Random Walks and Evolving Sets:
Faster Convergences and Limitations

Siu On Chan[*]   Tsz Chiu Kwok[†]   Lap Chi Lau[‡]

Abstract

Analyzing the mixing time of random walks is a well-studied problem with applications in random sampling and more recently in graph partitioning. In this work, we present new analysis of random walks and evolving sets using more combinatorial graph structures, and show some implications in approximating small-set expansion. On the other hand, we provide examples showing the limitations of using random walks and evolving sets in disproving the small-set expansion hypothesis.

1. We define a combinatorial analog of the spectral gap, and use it to prove the convergence of non-lazy random walks. A corollary is a tight lower bound on the small-set expansion of graph powers for any graph.

2. We prove that random walks converge faster when the robust vertex expansion of the graph is larger. This provides an improved analysis of the local graph partitioning algorithm using the evolving set process, and also derives an alternative proof of an improved Cheeger's inequality.

3. We give an example showing that the evolving set process fails to disprove the small-set expansion hypothesis. This refutes a conjecture of Oveis Gharan and shows the limitations of all existing local graph partitioning algorithms in approximating small-set expansion.

1   Introduction

Analyzing the mixing time of random walks is a fundamental problem with many applications in random sampling [LPW08]. The evolving set process is an elegant tool introduced by Morris and Peres [MP05] to provide sharp analyses of mixing time (see the survey [MT06]). Recently, random walks and the evolving set process have also been used in designing local algorithms for graph partitioning [ST13, ACL06, AP09, OT12, KL12].

[*]The Chinese University of Hong Kong, sochan@gmail.com.

[†]University of Waterloo, tckwok@uwaterloo.ca.

[‡]University of Waterloo, lapchi@uwaterloo.ca.

The evolving set process is the most powerful one among the local graph partitioning algorithms, and it was conjectured [Ove13, AOPT16] that the process can be used to disprove the small-set expansion hypothesis [RS10]. The theme of this paper is to study the power and limitations of this technique, from analyzing mixing time to local graph partitioning and approximating small-set expansion.

Random Walks and Mixing Time We consider random walks in a weighted undirected graph $G = (V, E)$ with a nonnegative weight $w(e)$ on each edge $e \in E$. Let $n = |V|$ and $m = |E|$. For simplicity, we assume that the graph G is regular and the weights are scaled such that the weighted degree of i is $\sum_{j:ij \in E} w(i,j) = 1$ for all $i \in V$ throughout this paper, but we will mention how to deal with general graphs in Section 2.5. Let A be the $n \times n$ symmetric random walk matrix of G with $A_{ij} = w(i,j)$. Let $p_0 \in \mathbb{R}^n$ be an initial probability distribution, and let $p_t := A^t p_0$ be the probability distribution after t steps of random walks. When G is connected and non-bipartite, it is well-known that p_t will converge to the uniform distribution. The mixing time is defined as

$$T_{\mathrm{mix}} = \min\{t : \|p_t - \frac{\vec{1}}{n}\|_1 \leqslant \frac{1}{2} \text{ for all initial } p_0\}.$$

One approach to analyze the mixing time is to look at the eigenvalues of the random walk matrix. Let the eigenvalues of A be $1 = \lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_n \geqslant -1$. By basic spectral graph theory, it can be shown that $1 > \lambda_2$ if and only if G is connected, and $\lambda_n > -1$ if and only if G is non-bipartite. This implies that, when G is connected and non-bipartite, p_t will converge to the first eigenvector, and thus the uniform distribution is the unique limiting distribution of the random walk. Let $\lambda := \max\{\lambda_2, |\lambda_n|\}$, and let $1 - \lambda$ be the spectral gap of the random walk matrix. A standard calculation shows that the mixing time is upper bounded by $O(\log(n)/(1 - \lambda))$.

For many problems, it is useful to have combinatorial characterizations of graphs with fast mixing time. For two subsets $S, T \subseteq V$, let $E(S, T)$ be the set of

edges with one vertex in S and another vertex in T, and let $w(S,T) := \sum_{e \in E(S,T)} w(e)$. The expansion of a set $S \subseteq V$ and the expansion of a graph G are defined as

$$\phi(S) := \frac{w(S, V-S)}{|S|} \quad \text{and} \quad \phi(G) := \min_{S:|S| \leqslant n/2} \phi(S).$$

Cheeger's inequality for graphs [Alo86, AM85] states that

$$\frac{1}{2}(1 - \lambda_2) \leqslant \phi(G) \leqslant \sqrt{2(1 - \lambda_2)},$$

and thus $1 - \lambda_2 = \Omega(\phi(G)^2)$. Having large conductance is not enough to guarantee fast mixing time, as λ_n may be very close to -1. There is a simple trick to bypass this issue: one can guarantee that $\lambda_n \geqslant 0$ by considering "lazy" random walks (with probability $1/2$ stay put), and this implies that the mixing time of lazy random walks is upper bounded by $O(\log(n)/\phi(G)^2)$.

