.$$

Recall that with probability at least $1 - \exp\left(-\frac{\epsilon^3 n}{41^2}\right)$, we have that Q does not empty in the interval I . With probability at least $1 - \exp\left(-\frac{\epsilon^3 n}{41^2}\right)$, we have that $\sum_{j=1}^{\frac{3\epsilon nd}{4}} X_j \leq \frac{31\epsilon n}{40}$. Furthermore, consider the sets S and T at the last moment before $\frac{\epsilon^2 nd}{20} + 1$ where Q was empty. If we had that $|S| \geq \frac{3\epsilon^2 n}{5}$, then by the expander mixing lemma at that moment

$$e_G(S, T) \geq \frac{3\epsilon^2 nd}{5} - 2\lambda\epsilon n > \frac{\epsilon^2 nd}{20},$$

a contradiction. By the first and fourth properties of the BFS, we may thus conclude that with probability at least $1 - \exp\left(-\frac{\epsilon^3 n}{42^2}\right)$, we have that $|S_0| \leq \frac{31\epsilon n}{40} + \frac{3\epsilon^2 n}{5}$, and that at most $\frac{3\epsilon^2 n}{5}$

vertices in S_0 do not belong to the component we are now uncovering. Finally, for the simplicity of computation, we stop the process before $|Q_i| \leq \frac{7\epsilon^2 n}{90}$, and hence we have that $\frac{\epsilon^4 n^2}{|Q_i|} \leq 15\epsilon^2 n$.

Therefore, with probability at least $1 - \exp\left(-\frac{\epsilon^3 n}{45^2}\right)$, we have that:

$$\begin{aligned} |T_i| - |B_i| &\geq n - |S_0| - \sum_{j=1}^i |Q_j| - \frac{\epsilon^4 n^2}{|Q_i|} \\ &\geq \left(1 - \frac{31\epsilon}{40} - 20\epsilon^2 - \frac{\sum_{j=1}^i |Q_j|}{n}\right) n. \end{aligned}$$

Let

$$N_i = \bigcup_{j=1}^i Q_j, \quad q_i = \frac{|Q_i|}{n}, \quad \text{and} \quad n_i = \frac{|N_i|}{n}.$$

Denote by i_0 the first iteration where $n_{i_0} \geq \frac{\frac{\epsilon}{5}}{1+\epsilon}$, and let i_1 be the last iteration before $q_i < \frac{7\epsilon^2}{90}$. By a typical Chernoff-type bound, we have with probability at least

$$1 - \exp\left(-\frac{\epsilon|Q_i|}{46^2}\right) \geq 1 - \exp\left(-\frac{\epsilon^3 n}{200^2}\right),$$

that

$$q_{i+1} \geq \left(1 + \frac{44\epsilon}{45}\right) \left(1 - \frac{31\epsilon}{40} - \sum_{j=1}^i q_j - 20\epsilon^2\right) q_i \geq \left(1 + \frac{\epsilon}{5} - (1+\epsilon) \sum_{j=1}^i q_j\right) q_i. \quad (3)$$

We continue by assuming that (3) holds and $q_0 \geq \frac{\epsilon^2}{9}$. Note that for any $i < i_0$, we have by (3) that $q_{i+1} \geq q_i \geq \frac{\epsilon^2}{9}$. Assume that $n_{i_1} < \frac{7\epsilon}{20}$. Then, for any $i_0 < i < i_1$, we have by (3) that

$$\begin{aligned} q_{i+1} &\geq \left(1 + \frac{\epsilon}{5} - (1+\epsilon) \frac{7\epsilon}{20}\right) q_i \\ &> \left(1 - \frac{\epsilon}{5}\right) q_i. \end{aligned}$$

Thus,

$$\begin{aligned} \sum_{i=i_0+1}^{i_1} q_i &\geq \frac{\frac{\epsilon^2}{9} - q_{i_1}}{\frac{\epsilon}{5}} \\ &\geq \frac{5\epsilon}{9} - \frac{7\epsilon}{18(1-\frac{\epsilon}{5})} \\ &> \frac{99\epsilon}{600}. \end{aligned}$$

However, we then have that

$$n_{i_1} \geq \frac{\frac{\epsilon}{5}}{1+\epsilon} + \frac{99\epsilon}{600} > \frac{7\epsilon}{20},$$

which contradicts our assumption on n_{i_1} . Hence, we may conclude that with probability at least $1 - \exp\left(-\frac{\epsilon^3 n}{200^2}\right)$, $|N_{i_1}| = n \cdot n_{i_1} \geq \frac{7\epsilon n}{20}$, and thus with probability at least $1 - \exp\left(-\frac{\epsilon^3 n}{201^2}\right)$, we have discovered a component of size at least

$$\frac{7\epsilon n}{20} + \frac{3\epsilon n}{4} = \frac{11\epsilon n}{10},$$

where we used both the above bound on $|N_{i_1}|$, our lower bound on $|S_0|$ as given in (2), and that there were at most $\frac{3\epsilon^2 n}{5}$ vertices in S_0 which do not belong to this component. \square

\square

3 Expansion and Expanders

In this section, we will prove Theorems 1, 2, and 3. The proofs of Theorems 1 and 2 are somewhat interwoven.

We begin with the proof of the first property of Theorem 1, which also implies the same asymptotic edge-expansion factor for connected subsets. Utilising Lemma 2.6 we then show that, in fact, the same edge-expansion factor holds **whp** for connected subsets of size up to $\frac{12\epsilon n}{11}$ (Lemma 3.1). Having the asymptotic bound on the edge-expansion of connected sets of any size between $\frac{16\ln n}{\epsilon^2}$ and $\frac{12\epsilon n}{11}$, we will be able to derive Theorem 2. This, together with an asymptotic bound on the sum of degrees of high-degree vertices (Lemma 2.4), will imply the second property of Theorem 1. We conclude the section with the proof of Theorem 3.

For the proof of the first property of Theorem 1, we will directly bound the probability of the event violating the statement of the theorem.

