Smoothed analysis on connected graphs

Michael Krivelevich[*]   Daniel Reichman[†]   Wojciech Samotij [‡]

November 12, 2013

Abstract

The main paradigm of smoothed analysis on graphs suggests that for any large graph G in a certain class of graphs, perturbing slightly the edges of G at random (usually adding few random edges to G) typically results in a graph having much "nicer" properties. In this work we study smoothed analysis on trees or, equivalently, on connected graphs. Given an n-vertex connected graph G, form a random supergraph G^* of G by turning every pair of vertices of G into an edge with probability $\frac{\varepsilon}{n}$, where ε is a small positive constant. This perturbation model has been studied previously in several contexts, including smoothed analysis, small world networks, and combinatorics.

Connected graphs can be bad expanders, can have very large diameter, and possibly contain no long paths. In contrast, we show that if G is an n-vertex connected graph then typically G^* has edge expansion $\Omega(\frac{1}{\log n})$, diameter $O(\log n)$, vertex expansion $\Omega(\frac{1}{\log n})$, and contains a path of length $\Omega(n)$, where for the last two properties we additionally assume that G has bounded maximum degree. Moreover, we show that if G has bounded degeneracy, then typically the mixing time of the lazy random walk on G^* is $O(\log^2 n)$. All these results are asymptotically tight.

1 Introduction

In this paper, we consider the following model of randomly generated graphs. We are given a fixed undirected graph $G = (V, E)$ on n vertices. For every pair $f \in \binom{V}{2}$, we add f to G, independently of all other pairs, with probability $\frac{\varepsilon}{n}$, where ε is a small (yet fixed) positive constant. Let R be the set of edges added and consider the random graph

$$G^* := (V, E \cup R).$$

This model can be viewed as a generalization of the classical Erdős-Rényi random graph, where one starts from an empty graph and adds edges between all possible pairs of vertices independently with a given probability. The focus on "small" ε means that we are interested in the effect of a rather gentle random perturbation. In particular, the average degree of G^* is (typically) close to that of G (assuming that G is connected, for example). Studying the effect of small perturbations on graphs, matrices, and other structures arises in diverse settings in several fields such as combinatorics, design and analysis of algorithms, linear algebra, and mathematical programming. We refer the reader to Section 1.3 for more details.

In this work, we study several properties of G^*, when G is *connected*. We first need a few definitions. For a graph $G = (V, E)$ and a subset $S \subseteq V$, we denote by ∂S the set of all edges of G with exactly one endpoint in S. We define $N(S)$ to be the set of all vertices in $V \setminus S$ that have a neighbor in S. When the graph G is not clear from the context, we will use the notation $\partial_G S$ and $N_G(S)$ to avoid ambiguity. The *edge*-isoperimetric number of G (also known as the *Cheeger constant*), denoted $c(G)$, is defined by

$$c(G) := \min\left\{ \frac{|\partial(U)|}{|U|} : 0 < |U| \leqslant \frac{|V|}{2} \right\}.$$

Similarly, the *vertex*-isoperimetric number of G, denoted $\iota(G)$, is defined by

$$\iota(G) := \min\left\{ \frac{|N(U)|}{|U|} : 0 < |U| \leqslant \frac{|V|}{2} \right\}.$$

Somewhat informally we shall refer to $c(G)$ and $\iota(G)$ as the edge and the vertex expansions of G, respectively. Observe that $\iota(G) \geqslant c(G)/\Delta(G)$ where $\Delta(G)$ is the maximum degree of G. Hence, when $\Delta(G)$ is bounded by a constant, then the vertex and edge expansions of G have the same order of magnitude. On the other hand, there are n-vertex graphs G for which $\iota(G) = O(c(G)/n)$. The Cheeger constant has been studied extensively as it is related to a host of combinatorial properties of the underlying graph. In particular, there is a strong connection between the Cheeger constant of G and the mixing time of the lazy random walk on G.

1.1 Our results

We begin by describing our results regarding the expansion properties of perturbed connected graphs. Our results concerning expansion allow us to deduce upper bounds on related parameters such as the diameter and the mixing time. The connection between expansion and long paths is perhaps less straightforward and therefore we prove the existence of long paths in randomly perturbed connected graphs without relying on their expansion properties. We do note

that there is a connection between expansion and long paths [26], and in fact one could have built on this connection in demonstrating the existence of long paths in our setting.

If G is disconnected, then clearly both the vertex and the edge expansion of G are 0. For every *connected* n-vertex graph G, it holds that $\iota(G) = \Omega(\frac{1}{n})$ as every subset $S \subseteq V$ has at least one neighbor outside S. Moreover, if G is a tree, then $\iota(G) = O(\frac{1}{n})$. Our first result is that for every *connected* graph with bounded maximum degree, the random perturbation G^* asymptotically almost surely[1] (a.a.s.) satisfies $\iota(G^*) = \Omega(\frac{1}{\log n})$.

Theorem 1. *For every $\varepsilon > 0$, there exists $\delta > 0$ such that the following holds. Let G be an n-vertex connected graph with maximum degree Δ. If $R \sim G(n, \frac{\varepsilon}{n})$, then a.a.s. the graph $G^* = G \cup R$ has vertex expansion at least $\frac{\delta}{\Delta^3 \log n}$.*

We note that in general one cannot remove restrictions on the maximum degree entirely. To see this, consider the case when $G = K_{1,n-1}$. After adding to G *any* εn edges, there will be an independent set S with at least $(1 - 2\varepsilon)n$ vertices such that $|N(S)| = 1$.

We obtain a similar bound on the *edge*-expansion without any assumptions on the maximum degree.

Theorem 2. *For every $\varepsilon > 0$ and $\alpha < 1$, there exists $\delta > 0$ such that the following holds. Let G be an n-vertex connected graph, choose $R \sim G(n, \frac{\varepsilon}{n})$, and let $G^* = G \cup R$. Then a.a.s. for every set $S \subseteq V(G)$ with $|S| \leqslant \alpha n$,*

$$|\partial_{G^*} S| \geqslant \frac{\delta}{\log(en/|S|)}|S|.$$

In particular, $c(G^*) \geqslant \frac{\delta}{\log(en)}$.

It should be noted that Theorem 2 implies that a.a.s. the vertex expansion of G is at least $\frac{\delta}{\Delta(G)\log(en)}$. This improves the bound obtained in Theorem 1 when $\Delta(G) \gg \left(\frac{\log n}{\log\log n}\right)^{1/3}$; to see this, observe that $G(n, \epsilon/n)$ a.a.s. contains vertices of degree $\Omega\left(\frac{\log n}{\log\log n}\right)$.

Furthermore, we prove an even stronger bound on the edge expansion of connected subsets of a perturbed connected graph.

Theorem 3. *For every $\varepsilon > 0$ and $\alpha < 1$, there exist $\delta > 0$ and $K > 0$ such that the following holds. Let G be an n-vertex connected graph, choose $R \sim G(n, \varepsilon/n)$, and let $G^* = G \cup R$. Then a.a.s. for every connected (in G^*) set $S \subseteq V(G)$ with $K \log n \leqslant |S| \leqslant \alpha n$,*

$$|\partial_{G^*} S| \geqslant \delta |S|.$$

One may wonder why we consider sets of size all the way up to αn for an arbitrary $\alpha < 1$ instead of restricting our attention only to sets of size at most $n/2$, as is customary in dealing with vertex and edge expansion. The reason is that this allows us later to give upper bounds on the conductance of sets of volume up to a half of the total volume, where volume is measured in terms of the degree sum rather than the number of vertices, which is crucial for the proof of Theorem 5 stated below.