Another approach to analyze the mixing time is to directly use the graph structures. Lovász and Simonovits [LS90] developed a combinatorial method to prove that the mixing time of lazy random walks is $O(\log(n)/\phi(G)^2)$. This method is more flexible in incorporating additional graph structures. Given a parameter $0 < \delta \leqslant 1/2$, the δ-small-set expansion is defined as

$$\phi_\delta(G) := \min_{S:|S| \leqslant \delta n} \phi(S).$$

Lovász and Kannan [LK99] proved that the mixing time of lazy random walks is

$$T_{\mathrm{mix}} \leqslant \int_{1/n}^{1/2} \frac{dx}{x \phi_x(G)^2}.$$

As we will discuss in more details shortly, this combinatorial approach can also be used to design local graph partitioning algorithms for approximating small-set expansion.

Evolving Sets The evolving set process is a Markov chain on subsets of V with the following transition rule: If the current set is S, choose U uniformly from $[0, 1]$ and the next set is defined as

$$\tilde{S} := \{y : w(y, S) \geqslant U\}.$$

Morris and Peres [MP05] used the evolving set process to strengthen Lovász and Kannan result to bound the uniform mixing time of lazy random walks by the expansion profile. An important definition in their analysis is the gauge of a set S and the gauge of a graph G, which are defined as

$$\psi(S) = 1 - \mathbb{E}[\sqrt{|\tilde{S}|/|S|}] \quad \text{and} \quad \psi(G) := \min_{S:|S| \leqslant n/2} \psi(S).$$

Morris and Peres [MP05] showed that the convergence rate of random walks is bounded by the gauge, and the mixing time of random walks is $O(\log(n)/\psi(G))$. They proved that $\psi(G) \geqslant \Omega(\phi^2(G))$ for lazy graphs, and this implies that the mixing time of lazy random walks is upper bounded by $O(\log(n)/\phi(G)^2)$. We refer the interested reader to [LPW08] for an excellent introduction of the evolving set process.

Local Graph Partitioning Algorithms Spielman and Teng [ST13] used random walks to design the first local graph partitioning algorithm, which outputs a set S of approximately optimal expansion with running time depends only on $|S|$ and polylog(n). Their analysis is based on the approach of Lovász and Simonovits [LS90] on analyzing mixing time. Andersen and Peres [AP09] and Oveis Gharan and Trevisan [OT12] used the evolving set process for local graph partitioning, and provide the current best known algorithm in terms of both the approximation ratio and the running time.

THEOREM 1.1. ([AOPT16]) *Given an unweighted d-regular graph, for any target set S^* and any $\epsilon > 0$, there is a subset S' with $|S'| \geqslant |S^*|/2$, such that if we start the evolving set process with $\{v\}$ for $v \in S'$, then with constant probability the algorithm returns a set S with $\phi(S) = O(\sqrt{\phi(S^*)/\epsilon})$ and $|S| = O(d^\epsilon |S^*|^{1+\epsilon})$, and the running time is $O(|dS^*|^{1+2\epsilon} \phi(S)^{-\frac{1}{2}} \log^2 n)$.*

Small Set Expansion The small set expansion hypothesis proposed by Raghavendra and Steurer [RS10] states that for any ϵ, there exists δ such that it is NP-hard to distinguish the following two cases:

1. There is a set S with $\phi(S) \leqslant \epsilon$ and $|S| \leqslant \delta n$;

2. $\phi(S) \geqslant 1 - \epsilon$ for every set S with $|S| \leqslant \delta n$.

This hypothesis is shown to be closely related to (and imply) the unique games conjecture [RS10], and they have been extensively studied in the past decade. The local graph partitioning algorithms provide bicriteria approximation algorithms for computing small set expansion $\phi_\delta(G)$. It is observed in [OT12, KL12] that if the output size guarantee of the above local graph partitioning algorithm is improved from $O(|S^*|^{1+\epsilon})$ to $O(|S^*|)$, then the small-set expansion hypothesis is false. Oveis Gharan [Ove13, AOPT16] suggested a plan to prove such an output size guarantee using the evolving set process. Besides the sum-of-squares approach [BBHKSZ12, BKS14], this is the only promising approach to attack the small-set expansion hypothesis.

1.1 Our Results

Combinatorial Analog of Spectral Gap We define a combinatorial analog of spectral gap with which we can directly analyze the mixing time of non-lazy random walks. Recall that $\phi(G)$ is defined as

$$\min_{S \subseteq V, |S| \leqslant n/2} \frac{w(S, V - S)}{|S|} = \min_{S \subseteq V, |S| \leqslant n/2} 1 - \frac{w(S, S)}{|S|}.$$

We define the combinatorial gap as

$$(1.1) \qquad \varphi(G) := \min_{S \subseteq V, T \subseteq V, |S| = |T| \leqslant n/2} 1 - \frac{w(S, T)}{|S|}.$$

Note that $\varphi(G)$ is small if there exists a near-bipartite component. We prove the following combinatorial analog of the spectral analysis of mixing time.

THEOREM 1.2. *For any graph G, $\psi(G) \geqslant \Omega(\varphi^2(G))$.*

By the aforementioned result of Morris and Peres [MP05], one immediate corollary is that the mixing time of non-lazy random walks is upper bounded by $O(\log(n)/\varphi^2(G))$.