Proof of Property 1 of Theorem 1. Consider the event \mathcal{A}_k ,

$$\mathcal{A}_k = \left\{ \exists S \subseteq V(G), |S| = k, G_p[S] \text{ is connected} \ \& \ |N_{G_p}(S)| < \frac{\epsilon^2 k}{40 \ln\left(\frac{1}{\epsilon}\right)} \right\}.$$

We will show that for $\frac{16\ln n}{\epsilon^2} \leq k \leq \frac{\epsilon^2 n}{15}$, $P[\mathcal{A}_k] = o\left(\frac{1}{n}\right)$, and therefore by the union bound over the at most n possible values of k , the probability of an event violating the statement of the theorem is $o(1)$.

Let S , $|S| = k$, be a connected set in G_p . Since it is connected, it must have a spanning tree. By Lemma 2.3, we have $n(ed)^{k-1}$ ways to choose a tree of size k , and we include its edges in G_p with probability p^{k-1} . Now, consider the auxiliary random bipartite graph $\Gamma(S, p)$, whose one side is S , the other side is $N_G(S)$, and we include every edge of G between S and $N_G(S)$ in $\Gamma(S, p)$ with probability p . Then, we have that $|N_{G_p}(S)| \geq \nu(\Gamma(S, p))$, where $\nu(H)$ is the matching number of H . Thus, it suffices to bound the probability that a maximum matching in $\Gamma(S, p)$ is smaller than $\frac{\epsilon^2 k}{40 \ln\left(\frac{1}{\epsilon}\right)}$.

Let us first bound the probability that $\nu(\Gamma(S, p)) = i$. This is at most the probability that $\Gamma(S, p)$ has a maximal by inclusion matching of size i . We have at most $\binom{e_G(S, N_G(S))}{i} \leq \binom{kd}{i}$ ways to choose a matching M of size i . We then need to include the edges of the matching, which happens with probability p^i . Due to the maximality of M , every edge of G between S and $N_G(S)$ disjoint from M is not in $\Gamma(S, p)$. Thus, we have at least $e_G(S, S^C) - 2id$ edges that do not fall into $\Gamma(S, p)$. As in our previous arguments, by the expander mixing lemma:

$$e_G(S, S^C) \geq \left(1 - \frac{\epsilon^2}{14}\right) dk,$$

using our assumptions on δ and that $k \leq \frac{\epsilon^2 n}{15}$. Thus,

$$P[\nu(\Gamma(S, p)) = i] \leq \binom{kd}{i} p^i (1-p)^{\left(1 - \frac{\epsilon^2}{14}\right) kd - 2id}.$$

Therefore, we obtain that

$$\begin{aligned}
P[\mathcal{A}_k] &\leq n(ed)^{k-1}p^k \sum_{i=0}^{\frac{\epsilon^2 k}{40 \ln(\frac{1}{\epsilon})}} \binom{kd}{i} p^i (1-p)^{(1-\frac{\epsilon^2}{14})kd-2id} \\
&\leq n \left((1+\epsilon) \exp \left(1 - (1+\epsilon) \left(1 - \frac{\epsilon^2}{14} \right) \right) \right)^k \sum_{i=0}^{\frac{\epsilon^2 k}{40 \ln(\frac{1}{\epsilon})}} \binom{kd}{i} p^i (1-p)^{-2id} \\
&\leq n \left((1+\epsilon) \exp \left(1 - (1+\epsilon) \left(1 - \frac{\epsilon^2}{14} \right) \right) \right)^k \left(1 + \sum_{i=1}^{\frac{\epsilon^2 k}{40 \ln(\frac{1}{\epsilon})}} \left(\frac{e^4 k}{i} \right)^i \right).
\end{aligned}$$

Observe that the ratio of consecutive terms in the above sum is:

$$\frac{\left(\frac{e^4 k}{i} \right)^i}{\left(\frac{e^4 k}{i+1} \right)^{i+1}} = \frac{(i+1)^{i+1}}{e^4 k^i} \leq \frac{1}{2},$$

and thus the sum is at most twice the final term. Hence,

$$\begin{aligned}
P[\mathcal{A}_k] &\leq n \left((1+\epsilon) \exp \left(1 - (1+\epsilon) \left(1 - \frac{\epsilon^2}{14} \right) \right) \right)^k \left(1 + 2 \left(\frac{e^4 40 \ln(\frac{1}{\epsilon})}{\epsilon^2} \right)^{\frac{\epsilon^2 k}{40 \ln(\frac{1}{\epsilon})}} \right) \\
&\leq 3n \left((1+\epsilon) \exp \left(-\epsilon + \frac{\epsilon^2}{14} + \frac{\epsilon^2}{40 \ln(\frac{1}{\epsilon})} \ln \left(\frac{41e^4 \ln(\frac{1}{\epsilon})}{\epsilon^2} \right) \right) \right)^k \\
&\leq 3n \left((1+\epsilon) \exp \left(-\epsilon + \frac{\epsilon^2}{4} \right) \right)^k \\
&\leq 3n \exp \left(-\frac{\epsilon^2 k}{8} \right) = o \left(\frac{1}{n} \right).
\end{aligned}$$

Note that the last inequality follows since $1+x \leq \exp\left(x - \frac{3x^2}{8}\right)$ for x small enough. Indeed, both functions attain the value of 1 when $x=0$, and the derivative of the latter is strictly larger for small $x \geq 0$. The equality holds since we assumed $k \geq \frac{16 \ln n}{\epsilon^2}$. Union bound over the at most n different values of k completes the proof. \square

We now turn to show that for edge-expansion, the same expansion factor holds **whp** for *connected* sets up to size $\frac{12\epsilon n}{11}$. For subsets larger than $\frac{\epsilon^2 n}{15}$, the arguments above do not suffice, as the expander mixing lemma does not yield a strong enough bound on the number of edges in the boundary. We therefore need to utilise Lemma 2.6, together with arguments inspired by those used for separators in [24].