Using Theorem 3, we derive the following upper bound on the diameter of a randomly perturbed connected graph. Observe that the diameter of a (non-perturbed) n-vertex connected graph may be as high as $n - 1$ (when the graph is a path on n vertices).

---

[1]That is, with probability tending to 1 as the number of vertices n tends to infinity.

Theorem 4. *For every $\varepsilon > 0$, there exists $C > 0$ such that the following holds. Let G be an n-vertex connected graph, choose $R \sim G(n, \frac{\varepsilon}{n})$, and let $G^* = G \cup R$. Then a.a.s. the diameter of G^* is at most $C \log n$.*

Using Theorem 3, we also prove upper bounds on the mixing times of lazy random walks on randomly perturbed connected graphs. Recall the notion of *degeneracy*. Given a positive integer D, a graph G is called D-*degenerate* if every subgraph of G contains a vertex of degree at most D. Observe that every graph G is $\Delta(G)$-degenerate and trees are 1-degenerate. Also, if G is D-degenerate, then every subset $S \subseteq V(G)$ spans at most $D|S|$ edges. Using the machinery developed by Fountoulakis and Reed [16], we are able to prove the following bound on the mixing time of the lazy random walk on a random pertubation of a connected graph with bounded degeneracy.

Theorem 5. *For all positive D and ε, there exists a constant M such that the following holds. Let G be an n-vertex D-degenerate connected graph, choose $R \sim G(n, \frac{\varepsilon}{n})$ and let $G^* = G \cup R$. Then a.a.s.*
$$T_{\mathrm{mix}}(G^*) \leqslant M \log^2 n.$$

For a precise definition of T_{mix}, we refer the reader to Section 2. The bound in Theorem 5 above is tight when G is the path on n vertices, as then a.a.s. G^* contains an induced subgraph which is a path of length $\Omega(\log n)$. Moreover, we cannot expect that for arbitrary connected graph G, the mixing time on the perturbed graph G^* will be $O(\log^2 n)$ as the following example demonstrates. Let G be the graph obtained by connecting two disjoint cliques of order $n/2$ with a single edge and let $R \sim G(n, \frac{1}{n})$. As the number of edges interconnecting the two cliques in the perturbed graph is a.a.s. $O(n)$, the conductance of G^* is $O(\frac{1}{n})$, which implies via standard results (e.g., [22]) that the mixing time of the lazy random walk on G^* is $\Omega(n)$.

Finally, we establish the existence of long paths in perturbed connected graphs with bounded maximum degree. Observe that a connected bounded degree graph with n vertices might contain only paths of length $O(\log n)$, as the case of the complete binary tree demonstrates.

Theorem 6. *For every $\varepsilon, \Delta > 0$, there exists $c > 0$ such that the following holds. Let G be an n-vertex connected graph with maximum degree bounded by Δ. Form a random graph $R \sim G(n, \frac{\varepsilon}{n})$, and let $G^* = G \cup R$. Then G^* a.a.s. contains a path of length cn.*

The assumption that the maximum degree is bounded is crucial, as it is easy to see that if $G = K_{1,n-1}$ and $\varepsilon < 1$, then a.a.s. the length of a longest path in $G \cup R$ is $O(\log n)$. This follows as it is known that a.a.s. each connected component of $G(n, \frac{\varepsilon}{n})$ has $O(\log n)$ vertices and the vertex set of any path in G^* induces at most two connected components in R.

1.2   Our techniques

While dealing with connected graphs with bounded maximum degree, we use a fairly basic result (see e.g., [19]) to decompose the graph to disjoint connected sets of comparable sizes. Treating each of these sets as a 'super-vertex' allows us to view the auxiliary graph induced by the random edges between sets as essentially the standard binomial random graph whose edge probability should be now compared to the number of super-vertices as opposed to the (much larger) number of vertices. Consequently, standard methods and results regarding the

threshold for connectivity and the existence of long paths in binomial random graphs can be used.

In order to deal with graphs with unbounded degrees, we prove a new upper bound on the number of connected subsets of given cardinality and number of vertices in their boundary. We believe that this bound (stated below), which we prove using an elementary argument, may be of independent interest. Recall that a subset of vertices of a graph is *connected*, if it induces a connected subgraph.

Proposition 7. *Let G be an arbitrary graph and let $v \in V(G)$. For integers a and b, let $\mathcal{C}(v, a, b)$ denote the collection of connected subsets A of $V(G)$ such that $v \in A$, $|A| = a$, and $|N(A)| = b$. Then*
$$|\mathcal{C}(v, a, b)| \leqslant \binom{a+b-1}{b}.$$

We remark that the bound in Proposition 7 is tight for all values of a and b. To see this, consider the case when $G = K_{1,a+b-1}$ and v is the center vertex.

In bounding the mixing time, we rely on an upper bound on the mixing time of a lazy random walk due to Fountoulakis and Reed [16]. This bound, which they used [17] to upper bound the mixing time of the lazy random walk on the giant component of $G(n, p)$, is suited for bounding the mixing time of random walks on graphs whose large vertex sets expand well but small sets (e.g., of logarithmic size) do not have to. Another attractive feature of the result of Fountoulakis and Reed is that it allows one to focus on the conductance of connected sets, which significantly simplifies union bound estimates. We note that the classical work of Jerrum and Sinclair [18] for upper-bounding the mixing time T_{mix} in terms of the conductance Φ of G (see Section 2 for precise definitions), namely
$$T_{\mathrm{mix}} \leqslant O\left(\frac{\log n}{\Phi^2}\right)$$
would give in our setting a weaker bound of $O(\log^3 n)$.

1.3 Motivation and related work

The study of random perturbations of graphs arose in several contexts. One of them is the field of *smoothed analysis*, which originated from the work of Spielman and Teng [29] on the smoothed complexity of the simplex algorithm. This field attempts to provide a theoretical explanation for the good performance of certain heuristics on "real-life" instances based on the assumption that they are likely to be subjected to random perturbations. It has been applied to a host of other problems such as numerical analysis and linear algebra [27, 31], machine learning [5], and satisfiability [9, 12]. It is closely related to the study of random perturbations of combinatorial structures and devising efficient algorithms for such "semi-random" instances, which had been considered in the past, see [6, 7, 13, 15, 21, 28, 30].

Another context where the study of random perturbations naturally arose, is the field of small world networks, see [11, 24, 25]. In an attempt to model social networks arising in "real-life" settings, one studies properties of networks composed of a (usually sparse) connected "base" graph along with a set of random edges, where every random edge is added independently with probability p. One well-known example is the Newmann–Watts small world model [24, 25]

(NW small world for short), where the base graph is the (n,k)-ring, i.e., the graph with vertex set $\{0, ..., n-1\}$ and edge set $\big\{\{i,j\} : i+1 \leqslant j \leqslant i+k\big\}$ (where addition is modulo n) and p is equal to $\frac{c}{n}$ for some constant $c > 0$.