Our proof of Theorem 1.2 is based on a new analysis of the approach by Lovász and Simonovits [LS90] using the bar chart in Figure 2.2. We believe that the new analysis is more intuitive and provides better insights into what combinatorial properties are needed for fast mixing.

Using Theorem 1.2 and the results in [KL14], another corollary is the following lower bound on small-set expansion of graph powers.

COROLLARY 1.1. *For any graph G and any integer $t \geqslant 1$,*
$$\phi_{\delta/4}(G^t) = \Omega(\min\{\sqrt{t} \cdot \phi_\delta(G), 1\}).$$

The same result is proved in [KL14] for lazy graphs, and here we prove it for all graphs. Note that it is not true that $\phi(G^t) = \Omega(\min\{\sqrt{t} \cdot \phi(G), 1\})$ when G is bipartite, but the above corollary shows that it is true for small-set expansion even when G is bipartite. As shown in [RS14], this result can be used to amplify hardness results for the small-set expansion problem.

Vertex Expansion The robust vertex expansion is defined by Kannan, Lovász and Montenegro [KLM06] as follows: For $S \subseteq V$, let $N_{1/2}(S) := \min\{|T| \mid T \subseteq V - S \text{ and } w(S, T) \geqslant \frac{1}{2}w(S, V - S)\}$. Define

$$\phi^V(S) := \min\left\{\frac{N_{1/2}(S)}{|S|}, 1\right\},$$

and

$$\phi^V(G) := \min_{S:|S| \leqslant n/2} \phi^V(S)$$

as the robust vertex expansion of a set S and the graph G. Also define

$$\Psi(S) := \phi(S) \cdot \phi^V(S),$$

and

$$\Psi(G) := \min_{S:|S| \leqslant n/2} \Psi(S)$$

as the minimum product of the edge expansion and the robust vertex expansion. It is proved in [KLL16] that

$$1 - \lambda_2 = \Omega(\Psi(G)),$$

and that the spectral partitioning algorithm and the local graph partitioning algorithm using personal pagerank vectors [ACL06] achieve better approximation when the robust vertex expansion is large. We analyze random walks and evolving sets and prove a similar result.

THEOREM 1.3. *For lazy graphs G, $\psi(G) \geqslant \Omega(\Psi(G))$.*

Morris and Peres [MP05] prove that the gauge is a lower bound of the spectral gap $\psi(G) \leqslant 1 - \lambda_2$. An interesting corollary is a random walk based proof of the improved Cheeger's inequality using vertex expansion proved in [KLL16] with better constant.

COROLLARY 1.2. *For lazy graphs G, $1 - \lambda_2 = \Omega(\Psi(G))$.*

Another corollary is an improved analysis of Theorem 1.1 when the robust vertex expansion of G is large.

COROLLARY 1.3. *Given an unweighted d-regular graph, for any target set S^* and any $\epsilon > 0$, there is a subset S' with $|S'| \geqslant |S^*|/2$, such that if we start the evolving set process with $\{v\}$ for $v \in S'$, then with constant probability the algorithm returns a set S with $\Psi(S) = O(\phi(S^*)/\epsilon)$ and $|S| = O(d^\epsilon |S^*|^{1+\epsilon})$, and the running time is $O(|dS^*|^{1+2\epsilon}\phi(S)^{-\frac{1}{2}}\log^2 n)$.*

Note that the conclusion $\Psi(S) = O(\phi(S^*)/\epsilon)$ implies that $\phi(S) = O_\epsilon(\phi(S^*)/\phi^V(S)) = O_\epsilon(\phi(S^*)/\phi^V(G))$. In particular, this implies that the evolving set algorithm is a constant factor approximation algorithm for finding small sparse cuts when $\phi^V(G)$ is a constant. Note that the assumption that the robust vertex expansion is large is satisfied by typical average case instances (such as the planted random model). This shows that the evolving set algorithm matches the improved analysis of the spectral partitioning algorithm in [KLL16], partially explaining the success of random walk based algorithms on non-worst-case instances. We refer the reader to [KLL16] for more discussions and motivations for robust vertex expansion.

Limitations The subexponential time algorithm for small-set expansion by Arora, Barak and Steurer [ABS10] uses eigenspace enumeration and random walks. The short code example in [BGHMRS12] shows the limitation of the eigenspace enumeration method. It is a natural question to ask whether random walks can be used to disprove the small-set expansion hypothesis. There were very few results showing the limitations of these random walks based algorithms (see [ZLM13] for the only such result that we know of). One main difference between the truncated random walk algorithm by Spielman and Teng [ST13] and the evolving set algorithm by Andersen and Peres [AP09] and Oveis Gharan and Trevisan [OT12] is that the random walk algorithm is deterministic while the evolving set algorithm involves much randomness. It is conjectured in [Ove13, AOPT16] that there is a small but nontrivial probability that all the sets explored by the evolving set process is of size $O(|S^*|)$, and argued that this would disprove the small-set expansion hypothesis. We present an example for which the evolving set algorithm fails with probability one, refuting the conjecture.