Lemma 3.1. *Let $\epsilon > 0$ be a small enough constant and let $\delta > 0$ be such that $\delta \leq \epsilon^4$. Let $G = (V, E)$ be an (n, d, λ) -graph with $\frac{\lambda}{d} \leq \delta$. Let $p = \frac{1+\epsilon}{d}$. Form G_p by retaining every edge of G independently with probability p . Let L_1 be the giant component of G_p . Then, there exists an absolute constant $c > 0$ such that **whp** for any $S \subseteq L_1$ such that $\frac{16 \ln n}{\epsilon^2} \leq |S| \leq \frac{12\epsilon n}{11}$ and $G_p[S]$ is connected, we have*

$$\left| \partial_{G_p}(S) \right| \geq \frac{c\epsilon^2 |S|}{\ln(\frac{1}{\epsilon})}.$$

Proof. For $\frac{16 \ln n}{\epsilon^2} \leq |S| \leq \frac{\epsilon^2 n}{15}$, this follows immediately from the first property of Theorem 1. We now assume that $\frac{\epsilon^2 n}{15} \leq |S| \leq \frac{12 \epsilon n}{11}$. Let $p_1 = \frac{1+\epsilon-\epsilon^2/50}{d}$, and let $\rho = \frac{p_1}{p}$. Let G_{p_1} be the subgraph obtained by retaining each edge of G with probability p_1 . Observe that G_{p_1} can also be obtained by drawing a random graph G_p , and then by retaining each edge with probability ρ .

Let \mathcal{A}_k be the following event addressing G_p :

$$\mathcal{A}_k = \left\{ \exists S \subseteq V(G), |S| = k, S \text{ is connected in } G_p \ \& \ \left| \partial_{G_p}(S) \right| < \frac{\epsilon^2 k}{400^2 \ln\left(\frac{1}{\epsilon}\right)} \right\}.$$

Let \mathcal{B}_k be the following event addressing G_{p_1} :

$$\mathcal{B}_k = \{ \exists S \subseteq V(G), |S| = k \ \& \ S \text{ is a connected component in } G_{p_1} \}.$$

Note that $\rho = 1 - \frac{\epsilon^2}{50} + O(\epsilon^3)$. Thus, going from G_p to G_{p_1} by retaining each edge of G_p with probability ρ , given \mathcal{A}_k , the probability that S is separated from L_1 and remains connected is at least:

$$\begin{aligned} (1 - \rho)^{\frac{\epsilon^2 k}{400^2 \ln\left(\frac{1}{\epsilon}\right)}} \rho^{k-1} &\geq \exp\left(-\left(\frac{\epsilon^2}{40} + \ln\left(\frac{1}{40\epsilon^2}\right) \frac{\epsilon^2}{400^2 \ln\left(\frac{1}{\epsilon}\right)}\right) k\right) \\ &\geq \exp\left(-\frac{\epsilon^2 k}{300^2}\right). \end{aligned}$$

Hence,

$$P[\mathcal{B}_k | \mathcal{A}_k] \geq \exp\left(-\frac{\epsilon^2 k}{300^2}\right).$$

Since p_1 can be rewritten as $p_1 = \frac{1+\epsilon'}{d}$ with $\epsilon' = \epsilon - \frac{\epsilon^2}{50}$, we have by Lemma 2.6 that for $\frac{\epsilon^2 n}{15} \leq k \leq \frac{12 \epsilon n}{11}$,

$$P[\mathcal{B}_k] \leq \exp\left(-\frac{\epsilon'^2 k}{220^2}\right) \leq \exp\left(-\frac{\epsilon^2 k}{250^2}\right).$$

Thus,

$$P[\mathcal{A}_k] \leq \exp\left(-\frac{\epsilon^2 k}{250^2} + \frac{\epsilon^2 k}{300^2}\right) = o\left(\frac{1}{n}\right).$$

Union bound over the at most n possible values of k completes the proof of the lemma, with $c = \frac{1}{400^2}$. \square

Before proving Theorem 2, we require the following lemma as well. Given a family of disjoint subsets of vertices, the volume of the family is the total number of vertices in the subsets.

Lemma 3.2. *Let $\epsilon > 0$ be a small enough constant and let $\delta \leq \epsilon^4$. Let $G = (V, E)$ be an (n, d, λ) -graph, with $\frac{\lambda}{d} \leq \delta$. Let $p = \frac{1+\epsilon}{d}$. Form G_p by retaining each edge of G independently with probability p . Let $c \leq \frac{1}{70}$ be a positive constant independent of ϵ . Define the random variable X to be the maximal volume of a family of vertex disjoint subsets S , such that $\frac{\ln\left(\frac{1}{\epsilon}\right)}{c\epsilon^2} \leq |S| \leq \frac{16 \ln n}{\epsilon^2}$, $G_p[S]$ is connected and $|\partial_{G_p}(S)| < \frac{c\epsilon^2}{\ln\left(\frac{1}{\epsilon}\right)}$. Then **whp** $X \leq 2\epsilon^3 n$.*

Proof. We begin with an upper bound on $\mathbb{E}X$. Clearly, evaluating the expected total volume of all trees in G_p of order $\frac{\ln(\frac{1}{\epsilon})}{c\epsilon^2} \leq k \leq \frac{16\ln n}{\epsilon^2}$ whose edge boundary in G_p is at most $\frac{c\epsilon^2 k}{\ln(\frac{1}{\epsilon})}$, provides an upper bound.