Durrett [11] showed that with high probability the mixing time of the lazy random walk on the NW small world is upper-bounded by $O(\log^3 n)$ and lower-bounded by $\Omega(\log^2 n)$. These results were improved by Addario-Berry and Lei [1] who proved that this mixing time is a.a.s. $O(\log^2 n)$. It is worth noting that our approach is similar to [1] in the sense that we bound the conductance of connected sets and then use this upper bound with the results of [16] to bound the mixing time. The crucial difference between our proof and theirs is the technique of counting connected sets with small boundary. While [1] uses a somewhat involved argument based on the Lagrange inversion formula, we use a more elementary approach based on Proposition 7.

Similar ideas were used in the study of the mixing time of the simple random walk on the giant component in a supercritical random graph $G(n, \frac{1+\varepsilon}{n})$. Fountolakis and Reed [16] and Benjamini, Kozma, and Wormald [3] showed that a.a.s. this mixing time is $O(\log^2 n)$. Moreover, there has been interest in probability theory in studying the robustness of the mixing time under random perturbations, see [4, 10].

Flaxman [15] examined the edge expansion of several models of randomly perturbed graphs. In particular, he considered the model studied in this work. He showed in particular that if $G = (V, E)$ is an n-vertex connected graph and $R \sim G(n, \frac{\varepsilon}{n})$, then a.a.s. all linear sized vertex subsets $S \subseteq V$, $|S| \leqslant n/2$, send outside at least a linear in n number of edges in $G^* = G \cup R$. The effect of adding random edges on the diameter of a given graph seems to have been considered first by Bollobás and Chung [8], who proved that adding a random matching to an n-vertex cycle result a.a.s. with a graph with diameter $(1 + o(1))\log_2 n$. The case of *directed* graphs was considered by Flaxman and Frieze [14]. They proved that if D is an n-vertex strongly connected digraph with maximum degree bounded by $n^{\frac{\varepsilon}{100}}$ and $R \sim D(n, \frac{\varepsilon}{n})$, then a.a.s. the diameter of $D \cup R$ is at most $100\varepsilon^{-1} \log n$. Our proof idea is different from theirs. Moreover, unlike their work, our upper bounds on the diameter of G^* hold unconditionally, regardless of the maximum degree in the base graph G.

1.4 Outline of the paper

In Section 2, we fix some notation, give a precise definition of the mixing time of a random walk, and state two auxiliary probabilistic lemmas that are used later in the paper. In Sections 3, 4, 5, 6, and 7, we prove Theorems 1, 2, 4, 5, and 6, respectively. Section 4 contains also the proofs of Theorem 3 and of Proposition 7. In Section 8, we state several concluding remarks.

2 Preliminaries

Let G be a graph with vertex set V. Given two disjoint sets $A, B \subseteq V$, we denote by $E(A, B)$ the set of all edges with one endpoint in A and one endpoint in B and by $E(A)$ the set of all edges entirely contained in A. We will denote the cardinality of $E(A)$ by $e(A)$. The degree of a vertex v in G is denoted by $\deg(v)$ and the maximum degree of G is denoted by $\Delta(G)$.

We denote by $[n]$ the set $\{1, \ldots, n\}$. When dealing with an n-vertex graph, we will implicitly assume that that its vertex set is $[n]$. We denote by $G(n, p)$ the classical binomial random graph with vertex set $[n]$ and edge probability p. Given a graph property $\mathcal{P}$ and a sequence (μ_n), where

μ_n is a probability distribution over n-vertex graphs, we will say that $\mathcal{P}$ holds asymptotically almost surely (a.a.s) if $\lim_{n\to\infty} \Pr_{G\sim\mu_n}(G \in \mathcal{P}) = 1$.

The lazy random walk on a graph $G = (V, E)$ is the Markov chain defined as follows. The set of states is V. For any vertex $u \in V$, the walk stays in u with probability $\frac{1}{2}$ and with probability $\frac{1}{2}$, it moves to a uniformly chosen random neighbor v of u (so that the transition probability $\Pr(u \to v)$ is $\frac{1}{2\deg(u)}$). When G is connected, this Markov chain is well-known to be irreducible and ergodic and hence it converges to a stationary distribution π which can be seen to equal $\pi(u) = \frac{\deg(u)}{2|E|}$ for every $u \in V$, see [22]. We will be interested in estimating how quickly this random walk on G converges to its stationary distribution π. To this end, we recall that the *total variation distance* d_{TV} between two distributions p_1, p_2 on V is defined by

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

Let P be the transition matrix of the random walk. The mixing time $T_{\mathrm{mix}}(G)$ is defined by

$$T_{\mathrm{mix}}(G) := \sup_{x_0} \min \left\{ t \colon d_{TV}(x_0 P^t, \pi) \leqslant \frac{1}{4} \right\},$$

where the supremum is taken over all probability distributions x_0 on V.

In the proofs of Theorems 4 and 5, we will need the following standard probabilistic estimate. We present its proof for the sake of completeness.

Lemma 8. *For every $C \geqslant 1$, if $p \leqslant C/n$, then a.a.s. for every non-empty set S of vertices in $G(n,p)$, we have $e(S) < 2C|S|$. In particular, if $p \leqslant \frac{1}{n}$, then a.a.s. for every non-empty set S of vertices in $G(n,p)$, we have $e(S) < 2|S|$.*

Proof. For a fixed set S of size k,

$$\Pr\left(e(S) \geqslant 2Ck\right) \leqslant \binom{\binom{k}{2}}{2Ck} p^{2Ck} \leqslant \left(\frac{ekp}{4C}\right)^{2Ck}$$

and hence, letting $\mathcal{E}$ denote the event that $e(S) \geqslant 2C|S|$ for some $S \neq \emptyset$,

$$\Pr(\mathcal{E}) \leqslant \Pr\left(e(G(n,p)) \geqslant Cn\right) + \sum_{k=5}^{n/2} \binom{n}{k} \left(\frac{ekp}{4C}\right)^{2Ck}$$

$$\leqslant \sum_{k=5}^{n/2} \left(\frac{en}{k} \cdot \left(\frac{ekp}{4C}\right)^{2C}\right)^k + o(1) = o(1).$$

To see the last inequality, note that if $k \leqslant \sqrt{n}$, then $np^{2C}k^{2C-1} \leqslant 1/\sqrt{n}$ and if $k \leqslant n/2$, then

$$\frac{en}{k} \cdot \left(\frac{ekp}{4C}\right)^{2C} \leqslant 2e \cdot \left(\frac{e}{8}\right)^{2C} \leqslant \frac{e^3}{32} \leqslant \frac{2}{3}. \qquad \square$$

We close this section with a version of Chernoff's inequality (see, e.g., [23]).

Lemma 9. *Suppose that $X = \sum_{i=1}^{m} X_i$ where every X_i is a $\{0,1\}$-random variable with $\Pr(X_i = 1) = p$ and the X_is are jointly independent. Then for arbitrary $\eta \in (0,1)$,*

$$\Pr(X < (1 - \eta)pm) \leqslant \exp(-pm\eta^2/2).$$

Finally, we remark that for the sake of clarity of the presentation, we will omit all floor and ceiling signs.

6

3 Vertex expansion

Proof of Theorem 1. Let

$$k = C\Delta \log n,$$

where $C = C(\varepsilon)$ is a large enough constant, which we will define later. Partition the vertex set of G into *disjoint* pieces $V_1, \dots, V_t$, such that for each $1 \leqslant i \leqslant t$, we have $k \leqslant |V_i| \leqslant \Delta k$ and $G[V_i]$ is *connected*. This is fairly straightforward, see, e.g., [19, Proposition 4.5]. Observe that

$$\frac{n}{\Delta k} \leqslant t \leqslant \frac{n}{k}.$$

Call each V_i a *blob*.