THEOREM 1.4. *Given any ϵ, there exists δ_ϵ such that for any $\delta > 0$, there is a graph G such that $\phi_\delta(G) \leqslant \epsilon$, but any subset of volume $\leqslant \delta_\epsilon n$ returned by the evolving set algorithm in [AP09, OT12] has expansion at least $1 - \epsilon$ with probability one.*

We remark that there are two difference parameters δ_ϵ and δ to emphasis the fact that the evolving set algorithm cannot find a set with small expansion even it looks at sets with larger volume.

The example is a k-ary ϵ-noisy hypercube, where the dimension cuts are of size n/k with expansion ϵ. We show that, however, the evolving set algorithm will only explore the Hamming balls, and the expansion is at least $1 - \epsilon$ for all Hamming balls of size $O(n/k)$. We note that this example also shows that the random walk algorithm [ST13, KL12] and the pagerank algorithm [ACL06, ZLM13] fail to disprove the small-set expansion hypothesis; see Section 3.

In addition to being the first lower bound to the evolving set algorithm, this example exposes the limitations of all known local graph partitioning algorithms. We believe that this example can be used as a basis to prove further lower bounds. An open question is whether the analysis of the $O(\sqrt{\phi(S)\log(|S|)})$-approximation of the evolving set algorithm in Theorem 1.1 is tight when $\epsilon = 1/\log(|S|)$.

1.2 Relations with Previous Work

Combinatorial Analog of Spectral Gap We note that the original analyses of Lovász and Simonovits [LS90] and Andersen and Peres [AP09] heavily rely on the laziness assumption and cannot be used to work with the combinatorial gap. The bar chart in Figure 2.2 is the new element introduced to analyze the combinatorial gap as well as the robust vertex expansion.

Trevisan [Tre12] defined the bipartiteness ratio $\beta(G)$ of a graph and proved that $\lambda_n + 1$ is related to $\beta(G)$ as if $1 - \lambda_2$ is related to $\phi(G)$ stated by Cheeger's inequality. After formulated and proved Theorem 1.2, we observe that Trevisan's result combined with the spectral argument can also be used to derive the corollary that the mixing time of non-lazy random walks is bounded by $O(\log(n)/\varphi(G)^2)$. However, we remark that Theorem 1.2 and Corollary 1.1 cannot be derived from Trevisan's result and the spectral approach, and also that the formulation of Theorem 1.2 is new.

Mixing Time and Local Graph Partitioning The results in Lovász and Kannan [LK99] and Kannan, Lovász and Montenegro [KLM06] show that the mixing time of lazy random walks is $O(\log(n)/\Psi(G))$, among other conditions that imply faster mixing. However, their results cannot be applied to analyze local graph partitioning algorithms as in Corollary 1.3.

Besides the random walk algorithm [ST13, KL12] and the evolving set algorithm [AP09, OT12], there is also a local graph partitioning algorithm using pagerank vectors [ACL06, ZLM13]. In terms of the approximation guarantee, the output size, and the running time, the pagerank algorithm is subsumed by the evolving set algorithm in [AP09, OT12].

In [KLL16], it was shown that the pagerank algorithm performs better when the robust vertex expansion is large. Similar results were not known for random walks and evolving sets, as the spectral techniques in [KLL16] are not applicable. These results are proved in this paper by a new analysis of the combinatorial approach of Lovász and Simonovits [LS90]. We remark that both the results and the techniques of this paper are different from [KLL16], especially the combinatorial analog of spectral gap, the counterexample for the evolving set algorithm, and the new analysis of Lovász and Simonovits approach using the barchart.

2 Faster Convergence

In this section, we prove the positive results about faster convergence rates of random walks and evolving sets. Our proofs are based on the combinatorial method of Lovász and Simonovits [LS90], and we will begin with an introduction of their techniques in Section 2.1, and

then we will discuss the proof outline and highlight the new idea in Section 2.2. Then, we will prove Theorem 1.2 about combinatorial analog of spectral gap in Section 2.3 and then prove Corollary 1.1 about small-set expansion of graph powers. Then, we will extend the technique to prove Theorem 1.3 about robust vertex expansion in Section 2.4 and then show its application in local graph partitioning. Finally, we will discuss how to extend the results to non-regular graphs in Section 2.5.

2.1 Lovász-Simonovits Curve For any vector $p \in \mathbb{R}^n$, Lovász and Simonovits [LS90] study the curve $C(p) : [0, n] \to \mathbb{R}$ that plots the cumulative sum of p defined as

$$(2.1) \qquad C(p, x) = \max_{c \in [0,1]^n : \sum_i c(i) = x} \sum_{i \in V} c(i) \cdot p(i).$$

In words, $C(p, x)$ is just the sum of the first x largest elements in p when x is a positive integer, and the curve $C(p, x)$ is defined for all $x \in [0, n]$ by piecewise linear extension. It is clear from the definition that $C(p, x)$ is a concave function of x. We are interested in studying the curve $C(A^t p, x)$ where A is a random walk matrix and p is a probability distribution. Notice that as $A^t p$ converges to the uniform distribution as t becomes larger, $C(A^t p)$ converges to the line x/n and vice versa. In [LS90], their method to bound the mixing time is to bound the difference between $C(A^t p)$ and the line x/n. When A is the lazy random walk matrix, they proved that

$$(2.2) \qquad C(A^t p, x) \leqslant \frac{x}{n} + \sqrt{x}(1 - \frac{\phi(G)^2}{8})^t,$$

and they show that this implies the mixing time of lazy random walks is $O(\log n / \phi(G)^2)$. The key lemma in their proof is the following inequality: For any lazy random walk matrix A, any $p \in \mathbb{R}^n$ and any integral x,

$$(2.3)$$
$$C(Ap, x) \leqslant \frac{1}{2}\big(C(p, x(1 - \phi(G))) + C(p, x(1 + \phi(G)))\big).$$