By Lemma 2.3, we have $n(ed)^{k-1}$ ways to choose the tree. We include its edges in G_p with probability p^{k-1} . Denote by m the number of edges in the boundary that fall into G_p . Then, we have $\binom{kd}{m}$ ways to choose the edges in the boundary that fall into G_p , and this happens with probability p^m . Observe that

$$\binom{kd}{m} p^m \leq \left(\frac{ek(1+\epsilon)}{m} \right)^m \leq \left(\frac{ek(1+\epsilon)}{\frac{c\epsilon^2 k}{\ln(\frac{1}{\epsilon})}} \right)^{\frac{c\epsilon^2 k}{\ln(\frac{1}{\epsilon})}},$$

since $(a/x)^x$ is strictly increasing for $0 < x < a/e$, and we assumed that $m \leq \frac{c\epsilon^2 k}{\ln(\frac{1}{\epsilon})}$. Furthermore, we have that all but m of the edges in $E_G(S, S^C)$ do not fall into G_p . By the expander mixing lemma, we obtain in a manner similar to the previous arguments that:

$$\begin{aligned} e_G(S, S^C) - m &\geq \left(1 - \frac{k}{n} - 2\delta\right) kd - m \\ &\geq (1 - \epsilon^3)kd, \end{aligned}$$

where we used our assumptions on k and δ and m . Finally, we consider all the at most k possible values of m . Hence, performing calculations similar to those we employed in proving Property 1 of Theorem 1, we obtain:

$$\begin{aligned} \mathbb{E}X &\leq \sum_{k=\frac{\ln(\frac{1}{\epsilon})}{c\epsilon^2}}^{\frac{16\ln n}{\epsilon^2}} k \cdot n(ed)^{k-1} p^{k-1} \cdot k(1-p)^{(1-\epsilon^3)kd} \left(\frac{ek(1+\epsilon)}{\frac{c\epsilon^2 k}{\ln(\frac{1}{\epsilon})}} \right)^{\frac{c\epsilon^2 k}{\ln(\frac{1}{\epsilon})}} \\ &\leq n \sum_{k=\frac{\ln(\frac{1}{\epsilon})}{c\epsilon^2}}^{\frac{16\ln n}{\epsilon^2}} k^2 \left[e(1+\epsilon) \exp(-(1+\epsilon)(1-\epsilon^3)) \left(\frac{e(1+\epsilon) \ln(\frac{1}{\epsilon})}{c\epsilon^2} \right)^{\frac{c\epsilon^2}{\ln(\frac{1}{\epsilon})}} \right]^k \\ &\leq n \sum_{k=\frac{\ln(\frac{1}{\epsilon})}{c\epsilon^2}}^{\frac{16\ln n}{\epsilon^2}} k^2 \exp\left(-\frac{\epsilon^2 k}{10}\right) \\ &\leq n \cdot 5 \frac{\ln^2(\frac{1}{\epsilon})}{c^2 \epsilon^4} \exp\left(-\frac{\epsilon^2}{10} \cdot \frac{\ln(\frac{1}{\epsilon})}{c\epsilon^2}\right) \\ &\leq n \cdot \frac{5}{c^2} \epsilon^{\frac{1}{10c} - 5} \leq \epsilon^3 n, \end{aligned}$$

assuming that $c \leq \frac{1}{90}$.

We now proceed to show that X is tightly concentrated. Indeed, consider the edge exposure martingale for G_p . Changing one edge in G_p can add or delete at most two sets to the family in question (recalling that it is a family of vertex disjoint subsets). Therefore, at each step of the exposure, we change the size of X by at most $\frac{32\ln n}{\epsilon^2}$. Thus, by a variant of the Azuma-Hoeffding inequality (see, for example, Theorem 3.9 of [29]), we have that:

$$\begin{aligned} P[X \geq 2\epsilon^3 n] &\leq P[X \geq \mathbb{E}X + \epsilon^3 n] \\ &\leq \exp\left(-\frac{\epsilon^6 n^2}{nd \cdot \frac{1+\epsilon}{d} \left(\frac{32\ln n}{\epsilon^2}\right)^2 + 2 \cdot \frac{32\ln n}{\epsilon^2} \cdot \epsilon^3 n}\right) = o\left(\frac{1}{n}\right). \end{aligned}$$

□

We are now ready to prove Theorem 2:

Proof of Theorem 2. Let $c_{(3.1)}$ be the constant guaranteed by Lemma 3.1. Consider the connected components of $G_p[S]$. By Lemma 3.1, **whp** every component of $G_p[S]$ whose size is $k \geq \frac{16 \ln n}{\epsilon^2}$ contributes at least $\frac{c_{(3.1)} \epsilon^2 k}{\ln(\frac{1}{\epsilon})}$ edges to $\partial_{G_p}(S)$, since none of these edges can be in S (otherwise, it would not be a component of $G_p[S]$). Furthermore, since $S \subseteq L_1$ and L_1 is connected, every connected component of $G_p[S]$ contributes at least one edge to $\partial_{G_p}(S)$. Therefore, any component of $G_p[S]$ whose size is at most $k \leq \frac{\ln(\frac{1}{\epsilon})}{c_{(3.1)} \epsilon^2}$ contributes at least $\frac{c_{(3.1)} \epsilon^2 k}{\ln(\frac{1}{\epsilon})}$ edges to $\partial_{G_p}(S)$ as well. We are thus left with connected components of $G_p[S]$, whose size is between $\frac{\ln(\frac{1}{\epsilon})}{c_{(3.1)} \epsilon^2}$ and $\frac{16 \ln n}{\epsilon^2}$.

For any $v \in S$, we define $C_v[S]$ to be the connected component of $G_p[S]$ which includes v . Let $B \subseteq S$ be the following set:

$$B = \left\{ v \in S : |C_v[S]| = k, \frac{\ln(\frac{1}{\epsilon})}{c_{(3.1)} \epsilon^2} \leq k \leq \frac{16 \ln n}{\epsilon^2}, \left| \partial(C_v[S]) \right| < \frac{c_{(3.1)} \epsilon^2 k}{\ln(\frac{1}{\epsilon})} \right\}.$$

By Lemma 3.2, we have that **whp** $|B| \leq 2\epsilon^3 n$.