The probabilistic statement about the random graph R we need is the following one: a.a.s. for every non-empty $I \subseteq [t]$ with $|I| \leqslant t/2$, there are at least $|I|/2$ blobs outside of I that are connected by an edge to the set $\bigcup_{i \in I} V_i$. Let

$$\rho := \frac{\varepsilon k^2}{2n}.$$

Clearly, the probability that two blobs are connected in $G \cup R$ is at least $1 - (1 - p)^{|V_i| \cdot |V_j|} \geqslant \rho$ (if the two blobs are connected in G, then this probability is 1). The probability P that there exists a non-empty set $I \subseteq [t]$ with $|I| \leqslant t/2$ such that the set $\bigcup_{i \in I} V_i$ has an edge (in R) to fewer than $|I|/2$ blobs outside of I satisfies

$$P \leqslant \sum_{1 \leqslant j \leqslant t/2} \binom{t}{j} \binom{t-j}{j/2} (1 - \rho)^{(t - \frac{3j}{2})j} \leqslant \sum_{1 \leqslant j \leqslant t/2} t^{\frac{3j}{2}+1} \cdot \exp\left(-\frac{\rho j t}{4}\right).$$

It is easy to verify that $P = o(1)$ if we take $C(\epsilon) \geqslant C'/\varepsilon$ for a sufficiently large absolute constant C'.

Suppose that R has the above property. We claim that $G \cup R$ has vertex expansion at least $\frac{\delta}{\Delta^3 \log n}$ for some positive constant $\delta = \delta(\varepsilon)$. Fix a set $A \subseteq [n]$ with $|A| \leqslant n/2$ and denote

$$
\begin{aligned}
I_0 &= I_0(A) = \{1 \leqslant i \leqslant t \colon V_i \subseteq A\}, \\
I_1 &= I_1(A) = \{1 \leqslant i \leqslant t \colon \emptyset \neq V_i \cap A \neq V_i\}, \\
I_2 &= I_2(A) = \{i \notin I_0 \colon V_i \text{ has a neighbor in } A\}.
\end{aligned}
$$

In other words, I_0 is the set of (indices of) blobs fully contained in A, I_1 is the set of blobs having a vertex in A but not falling completely inside A, and finally I_2 is the set of blobs outside I_0 having a neighbor in A.

It follows from the above stated property of R that a.a.s.

$$|I_2| \geqslant \min\left\{\frac{|I_0|}{2}, \frac{t - |I_0|}{3}\right\}.$$

This is clear when $|I_0| \leqslant t/2$; we simply take $I = I_0$. Else, we let $I = [t] \setminus (I_0 \cup I_2)$, note that $|I| \leqslant t - |I_0| \leqslant t/2$, and observe that no blob in I is connected to a blob in I_0 (and hence the neighborhood of I must be completely contained in I_2). Observe crucially that $|N(A)| \geqslant |I_1|$

7

since each set $V_i \cap A$ with $i \in I_1$ has at least one neighbor in $V_i \setminus A$ due to the connectivity of the blob $G[V_i]$. Also, $|N(A)| \geqslant |I_2 \setminus I_1|$ as for every blob V_i with $i \in I_2 \setminus I_1$, A has a neighbor in $V_i \setminus A$. Finally, note that $A \subseteq \bigcup_{i \in I_0 \cup I_1} V_i$, implying that

$$|I_0| + |I_1| \geqslant \frac{|A|}{\Delta k}. \tag{1}$$

If $|I_1| \geqslant \frac{|I_0|}{6\Delta}$, then it follows from (1) that $|I_1| \geqslant \frac{|A|}{\Delta(1+6\Delta)k}$ and thus $|N(A)| \geqslant |I_1| \geqslant \frac{|A|}{\Delta(1+6\Delta)k}$. If $|I_1| \leqslant \frac{|I_0|}{6\Delta}$, then we distinguish between two cases, depending on the value of $\min\{\frac{|I_0|}{2}, \frac{t-|I_0|}{3}\}$. If $\frac{|I_0|}{2} \leqslant \frac{t-|I_0|}{3}$, then $|I_2 \setminus I_1| \geqslant \frac{|I_0|}{2} - \frac{|I_0|}{6\Delta} = \frac{|I_0|}{3}$. As $|N(A)| \geqslant |I_2 \setminus I_1|$, we get by (1) and our assumption on $|I_1|$ that $|N(A)| \geqslant \frac{|I_0|}{3} \geqslant \frac{(1+1/(6\Delta))|A|}{3\Delta k}$. Else, if $\frac{t-|I_0|}{3} < \frac{|I_0|}{2}$, we observe that, as $|A| \leqslant \frac{n}{2}$ and $A^c \subseteq \bigcup_{i \notin I_0} V_i$, we have $|I_0| \leqslant (1-\frac{1}{\Delta})t$. In this case, $|I_2 \setminus I_1| \geqslant \frac{t-|I_0|}{3} - \frac{|I_0|}{6\Delta}$ and hence $|N(A)| \geqslant \frac{t}{6\Delta} \geqslant \frac{|A|}{6\Delta^2 k}$. Hence, G^* has vertex expansion at least $\frac{\delta}{\Delta^3 \log n}$, where $\delta = \delta'\varepsilon$ and δ' is an absolute positive constant. $\square$

Remark. Observe that the exact same proof as above works if instead of assuming that the graph G is connected and has maximum degree bounded by Δ, we assume that $\Delta(G) \leqslant \Delta$ and all connected components of G are at least as large as $C\Delta \log n$, where $C = C(\epsilon)$ is a large enough constant.

It is natural to ask what happens when $\epsilon(n)$ tends to zero with n. Similar ideas to those used in our proof of Theorem 1 apply in this case as well. We illustrate this in the case when $\varepsilon = n^{-a}$ for some $a \in (0,1)$. Observe that if $p = O(\frac{1}{n^2})$, then a.a.s. the number of random edges that are added is constant, hence if G is a tree, then a.a.s. G^* has expansion $O(\frac{1}{n})$. As the proof closely follows the lines of the above proof of Theorem 1, we only give an outline of the arguments, omitting some of the details.

Proposition 10. *Let G be an n-vertex connected graph with maximum degree Δ and set $\varepsilon = n^{-a}$ for some $a \in (0,1)$. If $R \sim G(n, \frac{\varepsilon}{n})$, then a.a.s. the graph $G^* = G \cup R$ has vertex expansion at least $\Omega(\frac{1}{\log n \cdot n^a \cdot \Delta^3})$.*

Proof. Let $k = C\Delta \log n \cdot n^a$ where C is a sufficiently large constant. Partition the vertex set of G into disjoint connected blobs with each blob of size between k and Δk. The number of blobs t again satisfies $\frac{n}{\Delta k} \leqslant t \leqslant \frac{n}{k}$. Similar arguments to those in the proof of Theorem 1 show that a.a.s. for every set I of size at most $t/2$, there are at least $|I|/2$ blobs outside of I that are connected by an edge to the set $\bigcup_{i \in I} V_i$. This implies, as before, that a.a.s. the expansion of G^* is $\Omega(\frac{1}{\log n \cdot n^a \cdot \Delta^3})$. $\square$

4 Edge expansion

In this section, we prove Proposition 7 and derive from it Theorem 2. We conclude by proving Theorem 3.