The bound in (2.2) follows from an inductive argument using (2.3); see [LS90, ST13, KL12] and also a slightly more general version in Lemma 2.7 in Section 2.4. We remark that their proof of (2.3) crucially relies on the assumption that there is a self-loop of weight $1/2$ on each vertex and is a bit magical.

2.2 Proof Outline Our main algorithmic result is Theorem 1.3, but its proof develops on the technique for Theorem 1.2. So, we begin with the outline of the proof of Theorem 1.2 in this subsection, and mention
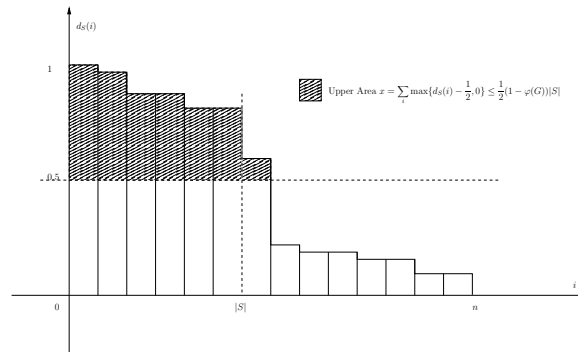

Figure 2.1: An illustration of the proof ideas.

the modifications needed to prove Theorem 1.3 at the end.

We will prove the following inequality similar to (2.3) using the combinatorial gap (without the laziness assumption that $A_{ii} \geqslant 1/2$ for all $i \in V$).

LEMMA 2.1. *For any random walk matrix A, any $p \in \mathbb{R}^n$ and any integral $x \leqslant n/2$,*

$$C(Ap, x) \leqslant \frac{1}{2}\big(C(p, x(1 - \varphi(G))) + C(p, x(1 + \varphi(G)))\big).$$

With Lemma 2.1 in place of (2.3), the same inductive argument that we mentioned before implies the convergence result in (2.2) with $\phi(G)$ replaced by $\varphi(G)$. It turns out that the analysis of the Lovász-Simonovits curve can be used to analyze the evolving set process as well, and the arguments in Lemma 2.1 can be adapted to prove Theorem 1.2. To prove Lemma 2.1, we consider an arbitrary $S \subseteq V$ and try to bound the total probability in S after one step of random walk $(Ap)(S) := \sum_{i \in S}(Ap)_i$. To bound $(Ap)(S)$, we look at where the probability in S is coming from. For each $i \in V$, let

$$d_S(i) := w(i, S)$$

be the total weight coming from i to S. Recall that we assume the weighted degree of each vertex is one. So, we have $d_S(i) \in [0, 1]$ for any i and

$$(2.4) \qquad \sum_{i \in V} d_S(i) = |S|.$$

The reason for this definition is that $(Ap)(S) = \sum_{i=1}^n d_S(i) \cdot p(i)$. We sort the vertices so that $d_S(1) \geqslant d_S(2) \geqslant \ldots \geqslant d_S(n)$. Let $T := \{1, 2, \ldots, |S|\}$ be the $|S|$ vertices with largest d_S values; note that T is in general not equal to S. See Figure 2.2 for an illustration of the proof setup.

The obstruction for mixing is when $d_S(1) = d_S(2) = \ldots = d_S(|S|) = 1$ and $p(1), \ldots, p(|S|)$ are the highest probability in p, in which case we would have $C(Ap, |S|) \geqslant (Ap)(S) = \sum_{i \in S} p(i) = C(p, |S|)$ and thus the curve is not dropping after one step of random walk. This could happen when $S = T$ in which case S is a disconnected component of G (corresponding to $\lambda_2 = 1$), or when S and T form a bipartite component (corresponding to $\lambda_n = -1$), and in these cases the random walk may not mix (depending on the initial distribution).

The combinatorial gap in (1.1) is defined precisely to exclude the obstruction. It states that any subset T with $|T| = |S|$ can only contribute $w(S, T) \leqslant (1 - \varphi(G))|S|$ to $w(S, V) = |S|$, so as to guarantee that the curve would drop, i.e. $C(Ap, |S|) = \max_{S':|S'|=|S|} Ap(S') < C(p, |S|)$. To prove Lemma 2.1, we look at the bar chart in Figure 2.2 horizontally and consider the telescoping sum

$$
\begin{aligned}
(Ap)(S) &= \sum_{i=1}^{n} d_S(i) \cdot p(i) \\
&= \sum_{i=1}^{n} (d_S(i) - d_S(i+1)) \sum_{j=1}^{i} p(j) \\
&\leqslant \sum_{i=1}^{n} (d_S(i) - d_S(i+1))) \cdot C(p, i).
\end{aligned}
$$