We can thus conclude that **whp** all but at most $2\epsilon^3 n$ of the vertices of S belong to connected components $C_v[S]$ of $G_p[S]$ such that $\left| \partial_{G_p}(C_v[S]) \right| \geq \frac{c_{(3.1)} \epsilon^2 |C_v[S]|}{\ln(\frac{1}{\epsilon})}$. Hence, for any $S \subseteq L_1$ such that $\frac{\epsilon^2 n}{15} \leq |S| \leq \frac{12\epsilon n}{11}$, we have that **whp**:

$$\left| \partial_{G_p}(S) \right| \geq \frac{c_{(3.1)} \epsilon^2 (|S| - 2\epsilon^3 n)}{\ln(\frac{1}{\epsilon})} \geq \frac{c_{(3.1)} \epsilon^2 |S|}{2 \ln(\frac{1}{\epsilon})},$$

completing the proof of the theorem with $c = c_{(3.1)}/2$. □

The second part of Theorem 1, linear sized (not necessarily connected) sets, will build upon Theorem 2:

Proof of Theorem 1: Property 2. By Theorem 2, for any subset $S \subseteq L_1$ such that $\frac{\epsilon^2 n}{15} \leq |S| \leq \frac{12\epsilon n}{11}$, we have that **whp**

$$\left| \partial_{G_p}(S) \right| \geq \frac{c\epsilon^2 |S|}{\ln(\frac{1}{\epsilon})}.$$

Let $D_t(N_{G_p}(S)) := \{v \in N_{G_p}(S) : d_{G_p}(v) > t\}$. By Lemma 2.4, **whp**,

$$\sum_{v \in D_{5 \ln(\frac{1}{\epsilon})}(N_{G_p}(S))} d_{G_p}(v) \leq 2n\epsilon^{\ln \ln(\frac{1}{\epsilon})}.$$

Thus, for any subset $S \subseteq L_1$ of relevant size, we have that **whp**

$$|N_{G_p}(S)| \geq \frac{\left| \partial_{G_p}(S) \right| - 2n\epsilon^{\ln \ln(\frac{1}{\epsilon})}}{5 \ln(\frac{1}{\epsilon})} \geq \frac{c\epsilon^2 |S|}{6 \ln^2(\frac{1}{\epsilon})},$$

completing the proof. □

We conclude this section with the proof of Theorem 3. The proof uses some ideas present in [23] and Theorem 1.4 of [11].

Proof of Theorem 3. Let $M \subseteq L_1$ be a maximal set such that $|M| < \frac{\epsilon^2 n}{15}$ and $|N_{G_p}(M)| < \frac{\epsilon^2 |M|}{c \ln^2(\frac{1}{\epsilon})}$. Let $L'_1 = L_1 \setminus M$. Assume there is some subset $B \subseteq L'_1$ with $|B| \leq \frac{|V(L'_1)|}{2}$ and $|N_{L'_1}(B)| < \frac{\epsilon^2 |B|}{c \ln^2(\frac{1}{\epsilon})}$. Then,

$$|N_{G_p}(M \cup B)| < \frac{\epsilon^2 |M|}{c \ln^2(\frac{1}{\epsilon})} + \frac{\epsilon^2 |B|}{c \ln^2(\frac{1}{\epsilon})} = \frac{\epsilon^2 |M \cup B|}{c \ln^2(\frac{1}{\epsilon})}.$$

Therefore, due to the maximality of M , we obtain $|M \cup B| \geq \frac{\epsilon^2 n}{15}$. However, by Theorem 1, **whp** every subset $S \subseteq L_1$ with $\frac{\epsilon^2 n}{15} \leq |S| \leq \frac{12\epsilon n}{11}$ has $|N_{G_p}(S)| \geq \frac{\epsilon^2 |S|}{c \ln^2(\frac{1}{\epsilon})}$. Hence, $|M \cup B| > \frac{12\epsilon n}{11}$. On the other hand, by our choice of B and M , we have that:

$$\begin{aligned} |M \cup B| &\leq |M| + \frac{|V(L_1)| - |M|}{2} \\ &= \frac{|V(L_1)| + |M|}{2} \\ &\leq \frac{|V(L_1)|}{2} + \frac{\epsilon n}{12}. \end{aligned}$$

By Theorem 1.1, **whp** $|V(L_1)| \leq 2\epsilon n$. Therefore, $|M \cup B| \leq \frac{13\epsilon n}{12} < \frac{12\epsilon n}{11}$, a contradiction. Thus, **whp** L'_1 has the desired expansion property, and by Theorem 1.1 **whp** $|L'_1| = |L_1| - |M| \geq \frac{23\epsilon n}{12} - \frac{\epsilon^2 n}{15} \geq \frac{11\epsilon n}{6}$, concluding the proof. \square

4 The Diameter of the Giant

Equipped with the expansion properties of connected subsets, we are now able to obtain a bound on the diameter of the largest component:

Proof of Theorem 4. Let $B(v, r)$ denote the ball of radius r around v in G_p . By Theorem 1.1, **whp** we have that $|L_1| < 2\epsilon n$. Hence, if we can show that for any $v \in V(L_1)$,

$$\left| B \left(v, C \frac{\ln(\frac{1}{\epsilon}) \ln n}{\epsilon^2} \right) \right| \geq \epsilon n,$$

for some large enough constant $C > 0$, then every two balls of such radius in L_1 intersect, and hence the diameter of the giant component of G_p is at most

$$2C \frac{\ln(\frac{1}{\epsilon}) \ln n}{\epsilon^2}.$$

By Theorem 1, for $\frac{16 \ln n}{\epsilon^2} \leq |B(v, r)| \leq \frac{\epsilon^2 n}{15}$, we have **whp** that $|N_{G_p}(B(v, r))| \geq \frac{c\epsilon^2 |B(v, r)|}{\ln(\frac{1}{\epsilon})}$.