Proof of Proposition 7. Assume that the vertices of G are labeled with distinct integers. We will describe an algorithm that, given an $A \in \mathcal{C}(v, a, b)$, outputs an encoding of A using a

sequence of $a - 1$ ones and b zeros in such a way that no two sets are encoded with the same sequence. This will clearly imply the statement of the proposition.

Let $S = \{v\}$ and $B = \emptyset$. The algorithm will grow the sets S and B, adding one vertex to one of the sets in each of its $a + b - 1$ iterations, making sure that the invariants $S \subseteq A$ and $B \subseteq N(A)$ hold in every iteration. It will stop when $S = A$ and $B = N(A)$, after having moved $a - 1$ vertices to S and b vertices to B. For the sake of brevity, we will denote by T the set $N(S) \setminus B$, updated after each iteration. Intuitively, in every iteration, S is the set of vertices that are known to belong to A, B is the set of vertices that belong to $N(A)$, and T is the remaining set of vertices for which we do not know yet whether they belong to A or to $N(A)$.

While $S \neq A$ or $B \neq N(A)$, we repeat the following. Let w be the vertex with the smallest label in T. Note that the assumption that A is connected implies that T is non-empty. Consider two cases. If $w \in A$, then move w to S and append 1 to the sequence encoding A. Otherwise, if $w \notin A$, then move w to B and append 0 to the sequence encoding A. Note that in this case $w \in N(A)$, since $S \subseteq A$ and $w \in N(S) \setminus A$.

A moment of thought reveals that the above encoding algorithm can easily be reversed and given v and the $\{0, 1\}$-sequence encoding A, one can recover the set A. This completes the proof. $\qquad\square$

Proof of Theorem 2. For positive integers s, m, and b, denote by $\mathcal{S}(s, m, b)$ the collection of all sets S of s vertices such that in the graph G, the set S induces exactly m connected components and the sum of their vertex boundaries is exactly b. In other words, $\mathcal{S}(s, m, b)$ consists of all sets $S \subseteq V(G)$ such that there is a partition $S = S_1 \cup \ldots \cup S_m$, where each S_i is connected, there are no edges of G connecting different S_i, and $|N_G(S_1)| + \ldots + |N_G(S_m)| = b$. Since G is connected, each $S \in \mathcal{S}(s, m, b)$ satisfies $|\partial_G S| \geqslant b \geqslant m \geqslant 1$ (but not necessarily $|N_G(S)| \geqslant b$). Therefore, it is enough to show that there exist positive constants K and δ such that a.a.s. for every s satisfying $s \geqslant K \log n$,

$$|\partial_R S| \geqslant \frac{\delta s}{\log(en/s)} \quad \text{for all } S \in \mathcal{S}(s, m, b) \text{ with } m \leqslant b \leqslant \frac{\delta s}{\log(en/s)}. \tag{2}$$

(For small sets S with $|S| \leqslant K \log n$, we have that $|\partial_{G^*} S| \geqslant |\partial_G S| \geqslant 1 \geqslant \frac{\delta s}{\log(en/s)}$, since we may assume that $K\delta \leqslant 1/2$.) In order to facilitate a union bound argument, we will estimate the size of $\mathcal{S}(s, m, b)$ with small b and m using Proposition 7. To this end, we first argue that each set in $\mathcal{S}(s, m, b)$ can be uniquely described by the following:

  (i) a set $W = \{v_1, \ldots, v_m\}$ of m vertices of G,

  (ii) a partition $s = s_1 + \ldots + s_m$, where $s_i \geqslant 1$ for each i,

  (iii) a partition $b = b_1 + \ldots + b_m$, where $b_i \geqslant 1$ for each i, and

  (iv) a set S_i in $\mathcal{C}(v_i, s_i, b_i)$ for each $i \in [m]$.

To see this, note that we may assume that there is a canonical linear ordering on the vertices of G. The representation of S in $\mathcal{S}(s, m, b)$ as (i)–(iv) is natural. Indeed, given such an S, we find the unique partition $\{S_1, \ldots, S_m\}$ into connected components of $G[S]$ and arbitrarily choose one vertex from each S_i to form W. We order the sets $S_1, \ldots, S_m$ according to the canonical

linear ordering on their representatives $v_1, \ldots, v_m$. Finally, we let $s_i = |S_i|$ and $b_i = |\partial_G S_i|$. Observe that this mapping is not only injective, but actually each set S can be represented in $s_1 \cdot \ldots \cdot s_m$ different ways.

It follows from Proposition 7, as well as from the inequality $\binom{x}{y}\binom{w}{z} \leqslant \binom{x+w}{y+z}$, that

$$
|\mathcal{S}(s,m,b)| \leqslant \binom{n}{m} \sum_{(s_i),(b_i)} \prod_{i=1}^{m} \binom{s_i + b_i - 1}{b_i} \leqslant \binom{n}{m} \sum_{(s_i),(b_i)} \binom{\sum_i (s_i + b_i - 1)}{\sum_i b_i}
$$
$$
\leqslant \binom{n}{m}\binom{s-1}{m-1}\binom{b-1}{m-1}\binom{s+b-m}{b} \leqslant \binom{n}{m}\binom{s}{m}\binom{b}{m}\binom{s+b}{b}.
$$

Consequently, if $m \leqslant b \leqslant \delta s/\log(en/s) \leqslant s$, then it follows from the well-known estimate $\binom{x}{y} \leqslant \left(\frac{ex}{y}\right)^y$ and the fact that the function $y \mapsto (ex/y)^y$ is increasing on the interval $(0, x]$ that

$$
|\mathcal{S}(s,m,b)| \leqslant \left(\frac{en}{m}\right)^m \left(\frac{es}{m}\right)^m \left(\frac{eb}{m}\right)^m \left(\frac{e(s+b)}{b}\right)^b \leqslant \left(\frac{e^4 n s(s+b)}{b^3}\right)^b
$$
$$
\leqslant \left(\frac{2e^4 n(\log(en/s))^3}{\delta^3 s}\right)^{\frac{\delta s}{\log(en/s)}} \leqslant \exp\left(C\delta\log(1/\delta)s\right),
$$

where C is some absolute constant.

On the other hand, by Chernoff's inequality, for a fixed set S with $|S| = s \leqslant \alpha n$,

$$
\Pr(|\partial_R S| < \delta s) \leqslant \Pr\left(\mathrm{Bin}(s(n-s), \varepsilon/n) < \delta s\right)
$$
$$
\leqslant \Pr\left(\mathrm{Bin}((1-\alpha)sn, \varepsilon/n) < \delta s\right) \leqslant \exp\left(-(1-\alpha)\varepsilon s/8\right).
$$

provided that $\delta < (1-\alpha)\varepsilon/2$.