Suppose we put a threshold $1/2$ and consider the upper area $x = \sum_{i=1}^{n} \max\{d_S(i) - \frac{1}{2}, 0\}$ and the lower area $y = \sum_{i=1}^{n} \min\{d_S(i), \frac{1}{2}\}$; see Figure 2.2. By concavity of the curve C, we will prove in Lemma 2.2 that

$$
\begin{aligned}
(Ap)(S) &\leqslant \frac{1}{2}(C(p, 2x) + C(p, 2y)) \\
&= \frac{1}{2}(C(p, 2x) + C(p, 2|S| - 2x)).
\end{aligned}
$$

The definition of combinatorial gap in (1.1) forces d_S to spread out, and we will prove in Lemma 2.3 that it implies that

$$
x \leqslant \frac{1}{2}(1 - \varphi(G))|S|.
$$

Combining these two steps gives Lemma 2.1. Once we prove Lemma 2.1, similar calculations can be used to prove Theorem 1.2 about the gauge in the evolving set process.

Vertex Expansion The proof of Theorem 1.3 is also based on the idea of using the bar chart in Figure 2.2. Intuitively, larger vertex expansion will lead to faster mixing, as the probability coming into a set S is from many different vertices, and so it cannot be the case that all probability in S come from a small number of

vertices with high probability. In fact, the definition of $N_{1/2}(S)$ can be rewritten as

$$
\begin{aligned}
N_{1/2}(S) = \min\{x \mid{}& C(d_S, |S| + x) - C(d_S, |S|) \\
&\geqslant \frac{1}{2} w(S, V - S)\}.
\end{aligned}
$$

This allows us to use robust vertex expansion to put an upper bound of the area of the largest vertices; see Lemma 2.5:

$$
C(d_S, (1 + \phi^V(S))|S|) \leqslant (1 - \frac{\phi(S)}{2})|S|.
$$

Because of the asymmetry of the area upper bound, we have to choose a different threshold (instead of $1/2$) to prove in Lemma 2.6 that if $C(d_S, a|S|) \leqslant b|S|$ for any S with $|S| \leqslant n/2$ then

$$
C(Ap, |S|) \leqslant \left(\frac{a-b}{a-b^2}\right) C(p, b|S|) + \left(\frac{b-b^2}{a-b^2}\right) C(p, \frac{a|S|}{b}),
$$

which is the most technical proof in the algorithmic results. Then we will modify the inductive argument accordingly to establish a similar bound as in (2.2). The proof can be adapted to analyze the gauge in the evolving set process to prove Theorem 1.3.

2.3 Combinatorial Analog of Spectral Gap As described in the outline, we will prove the following two lemmas.

LEMMA 2.2. *For any subset $S \subseteq V$ with $|S| \leqslant n/2$, let $x = \sum_{i \in V} \max\{d_S(i) - \frac{1}{2}, 0\}$, then*

$$
(Ap)(S) \leqslant \frac{1}{2}(C(p, 2x) + C(p, 2(|S| - x))),
$$

for any random walk matrix A and any vector $p \in \mathbb{R}^n$.

LEMMA 2.3. *For any subset $S \subseteq V$ with $|S| \leqslant n/2$, let $x = \sum_{i \in V} \max\{d_S(i) - \frac{1}{2}, 0\}$, then*

$$
x \leqslant \frac{1}{2}(1 - \varphi(G))|S|.
$$

First, we assume the lemmas are correct and derive Lemma 2.1.

Proof. Let S be a subset of V. By Lemma 2.2, Lemma 2.3 and the concavity of C, we have

$$
\begin{aligned}
(Ap)(S) &\leqslant \frac{1}{2}(C(p, 2x) + C(p, 2(|S| - x))) \\
&\leqslant \frac{1}{2}(C(p, (1 - \varphi(G))|S|) \\
&\qquad + C(p, (1 + \varphi(G))|S|)),
\end{aligned}
$$

where the last inequality holds since for any concave function f and any values $a \leqslant b \leqslant c \leqslant d$ with $a + d = b + c$, we have $f(a) + f(d) \leqslant f(b) + f(c)$. Since the argument applies to all subset $S \subseteq V$ with $|S| \leqslant n/2$, we have

$$C(Ap, |S|) = \max_{S' \subseteq V, |S'| = |S|} (Ap)(S')$$
$$\leqslant \frac{1}{2} \big(C(p, (1 - \varphi(G))|S|) + C(p, (1 + \varphi(G))|S|) \big).$$

Next we prove Lemma 2.2, which follows from the concavity of the curve C.