Thus, we have for $r \geq \frac{16 \ln n}{\epsilon^2}$ that **whp**:

$$|B(v, r+1)| \geq \min \left\{ \frac{\epsilon^2 n}{15}, \left(1 + \frac{c\epsilon^2}{\ln(\frac{1}{\epsilon})} \right) |B(v, r)| \right\}.$$

Hence,

$$\left| B \left(v, \left(\frac{16}{\epsilon^2} + \frac{1}{\ln \left(1 + \frac{c\epsilon^2}{\ln(1/\epsilon)} \right)} \right) \ln n \right) \right| \geq \frac{\epsilon^2 n}{15}.$$

Set $r_1 = \left(\frac{16}{\epsilon^2} + \frac{1}{\ln\left(1 + \frac{c\epsilon^2}{\ln(1/\epsilon)}\right)} \right) \ln n$. By Theorem 1, for any subset $S \subseteq L_1$ with $\frac{\epsilon^2 n}{15} \leq |S| \leq \frac{12\epsilon n}{11}$, we have that **whp** $|N_{G_p}(S)| \geq \frac{c\epsilon^2 |B(v,r)|}{\ln^2(\frac{1}{\epsilon})}$. Therefore, for any $r \geq r_1$, we have that **whp**:

$$|B(v, r+1)| \geq \min \left\{ \frac{12\epsilon n}{11}, \left(1 + \frac{c\epsilon^2}{\ln^2(\frac{1}{\epsilon})} \right) |B(v, r)| \right\},$$

and thus for some constant $K := K(\epsilon)$ large enough, **whp**

$$\left| B(v, r_1 + K) \right| \geq \frac{12\epsilon n}{11} > \epsilon n.$$

Thus, **whp**, the diameter of L_1 is at most

$$\begin{aligned} 2(r_1 + K) &= 2 \left(\frac{16}{\epsilon^2} + \frac{1}{\ln\left(1 + \frac{c\epsilon^2}{\ln(1/\epsilon)}\right)} \right) \ln n + 2K \\ &\leq C' \frac{\ln(\frac{1}{\epsilon}) \ln n}{\epsilon^2}, \end{aligned}$$

for large enough absolute constant C' . □

5 Mixing Time of the Lazy Random Walk

In this section we prove Theorem 5.

We start with some definitions and brief background (see [28] for extensive background on Markov chains and mixing time). Given a graph G , the *lazy simple random walk* on G is a Markov chain starting at a vertex v_0 chosen according to some distribution σ , such that for any vertex $v \in V(G)$ the walk stays at v with probability $\frac{1}{2}$, and otherwise moves to a uniformly chosen random neighbour u of v . Hence, the transition probability from v to u satisfies $P(v \rightarrow u) = \frac{1}{2d(v)}$. For G connected, this Markov chain is irreducible and ergodic and as such has a limit distribution, which we call the *stationary distribution* π , given by $\pi(v) = \frac{d(v)}{2e(G)}$ for any $v \in V(G)$ (see [28]). We are interested in estimating how quickly this Markov chain converges to its limit distribution. For that, recall that the *total variation distance* d_{TV} between two distributions p_1 and p_2 on $V(G)$ is defined by:

$$d_{TV}(p_1, p_2) := \max_{A \subseteq V(G)} \left| p_1(A) - p_2(A) \right|.$$

Let $P^t(v, \cdot)$ denote the distribution on $V(G)$ given by starting the lazy random walk at $v \in V(G)$ and running for t steps. Letting

$$d(t) := \max_{v \in V(G)} d_{TV}(P^t(v, \cdot), \pi),$$

the *mixing time* of the lazy random walk is then defined by:

$$t_{mix} := \min \left\{ t : d(t) \leq \frac{1}{4} \right\}.$$

Now, for any $S \subseteq V(G)$, let

$$\pi(S) := \sum_{v \in S} \pi(v) = \frac{2e(S) + e(S, S^C)}{2e(G)}.$$

We further define:

$$Q(S) := \sum_{v \in S, u \in S^C} \pi(v)P(v \rightarrow u) = \frac{e(S, S^C)}{4e(G)}.$$

The *conductance* $\Phi(S)$ of S is then given by:

$$\Phi(S) := \frac{Q(S)}{\pi(S)\pi(S^C)} = \frac{e(S, S^C)}{2(2e(S) + e(S, S^C))\pi(S^C)},$$

and we note that since $Q(S) = Q(S^C)$, we have that $\Phi(S) = \Phi(S^C)$. Let $\pi_{\min} = \min_{v \in V(G)} \pi(v)$. For $x > \pi_{\min}$, we define:

$$\Phi(x) := \min \{ \Phi(S) : S \subseteq V(G), x/2 \leq \pi(S) \leq x, S \text{ is connected in } G \},$$

if there is no such subset S , we set $\Phi(x) = 1$. The following theorem due to Fountoulakis and Reed [14] bounds the mixing time through the conductance of connected sets:

Theorem 5.1 (Theorem 1 of [14]). *There exists an absolute constant K such that*

$$t_{mix} \leq K \sum_{j=1}^{\log_2 \pi_{\min}^{-1}} \Phi^{-2}(2^{-j}).$$

Throughout the rest of this section, we consider the mixing time of the lazy random walk on the giant component. Formally, let $\epsilon > 0$ be a small enough constant and let $\delta > 0$ be such that $\delta < \epsilon^4$. Let G be an (n, d, λ) -graph with $\frac{\lambda}{d} \leq \delta$. Let $p = \frac{1+\epsilon}{d}$. Let G_p be the graph obtained by retaining each edge of G independently with probability p . By Theorem 1.1, **whp** there is a unique giant component in G_p which we denote by L_1 . Below, $e(S)$ will stand for $e_{G_p}(S)$ and $e(S, S^C)$ will stand for $|\partial_{G_p}(S)|$. Equipped with these notation, we now establish a couple of lemmas.

Lemma 5.2. *Whp for any $S \subseteq L_1$ such that S is connected in G_p and $e(S) + e(S, S^C) \geq \frac{160 \ln n}{\epsilon^2}$, we have that $|S| \geq \frac{16 \ln n}{\epsilon^2}$.*

Proof. We prove the contrapositive. Since $S \subseteq L_1$, by Theorem 1.1 **whp** S is contained in some connected subset $S' \subseteq L_1$ with $|S'| = \frac{16 \ln n}{\epsilon^2}$. Observing that $e(S) + e(S, S^C)$ increases when moving from S to $S' \supseteq S$, we have by Lemma 2.5 that **whp**

$$e(S) + e(S, S^C) \leq e(S') + e(S', S'^C) \leq \frac{160 \ln n}{\epsilon^2}.$$

□

Using Lemma 5.2, we are now able to bound the conductance of relevant connected sets. We will make use of the notion of *excess* of a graph. Recall that the excess of a connected graph $G = (V, E)$ is defined as $|V| - |E| + 1$.