Finally, choose positive constants K and δ such that

$$
K \geqslant \frac{64}{\varepsilon(1-\alpha)}, \qquad C\delta\log(1/\delta) \leqslant \frac{\varepsilon(1-\alpha)}{16}, \qquad \text{and} \qquad K\delta leq 1/2.
$$

Taking a union bound over all triples b, m, and s satisfying $K \log n \leqslant s \leqslant \alpha n$ and $m \leqslant b \leqslant \delta s/\log(en/s)$, we get that

$$
\Pr(\text{property (2) fails}) \leqslant n^3 \exp\left[\left(C\delta\log(1/\delta) - \frac{\varepsilon(1-\alpha)}{8}\right)s\right] = o(1). \qquad \square
$$

Proof of Theorem 3. Due to the obvious monotonicity we can assume that $\epsilon < 1$. Recall the definition of $\mathcal{S}(s,m,b)$ from the proof of Theorem 2. It clearly suffices to show that a.a.s., for every s with $K \log n \leqslant s \leqslant \alpha n$,

$$
|\partial_R S| \geqslant \delta s \quad \text{for all } \textit{connected } S \in \mathcal{S}(s,m,b) \text{ with } m \leqslant b < \delta s, \tag{3}
$$

where connected means connected in the graph G^*.

Let us denote by $\mathcal{S}'(s,m,b)$ the collection of all ordered pairs

- $S = S_1 \cup \ldots \cup S_m \in \mathcal{S}(s,m,b)$, where $S_1, \ldots, S_m$ are connected components of $G[S]$,

- $m-1$ pairs of vertices of S whose addition to G makes $G[S]$ connected.

A moment of thought reveals that for fixed s, the probability that (3) does not hold is bounded by

$$\sum_{m=1}^{\delta s}\sum_{b=m}^{\delta s} |\mathcal{S}'(s,m,b)| \cdot (\varepsilon/n)^{m-1} \cdot \Pr\left(\mathrm{Bin}(s(n-s),\varepsilon/n) \leqslant \delta s\right). \tag{4}$$

Therefore, it suffices to prove the following.

Claim. *There exists an absolute constant C such that for all s, m, and b with $m \leqslant b \leqslant \delta s$,*

$$|\mathcal{S}'(s,m,b)| \leqslant n^m \exp(C\delta \log(1/\delta)s).$$

Indeed, if $K \log n \leqslant s \leqslant \alpha n$, then by Chernoff's inequality,

$$\Pr\left(\mathrm{Bin}(s(n-s),\varepsilon/n) < \delta s\right) \leqslant \Pr\left(\mathrm{Bin}((1-\alpha)sn,\varepsilon/n) < \delta s\right)$$
$$\leqslant \exp\left(-(1-\alpha)\varepsilon s/8\right),$$

provided that $\delta < (1-\alpha)\varepsilon/2$. Hence, (4) is bounded from above by

$$s^2 n \exp\left[\left(C\delta \log(1/\delta) - \frac{(1-\alpha)\varepsilon}{8}\right)s\right].$$

If we choose K and δ as in the proof of Theorem 2, a union bound over $K \log n < s \leqslant \alpha n$ yields that (4) is indeed $o(1)$.

Hence, it suffices to prove the claim. To this end, we will argue that each element of $\mathcal{S}(s,m,b)$ can be uniquely described by the following:

(i) a set $W = \{v_1, \ldots, v_m\}$ of vertices of G,

(ii) a partition $s = s_1 + \ldots + s_m$, where $s_i \geqslant 1$ for each i,

(iii) a partition $b = b_1 + \ldots + b_m$, where $b_i \geqslant 1$ for each i,

(iv) a set S_i in $\mathcal{C}(v_i, s_i, b_i)$ for each $i \in [m]$,

(v) a partition $m - 1 = d_1 + \ldots + d_m$, where $d_i \geqslant 0$ for each i,

(vi) a multiset D_i of d_i elements from S_i for each $i \in [m]$,

(vii) a permutation $f \colon [m-1] \to [m-1]$.

Assuming that this is indeed the case, by Proposition 7 we have

$$|\mathcal{S}'(s,m,b)| \leqslant \binom{n}{m}(m-1)! \sum_{(s_i),(b_i),(d_i)} \prod_{i=1}^{m}\left[\binom{s_i+b_i-1}{b_i}\binom{s_i+d_i-1}{d_i}\right]$$
$$\leqslant \frac{n^m}{m}\binom{s-1}{m-1}^2\binom{b-1}{m-1}\binom{2m-2}{m-1}\binom{s+b-m}{b} \leqslant (2n)^m\binom{s}{m}^2\binom{b}{m}\binom{s+b}{b}.$$

11

Consequently, if $m \leqslant b \leqslant \delta s$, then

$$|\mathcal{S}'(s,m,b)| \leqslant n^m \left(\frac{2e^4 s^2 (s+b)}{b^3} \right)^b \leqslant n^m \left(\frac{3e^4}{\delta^3} \right)^{\delta s} \leqslant n^m \exp\left(C\delta \log(1/\delta) s \right),$$

where C is some absolute constant.

Finally, we show that each $S \in \mathcal{S}'(s,m,b)$ may be uniquely described by (i)–(vii). First, observe that (i)–(iv) uniquely describe the set $S = S_1 \cup \ldots \cup S_m$, together with a root vertex v_i in each connected component S_i, whose use will be explained later. As in the proof of Theorem 2, one may assume some canonical linear ordering $\preceq$ on the set of vertices of G. Given this ordering, one may canonically order the sets $S_1, \ldots, S_m$ according to the canonical ordering $\preceq$ on the set $\{\min_{\preceq} S_1, \ldots, \min_{\preceq} S_m\}$ of representatives of each S_i. Now, note that the $m-1$ pairs of vertices of S whose addition to G makes $G[S]$ connected naturally define a tree T on the vertex set $\{S_1, \ldots, S_m\}$. Root this tree at S_m and orient all of its edges away from the root. Now, start with $v_m \in S_m$ and for each $i \in [m-1]$ let $v_i \in S_i$ be the unique vertex of S_i that lies in the pair of vertices of S that corresponds to the unique edge of T going into S_i. Next, for each $i \in [m]$, let d_i be the outdegree of S_i in T and let D_i be the multiset of d_i vertices of S_i that lie in the pairs of vertices of S that correspond to the d_i edges of T going out of S_i. Finally, let $D = D_1 \cup \ldots \cup D_m$ and observe that the $m-1$ pairs of vertices of S that correspond to the edges of T define a bijection between D and $\{v_1, \ldots, v_{m-1}\}$. Namely, if $D = \{w_1, \ldots, w_{m-1}\}$, where $w_1 \preceq \ldots \preceq w_{m-1}$, then this bijection can be described by a permutation $f \colon [m-1] \to [m-1]$ defined by letting $f(i)$ be the unique j such that $\{v_i, w_j\}$ is one of the $m-1$ pairs of vertices whose addition to G makes $G[S]$ connected. This concludes the proof of the theorem. $\qquad\square$

5 Diameter

In this section, we prove Theorem 4. Since adding edges to a graph can only decrease its diameter, it suffices to consider the case when G is a tree and $\varepsilon \leqslant 1/3$. Since $e(G) = n - 1$, it follows from Chernoff's inequality (Lemma 9) that a.a.s. G^* has at most $(1+\varepsilon)n$ edges. Hence, it is enough to prove that there is a constant $C = C(\varepsilon)$ such that a.a.s.

$$e\big(B(v, C\log n)\big) > \frac{(1+\varepsilon)n}{2} \qquad \text{for every } v \in V(G), \tag{5}$$

where $B(v,r)$ denotes the G^*-ball of radius r around v. Indeed, (5) implies that for every $u, v \in V(G)$, we have that $B(u, C\log n) \cap B(v, C\log n) \neq \emptyset$, and consequently $\operatorname{diam}(G^*) \leqslant 2C\log n$.