Proof. Recall that $(Ap)(S) = \sum_i p(i) \cdot d_S(i)$. For convenience, we define the boundary values to be $d_S(0) = 1$ and $d_S(n + 1) = 0$, and also define $p(0) = 0$. Then, there is an index k such that $d_S(k) > 1/2 \geqslant d_S(k + 1)$. By looking at the bar chart in Figure 2.2 horizontally as in the outline and considering the telescoping sum, we have

$$(Ap)(S) = \sum_{i=0}^{n} (d_S(i) - d_S(i+1))(\sum_{j=1}^{i} p(j))$$
$$\leqslant \sum_{i=0}^{n} (d_S(i) - d_S(i+1)) \cdot C(p, i)$$
$$= \bigg(\sum_{i=0}^{k-1} (d_S(i) - d_S(i+1)) \cdot C(p, i)$$
$$+ (d_S(k) - \frac{1}{2}) \cdot C(p, k) \bigg)$$
$$+ \bigg((\frac{1}{2} - d_S(k+1)) \cdot C(p, k)$$
$$+ \sum_{i=k+1}^{n} (d_S(i) - d_S(i+1)) \cdot C(p, i) \bigg).$$

Recall that $C(p)$ is concave, and for any concave function f, we have by Jensen's inequality

$$\sum_i a_i f(i) \leqslant (\sum_i a_i) f\big(\sum_i \frac{a_i}{\sum_i a_i} i \big).$$

Since

$$\sum_{i=0}^{k-1} (d_S(i) - d_S(i+1)) + (d_S(k) - \frac{1}{2})$$
$$= d_S(0) - \frac{1}{2} = \frac{1}{2}$$

and

$$(\frac{1}{2} - d_S(k+1)) + \sum_{i=k+1}^{n} (d_S(i) - d_S(i+1))$$
$$= \frac{1}{2} - d_S(n+1) = \frac{1}{2},$$

by applying Jensen's inequality with $f = C(p)$, we have

$$(Ap)(S) \leqslant \frac{1}{2} C\bigg(p, 2\Big(\sum_{i=0}^{k-1} (d_S(i) - d_S(i+1)) \cdot i$$
$$+ (d_S(k) - \frac{1}{2}) \cdot k \Big) \bigg)$$
$$+ \frac{1}{2} C\bigg(p, 2\Big((\frac{1}{2} - d_S(k+1)) \cdot k$$
$$+ \sum_{i=k+1}^{n} (d_S(i) - d_S(i+1)) \cdot i \Big) \bigg).$$

Finally, note that the first sum

$$\sum_{i=0}^{k-1} (d_S(i) - d_S(i+1)) \cdot i + (d_S(k) - \frac{1}{2}) \cdot k$$
$$= \sum_{i=1}^{k} (d_S(i) - \frac{1}{2})$$
$$= \sum_{i=1}^{n} \max(d_S(i) - \frac{1}{2}, 0) = x.$$

Since $\sum_{i=0}^{n} (d_S(i) - d_S(i+1)) \cdot i = \sum_{i=1}^{n} d_S(i) = |S|$ by (2.4), the second sum is $|S| - x$, and so we have

$$(Ap)(S) \leqslant \frac{1}{2} C(p, 2x) + C(p, 2(|S| - x)).$$

Now we prove Lemma 2.3, which uses the definition of the combinatorial gap in (1.1).

Proof. Recall that we sort the vertices such that $1 \geqslant d_S(1) \geqslant \ldots \geqslant d_S(n) \geqslant 0$. Let k be the index such that $d_S(k) > 1/2 \geqslant d_S(k+1)$. Let $T = \{1, \ldots, |S|\}$ be the subset of the first $|S|$ vertices. We consider two cases. The first case is when $k \leqslant |S|$, in which

$$x = \sum_{i=1}^{n} \max\{d_S(i) - \frac{1}{2}, 0\} = \sum_{i=1}^{k} (d_S(i) - \frac{1}{2})$$
$$\leqslant \sum_{i=1}^{k} \frac{d_S(i)}{2} \leqslant \sum_{i=1}^{|S|} \frac{d_S(i)}{2} \leqslant \frac{1}{2}(1 - \varphi(G))|S|,$$

where the first inequality holds as $d_S(i) \in [0, 1]$, and the last inequality holds by using (1.1) to obtain $\varphi(G) \leqslant 1 - w(S, T)/|S| = 1 - \sum_{i=1}^{|S|} d_S(i)/|S|$.

The second case is when $k > |S|$. Note that $d_S(|S|) \leqslant 1 - \varphi(G)$, as otherwise T would violate (1.1). Then, for any $i \geqslant |S|$, we have $d_S(i) \leqslant d_S(|S|) \leqslant 1 - \varphi(G)$. Therefore, for any $i \geqslant |S|$,

$$d_S(i) - \frac{1}{2} \leqslant d_S(i)\left(\frac{\frac{1}{2} - \varphi(G)}{1 - \varphi(G)}\right) \leqslant \frac{d_S(i) \cdot (1 - \varphi(G))}{2},$$

and we have

$$
\begin{aligned}
x &= \sum_{i=1}^{k}(d_S(i) - \frac{1}{2}) \\
&\leqslant \sum_{i=1}^{|S|}(d_S(i) - \frac{1}{2}) + \sum_{i=|S|+1}^{k}(d_S(i) - \frac{1}{2}) \\
&\leqslant \sum_{i=1}^{|S|}(d_S(i) - \frac{1}{2}) + (\frac{1 - \varphi(G)}{2})\sum_{i=|S|+1}^{k}d_S(i) \\
&= \left(w(S,T) - \frac{|S|}{2}\right) + (\frac{1 - \varphi(G)}{2})(|S| - w(S,T)) \\
&= (\frac{1 + \varphi(G)}{2})w(S,T) - \frac{|S|}{2} + (\frac{1 - \varphi(G)}{2})|S| \\
&\leqslant (\frac{1 + \varphi(G)}{2})(1 - \varphi(G))|S| - \frac{|S|}{2} + (\frac{1 - \varphi(G)}{2})|S| \\
&\leqslant (\frac{1 - \varphi(G)}{2})|S|,
\end{aligned}
$$

where the second last inequality is by (1.1). Hence, in any case, we have $x \leqslant \frac{1}{2}(1 - \varphi(G))|S|$.