Lemma 5.3. *There exist positive constants c and C such that **whp**, for every $S \subseteq L_1$ with S connected in G_p and $\frac{C \ln n}{\epsilon^3 n} \leq \pi(S) \leq \frac{1}{2}$,*

$$\Phi(S) \geq c \frac{\epsilon^2}{\ln\left(\frac{1}{\epsilon}\right)}.$$

Proof. It can be easily read from the proof of Theorem 2 in [17] (in particular, from the asymptotic number of edges in the giant which is estimated therein) that **whp** the excess of L_1 is at most $\epsilon^2 n$. Furthermore, S is connected and thus $e(S) \geq |S| - 1$. Since $\pi(S) = \frac{2e(S) + e(S, S^C)}{2e(L_1)} \leq \frac{1}{2}$, we obtain that **whp**,

$$|S| \leq 1 + e(S) \leq 1 + \frac{e(L_1) - e(S, S^C)}{2} \leq 1 + \frac{|L_1| + \epsilon^2 n}{2} \leq \frac{12\epsilon n}{11},$$

where the last inequality holds **whp** by Theorem 1.1.

On the other hand, L_1 is connected, and hence $e(L_1) \geq |L_1| - 1$. Since $\pi(S) \geq \frac{C \ln n}{\epsilon^3 n}$, we obtain that **whp**

$$e(S) + e(S, S^C) \geq \frac{2e(S) + e(S, S^C)}{2} \geq \frac{C \ln n \cdot e(L_1)}{\epsilon^3 n} \geq \frac{C \ln n}{\epsilon^2},$$

where the last inequality once again holds **whp** by Theorem 1.1 and $e(L_1) \geq |L_1| - 1$. Hence, choosing $C = 160$, we obtain by Lemma 5.2 that **whp** $|S| \geq \frac{16 \ln n}{\epsilon^2}$.

All in all, we have that **whp** $\frac{16 \ln n}{\epsilon^2} \leq |S| \leq \frac{12\epsilon n}{11}$. Thus, by Theorem 1 and Theorem 2, **whp** $e(S, S^C) \geq \frac{c' \epsilon^2 |S|}{\ln(\frac{1}{\epsilon})}$, and by Lemma 2.5 **whp** $2e(S) + e(S, S^C) \leq 2(e(S) + e(S, S^C)) \leq 20|S|$. Therefore, **whp**

$$\begin{aligned} \Phi(S) &= \frac{e(S, S^C)}{2(2e(S) + e(S, S^C)) \pi(S^C)} \\ &\geq \frac{c' \epsilon^2 |S|}{\ln(\frac{1}{\epsilon}) \cdot 2 \cdot 20|S|}. \end{aligned}$$

Hence we obtain that $\Phi(S) \geq c \frac{\epsilon^2}{\ln(\frac{1}{\epsilon})}$, for a fitting choice of c . □

We are now ready to prove Theorem 5.

Proof of Theorem 5. By Theorem 5.1, there is an absolute constant K such that

$$t_{mix} \leq K \sum_{j=1}^{\log_2 \pi_{\min}^{-1}} \Phi^{-2}(2^{-j}).$$

Let C be as defined in Lemma 5.3. Observe that if $2^{-j} > \frac{2C \ln n}{\epsilon^3 n}$, then

$$-j > -\log_2 \left(\frac{\epsilon^3 n}{2C \ln n} \right),$$

and hence $j < \log_2 n$. As such, we may define J to be the set of indices j satisfying $2^{-j} \leq \frac{2C \ln n}{\epsilon^3 n}$ and note that $|J^C| < \log_2 n$. Now, for $i \in J^C$, we have that $2^{-i} > \frac{2C \ln n}{\epsilon^3 n}$, and therefore $\Phi(2^{-i})$ denotes the minimum conductance of connected sets S with $\pi(S) \geq \frac{C \ln n}{\epsilon^3 n}$, by our definition. Furthermore, since $j \geq 1$, we only consider sets S such that $\pi(S) \leq \frac{1}{2}$. Hence, for $i \in J^C$, we may use Lemma 5.3 to obtain that **whp** $\Phi(2^{-i}) \geq c \frac{\epsilon^2}{\ln(\frac{1}{\epsilon})}$. Thus, **whp**

$$\begin{aligned} \sum_{j=1}^{\log_2 \pi_{\min}^{-1}} \Phi^{-2}(2^{-j}) &\leq |J^C| \cdot \left(c \frac{\epsilon^2}{\ln(\frac{1}{\epsilon})} \right)^{-2} + \sum_{j \in J} \Phi^{-2}(2^{-j}) \\ &\leq O(\log n) + \sum_{j \in J} \Phi^{-2}(2^{-j}). \end{aligned}$$

In order to bound $\sum_{j \in J} \Phi^{-2}(2^{-j})$, observe that since L_1 is connected, for every subset $S \subset V(L_1)$ we have:

$$\Phi(S) = \Phi(S^C) \geq \frac{e(S, S^C)}{4e(L_1)\pi(S)} \geq \frac{1}{4e(L_1)\pi(S)}.$$

Recall that the excess of L_1 is **whp** at most $\epsilon^2 n$, and therefore by Theorem 1.1 **whp** $e(L_1) \leq 3\epsilon n$. Furthermore, when considering $\Phi(2^{-j})$, we restrict ourselves to $\pi(S) \leq 2^{-j}$. Thus, **whp**

$$\Phi(S) \geq \frac{1}{4e(L_1)\pi(S)} \geq \frac{2^j}{12\epsilon n}.$$

Hence, we have that **whp**

$$\begin{aligned} \sum_{j \in J} \Phi^{-2}(2^{-j}) &\leq 2 \max_{j \in J} \{2^{-2j}\} 12^2 \epsilon^2 n^2 \\ &\leq 2 \cdot 12^2 \epsilon^2 n^2 \left(\frac{2C \ln n}{\epsilon^3 n} \right)^2 \\ &\leq C' \frac{\ln^2 n}{\epsilon^4}, \end{aligned}$$

completing the proof. □

Acknowledgement The authors wish to thank the anonymous referees for their careful reading and their many insightful comments and suggestions.