Fix some $v \in V(G)$, let K and δ be as in Theorem 3 with $\alpha = 3/4$, and condition on the event that G^* satisfies the assertion of this theorem. Moreover, condition on the event that R satisfies the assertion of Lemma 8 with $C = 1$. This implies that $e(S)/3 \leqslant |S| \leqslant e(S) + 1$ for every connected set S in G^*. Since G^* is connected, we clearly have that $|B(v,r)| \geqslant r + 1$. Hence, if $r \geqslant K\log n$, we have that

$$e(B(v, r+1)) \geqslant \min\left\{ \frac{3n}{4} - 1, \left(1 + \frac{\delta}{3}\right) e\big(B(v,r)\big) \right\}.$$

Letting $C = K + \frac{1}{\log(1+\delta/3)}$, we have that

$$e(B(v, C \log n)) \geqslant \frac{3n}{4} - 1 > \frac{2n}{3} \geqslant \frac{(1+\varepsilon)n}{2},$$

as claimed. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

6   Mixing time

In this section, we prove Theorem 5. Let ε be a positive real, let D be a positive integer, and assume that G is a connected D-degenerate graph. Let $G^* = G \cup R$, where $R \sim G(n, \varepsilon/n)$.

Our argument for bounding the mixing time is based on the approach of Fountoulakis and Reed [16, 17]. The main idea there is that one can bound the mixing time of an abstract irreducible, reversible, and aperiodic Markov chain in terms of the *conductances* of *connected* sets of states of various sizes. For simplicity, we only state their results in the setting of the lazy random walk on the graph G^*. For $S \subseteq V$, let $\pi(S)$ equal $\sum_{v \in S} \pi(v)$. It can be verified that $\pi(S) = \frac{2e_{G^*}(S) + |\partial_G S|}{2e(G^*)}$. We define

$$Q(S) = \sum_{u \in S, v \notin S} \pi(u) \Pr(u \to v) = \frac{|\partial_{G^*} S|}{4e(G^*)}$$

and note that $Q(S) = Q(S^c)$. The *conductance* $\Phi(S)$ of S is

$$\Phi(S) = \frac{Q(S)}{\pi(S)\pi(S^c)} = \frac{|\partial_{G^*} S|}{2 \cdot (2e_{G^*}(S) + |\partial_{G^*} S|) \cdot \pi(S^c)}.$$

Let $\pi_{\min} = \min_{v \in V(G)} \pi(v)$. For $p > \pi_{\min}$, we denote by $\Phi(p)$ the minimum conductance of a connected (in G^*) set S with $p/2 \leqslant \pi(S) \leqslant p$ (if there is no such S, we define $\Phi(p) = 1$). Fountoulakis and Reed [16] proved the following result.

Theorem 11. *There exists an absolute constant C such that*

$$T_{\mathrm{mix}}(G^*) \leqslant C \sum_{j=1}^{\lceil \log_2 \pi_{\min}^{-1} \rceil} \Phi^{-2}(2^{-j}).$$

In the remainder of the proof, we will estimate the sum in Theorem 11. We claim that it is enough to prove the following.

Lemma 12. *There exist positive constants δ^* and K^* such that a.a.s. for every connected (in G^*) set S with $\frac{K^* \log n}{n} \leqslant \pi(S) \leqslant 1/2$,*

$$\Phi(S) \geqslant \delta^*.$$

Indeed, suppose that the assertion of Lemma 12 holds for some δ^* and K^*. Let J be the set of indices j satisfying $2^{-j} \leqslant \frac{2K^* \log n}{n}$ and note the $|J^c| < \log_2 n$, as $2^{-j} > \frac{2K^* \log n}{n}$ implies that $j < \log_2 n$. Since G^* is connected, we have that for every set S,

$$\Phi(S) \geqslant \frac{|\partial_{G^*} S|}{4e(G^*) \cdot \pi(S)} \geqslant \frac{1}{4e(G^*) \cdot \pi(S)}.$$

Condition on the event that R satisfies the assertion of Lemma 8 with $C = \max\{\varepsilon, 1\}$. Let $D^* = D + 2C$ and observe that the degeneracy assumption implies that

$$e_{G^*}(S) \leqslant D^*|S| \quad \text{for every } S \subseteq V(G). \tag{6}$$

In particular, $e(G^*) \leqslant D^*n$ and hence, letting $M = 129(K^*)^2(D^*)^2$,

$$\sum_{j=1}^{\lceil \log_2 \pi_{\min}^{-1} \rceil} \Phi^{-2}(2^{-j}) \leqslant |J^c| \cdot (\delta^*)^{-2} + \sum_{j \in J} 2^{-2j}(4e(G^*))^2$$

$$\leqslant O(\log n) + 2 \cdot \max_{j \in J}\{2^{-2j}\} \cdot 16(D^*)^2 n^2 \leqslant M \log^2 n,$$

provided that n is sufficiently large, where we used the definition of J and the inequality $\sum_{j \geqslant i} 2^{-2j} \leqslant 2^{-2i+1}$.

Therefore, it suffices to prove Lemma 12. We first show that any connected set S with $\pi(S) \leqslant 1/2$ has at most $n - \Omega(n)$ elements.

Claim. *Every connected (in G^*) set $S \subseteq V(G)$ with $\pi(S) \leqslant 1/2$ satisfies*

$$|S| \leqslant \frac{D^*n + 1}{D^* + 1}.$$

Proof. Since $\pi(S) \leqslant 1/2$ implies that $\pi(S) \leqslant \pi(S^c)$, we have

$$2e_{G^*}(S) = 2e(G^*)\pi(S) - |\partial_{G^*}S| \leqslant 2e(G^*)\pi(S^c) - |\partial_{G^*}S|$$
$$= 2e_{G^*}(S^c) \leqslant 2D^*|S^c| = 2D^*(n - |S|).$$

Since S is connected in G^*, we obtain $e_{G^*}(S) \geqslant |S| - 1$ and the claim follows. $\qquad\square$

Let δ and K be as in Theorem 3 with $\alpha = \frac{D^*}{D^*+1}$ and condition on the event that G^* satisfies the assertion of this theorem. Let

$$K^* = (2D^* + 1)K \qquad \text{and let} \qquad \delta^* = \min\left\{\frac{1}{2D^* + 2}, \frac{\delta}{2D^* + 2\delta}\right\}.$$

It follows from (6) that for every connected set S with $\pi(S) \leqslant 1/2$, we have

$$\Phi(S) \geqslant \frac{|\partial_{G^*}S|}{2 \cdot (e_{G^*}(S) + |\partial_{G^*}S|)} \geqslant \frac{|\partial_{G^*}S|}{2D^*|S| + 2|\partial_{G^*}S|}.$$

Note that if $|\partial_{G^*}S| \geqslant |S|$, then $\Phi(S) \geqslant \delta^*$, so we may assume otherwise. In particular, if $\pi(S) \geqslant K^* \log n/n$, then

$$K^* \log n \leqslant 2\pi(S)e(G^*) = 2e_{G^*}(S) + |\partial_{G^*}S| \leqslant (2D^* + 1)|S|$$

and hence $|\partial_{G^*}S| \geqslant \delta|S|$. It follows that $\Phi(S) \geqslant \delta^*$. This concludes the proof of Lemma 12 and therefore the proof of Theorem 5. $\qquad\square$

7 Long Paths

Here we show that after adding random edges, each with probability $\frac{\epsilon}{n}$, to a connected n-vertex graph with bounded maximum degree, we a.a.s. get a path whose length is linear in n.