References

- [1] M. Ajtai, J. Komlós, and E. Szemerédi, *Largest random component of a k -cube*, *Combinatorica* 2 (1982), 1–7.
- [2] N. Alon and F. R. K. Chung, *Explicit construction of linear sized tolerant networks*, *Discrete Mathematics* 72 (1988), 15–19.
- [3] N. Alon and J. H. Spencer, **The probabilistic method**, 4th Ed., Wiley, New York, 2016.
- [4] I. Benjamini, G. Kozma, and N. Wormald, *The mixing time of a random graph*, *Random Structures & Algorithms* 45 (2014), 383–407.
- [5] A. Beveridge, A. Frieze, and C. McDiarmid, *Random minimum length spanning trees in regular graphs*, *Combinatorica* 18(3) (1998), 311–333.
- [6] B. Bollobás, **Random graphs**, 2nd Ed., Cambridge University Press, Cambridge, 2001.
- [7] B. Bollobás, Y. Kohayakawa, and T. Łuczak, *The evolution of random subgraphs of the cube*, *Random Structures & Algorithms* 3 (1992), 55–90.
- [8] B. Bollobás and O. Riordan, **Percolation**, Cambridge University Press, Cambridge, 2006.
- [9] S. R. Broadbent and J. M. Hammersley, *Percolation processes*, *Mathematical Proceedings of the Cambridge Philosophical Society* 53 (1957), 629–641.
- [10] J. Ding, E. Lubetzky, and Y. Peres, *Anatomy of the giant component: The strictly supercritical regime*, *European Journal of Combinatorics* 35 (2014), 155–168.

- [11] J. Erde, M. Kang, and M. Krivelevich, *Expansion in supercritical random subgraphs of the hypercube and its consequences*, *Annals of Probability* 51 (2023), 127–156.
- [12] P. Erdős and A. Rényi, *On the evolution of random graphs*, *A Magyar Tudományos Akadémia. Matematikai Kutató Intézetének Közleményei* 5 (1960), 17–61.
- [13] D. Fernholz and V. Ramachandran, *The diameter of sparse random graphs*, *Random Structures & Algorithms* 31 (2007), 482–516.
- [14] N. Fountoulakis and B. A. Reed, *Faster mixing and small bottlenecks*, *Probability Theory and Related Fields* 137 (2006), 475–486.
- [15] N. Fountoulakis and B. A. Reed, *The evolution of the mixing rate of a simple random walk on the giant component of a random graph*, *Random Structures & Algorithms* 33 (2008), 68–86.
- [16] A. Frieze and M. Karoński, **Introduction to random graphs**, Cambridge University Press, Cambridge, 2016.
- [17] A. Frieze, M. Krivelevich, and R. Martin, *The emergence of a giant component in random subgraphs of pseudo-random graphs*, *Random Structures & Algorithms* 24 (2004), 42–50.
- [18] G. Grimmett. **Percolation**, 2nd Ed., Springer, Berlin, 1999.
- [19] S. Hoory, N. Linial, and A. Wigderson, *Expander graphs and their applications*, *Bulletin of the American Mathematical Society* 43 (4) (2006), 439–561.
- [20] S. Janson, T. Łuczak, and A. Ruciński, **Random graphs**, Wiley, New York, 2000.
- [21] H. Kesten, **Percolation theory for mathematicians**, Springer, New York, 1982.
- [22] M. Krivelevich, *Finding and using expanders in locally sparse graphs*, *SIAM Journal on Discrete Mathematics*, 32(1) (2018), 611–623.
- [23] M. Krivelevich, *Expanders — how to find them, and what to find in them*, *Surveys in Combinatorics 2019*, A. Lo et al., Eds., London Mathematical Society Lecture Notes 456 (2019), 115–142.
- [24] M. Krivelevich, E. Lubetzky, and B. Sudakov, *Asymptotics in percolation on high-girth expanders*, *Random Structures & Algorithms* 56 (2020), 927–947.
- [25] M. Krivelevich, D. Reichman, and W. Samotij. *Smoothed analysis on connected graphs*, *SIAM Journal of Discrete Mathematics*, 29(3) (2015), 1654–1669.
- [26] M. Krivelevich and B. Sudakov, *Pseudo-random graphs*, *More sets, graphs and numbers*, E. Györi, G. O. H. Katona and L. Lovász, Eds., Bolyai Society Mathematical Studies vol. 15 (2006), 199–262.
- [27] M. Krivelevich and B. Sudakov, *The phase transition in random graphs — a simple proof*, *Random Structures & Algorithms* 43 (2013), 131–138.
- [28] D. A. Levin, Y. Peres, and E. L. Wilmer, **Markov chains and mixing times**, Providence, RI: American Mathematical Society, 2017.
- [29] C. McDiarmid, *Concentration*, *Algorithms and Combinatorics*, M. Habib, C. McDiarmid, J. Ramirez-Alfonsin, B. Reed, Eds., *Probabilistic Methods for Algorithmic Discrete Mathematics* vol. 16 (1998), 195–248.

- [30] E. Ofek, *On the expansion of the giant component in percolated (n, d, λ) graphs*, *Combinatorics, Probability & Computing* 16 (2007), 445–457.
- [31] O. Riordan and N. Wormald, *The diameter of sparse random graphs*, *Combinatorics, Probability & Computing* 19 (2010), 835–926.