Proof of Theorem 6. Let k be a sufficiently large constant. Similarly as in the proof of Theorem 1, let us chop the vertex set of the graph G into connected pieces $V_1, \ldots, V_t$ such that for each $1 \leqslant i \leqslant t$, we have $k \leqslant |V_i| \leqslant \Delta k$. As in Theorem 1, the probability that two blobs are connected (in G^*) is at least $\frac{\varepsilon k^2}{2n}$. Hence, if k is sufficiently large, then the auxiliary graph naturally induced by the blobs (obtained by treating each blob as a super-vertex and connecting two super-vertices if there is an edge of G^* connecting the two blobs) contains the random graph $G(t, C/t)$, where $C \to \infty$ as $k \to \infty$. It is well known ([2], see also [20]) that if $C > 1$, then $G(t, C/t)$ a.a.s. contains a path P_0 of length $\Omega(t)$. Since the blobs are connected and $t \geqslant n/(\Delta k)$, one can turn P into a path in G^*, whose length is at least as large as the length of P. Indeed, we may use the edges of P to move between the blobs and the edges of G to connect the entry and the exit points of P within each blob traversed by P. □

8 Concluding remarks

In this paper, we studied the model of randomly perturbed connected graphs. We presented two different approaches to analyzing the expansion properties of such graphs and obtained lower bounds for both the vertex and the edge expansions under mild assumptions on the base graph. One of the approaches is based on the idea of decomposing a connected graph with bounded maximum degree into connected subgraphs of comparable size, the other approach builds on a new general upper bound on the number of connected subsets with small vertex boundary. Using these results, we established several other interesting properties of randomly perturbed connected graphs: bounds on the diameter and the mixing time of the lazy random walk, and the containment of long paths. It would be interesting to study other parameters of this model.

It seems that randomly perturbed connected n-vertex graph with bounded degeneracy shares some similarities with the giant component in the supercritical Erdős-Rényi random graph $G(n, \frac{1+\varepsilon}{n})$. In particular, a.a.s. they both have diameter $O(\log n)$, mixing time $O(\log^2 n)$, and contain paths of length $\Omega(n)$. It could be interesting to explore this analogy further and to check whether the methods used in this work to study the model of randomly perturbed graphs can be applied to the other model.

References

[1] L. Addario-Berry and T. Lei, *The mixing time of the Newman–Watts small world.*, Proceedings of the 23rd Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '12, 2012, pp. 1661–1668.

[2] M. Ajtai, J. Komlós, and E. Szemerédi, *The longest path in a random graph*, Combinatorica **1** (1981), 1–12.

[3] I. Benjamini, G. Kozma, and N. Wormald, *The mixing time of the giant component of a random graph*, arXiv:math/0610459 [math.PR].

[4] I. Benjamini and E. Mossel, *On the mixing time of a simple random walk on the super critical percolation cluster*, Probab. Theory Related Fields **125** (2003), 408–420.

[5] A. Blum and J. Dunagan, *Smoothed analysis of the perceptron algorithm for linear programming*, Proceedings of the 13th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '02, 2002, pp. 905–914.

[6] A. Blum and J. Spencer, *Coloring random and semi-random k-colorable graphs*, J. Algorithms **19** (1995), 204–234.

[7] F. Bohman, A. Frieze, and R. Martin, *How many random edges make a dense graph hamiltonian?*, Random Structures Algorithms **22** (2003), 33–42.

[8] B. Bollobás and F. R. Chung, *The diameter of a cycle plus a random matching.*, SIAM J. Discrete Math. **1** (1988), 328–333.

[9] A. Coja-Oghlan, U. Feige, A. Frieze, M. Krivelevich, and D. Vilenchik, *On Smoothed k-CNF Formulas and the Walksat Algorithm*, Proceedings of the 20th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '09, 2009, pp. 451–460.

[10] J. Ding and Y. Peres, *Sensitivity of mixing times*, arXiv:1304.0244 [math.PR].

[11] R. Durrett, *Random graph dynamics*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, 2010.

[12] U. Feige, *Refuting Smoothed 3CNF Formulas*, Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science, FOCS' 07, 2007, pp. 407–417.

[13] U. Feige and J. Kilian, *Heuristics for semirandom graph problems*, J. Comput. System Sci. **63** (2001), 639–671.

[14] A. Flaxman and A. Frieze, *The diameter of randomly perturbed digraphs and some applications*, Random Structures Algorithms **30** (2007), 484–504.

[15] A. D. Flaxman, *Expansion and lack thereof in randomly perturbed graphs*, Internet Math. **4** (2007), 131–147.

[16] N. Fountoulakis and B. A. Reed, *Faster mixing and small bottlenecks*, Probab. Theory Related Fields **137** (2007), 475–486.

[17] _____, *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.

[18] M. Jerrum and A. Sinclair, *Conductance and the rapid mixing property for markov chains: the approximation of the permanent resolved*, Proceedings of the 20th Annual ACM Symposium on Theory of Computing, STOC '88, 1988, pp. 235–244.

[19] M. Krivelevich and A. Nachmias, *Coloring complete bipartite graphs from random lists*, Random Structures Algorithms **29** (2006), 436–449.

[20] M. Krivelevich and B. Sudakov, *The phase transition in random graphs – a simple proof*, Random Structures Algorithms **43** (2013), 1–15.

[21] M. Krivelevich, B. Sudakov, and P. Tetali, *On smoothed analysis in dense graphs and formulas*, Random Structures Algorithms **29** (2006), 180–193.

[22] D. A. Levin, Y. Peres, and E. L. Wilmer, *Markov chains and mixing times*, American Mathematical Society, Providence, RI, 2009.

[23] M. Mitzenmacher and E. Upfal, *Probability and Computing, randomized algorithms and probabilistic analysis*, Cambridge University Press, Cambridge, 2005.

[24] M. E. J. Newman and D. J. Watts, *Renormalization group analysis of the small-world network model*, Phys. Lett. A **263** (1999), 341–346.

[25] _____, *Scaling and percolation in the small-world network model*, Phys. Rev. E (3) **60** (1999), 7332–7342.

[26] L. Pósa, *Hamiltonian circuits in random graphs*, Discrete Math. **14** (1976), 359–364.

[27] A. Sankar, D. Spielman, and S. H. Teng, *Smoothed analysis of the condition numbers and growth factors of matrices*, SIAM J. Matrix Anal. Appl. **28** (2006), 446–476.

[28] J. Spencer and G. Tóth, *Crossing numbers of random graphs*, Random Structures Algorithms **21** (2002), 347–358.

[29] D. Spielman and S. H. Teng, *Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time*, J. ACM **51** (2004), 385–463.

[30] B. Sudakov and J. Vondrák, *How many random edges make a dense hypergraph non-2-colorable?*, Random Structures Algorithms **32** (2008), 290–306.

[31] T. Tao and V. Vu, *The condition number of a randomly perturbed matrix*, STOC'07—Proceedings of the 39th Annual ACM Symposium on Theory of Computing, ACM, New York, 2007, pp. 248–255.