Evolving Sets Finally, we prove Theorem 1.2 about the gauge of the evolving set process, which involves very similar calculations as in the proof of Lemma 2.1. Recall that the gauge is defined as $\psi(S) = 1 - \mathbb{E}[\sqrt{|\tilde{S}|/|S|}]$. The following claim follows from Morris and Peres [MP05].

CLAIM 1. ([MP05], EQUATION 27) *For any $t \in [0,1]$, we have*

$$t \cdot \mathbb{E}[|\tilde{S}| \mid U \leqslant t] = \sum_{i \in V}\min\{t, d_S(i)\}.$$

Proof. Note that $\mathbb{P}[i \in \tilde{S} \mid U \leqslant t] = \min\{1, d_S(i)/t\}$. It follows that

$$
\begin{aligned}
\mathbb{E}[|\tilde{S}| \mid U \leqslant t] &= \sum_{i \in V}\mathbb{P}[i \in \tilde{S} \mid U \leqslant t] \\
&= \sum_{i \in V}\min\{1, d_S(i)/t\}.
\end{aligned}
$$

When $t = 1/2$, Claim 1 implies that

$$\frac{1}{2}\mathbb{E}[|\tilde{S}| \mid U \leqslant \frac{1}{2}] = \sum_{i \in V}\min\{\frac{1}{2}, d_S(i)\},$$

which is the lower area under the threshold $1/2$ (see Figure 2.2). And

$$
\begin{aligned}
\frac{1}{2}\mathbb{E}[|\tilde{S}| \mid U \geqslant \frac{1}{2}] &= \mathbb{E}[\tilde{S}] - \frac{1}{2}\mathbb{E}[|\tilde{S}| \mid U \leqslant \frac{1}{2}] \\
&= \sum_{i \in V}d_S(i) - \sum_{i \in V}\min\{\frac{1}{2}, d_S(i)\} \\
&= \sum_{i \in V}\max\{d_S(i) - \frac{1}{2}, 0\},
\end{aligned}
$$

which is the upper area above the threshold $1/2$, and this is at most $\frac{1}{2}(1 - \varphi(G))|S|$ by Lemma 2.3. We are ready to prove Theorem 1.2.

Proof. By definition,

$$
\begin{aligned}
\psi(S) &= 1 - \mathbb{E}[\sqrt{|\tilde{S}|/|S|}] \\
&= 1 - \frac{1}{2}\mathbb{E}[\sqrt{|\tilde{S}|/|S|} \mid U \leqslant \frac{1}{2}] - \frac{1}{2}\mathbb{E}[\sqrt{|\tilde{S}|/|S|} \mid U \geqslant \frac{1}{2}].
\end{aligned}
$$

Let $x = \sum_i \max\{d_S(i) - \frac{1}{2}, 0\}$ be the upper area. We know that $x \leqslant \frac{1}{2}(1 - \varphi(G))|S|$ by Lemma 2.3. By concavity,

$$
\begin{aligned}
\frac{1}{2}\mathbb{E}[\sqrt{|\tilde{S}|/|S|} \mid U \geqslant \frac{1}{2}] &\leqslant \frac{1}{2}\sqrt{\frac{1}{|S|}\mathbb{E}[|\tilde{S}| \mid U \geqslant \frac{1}{2}]} \\
&= \frac{1}{2}\sqrt{\frac{2x}{|S|}}.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
\frac{1}{2}\mathbb{E}[\sqrt{|\tilde{S}|/|S|} \mid U \leqslant \frac{1}{2}] &\leqslant \frac{1}{2}\sqrt{\frac{1}{|S|}\mathbb{E}[|\tilde{S}| \mid U \leqslant \frac{1}{2}]} \\
&= \frac{1}{2}\sqrt{\frac{2(|S| - x)}{|S|}}.
\end{aligned}
$$

Plugging these two inequalities back into the first one, we have

$$
\begin{aligned}
\psi(S) &\geqslant 1 - \frac{1}{2}(\sqrt{2 - \frac{2x}{|S|}} + \sqrt{\frac{2x}{|S|}}) \\
&\geqslant 1 - \frac{1}{2}(\sqrt{1 + \varphi(G)} + \sqrt{1 - \varphi(G)}) \geqslant \varphi(G)^2/8,
\end{aligned}
$$

where the second inequality is because $\sqrt{2 - 2x/|S|} + \sqrt{2x/|S|} \leqslant \sqrt{2 - 2y/|S|} + \sqrt{2y/|S|}$ for $x \leqslant y \leqslant |S|/2$ and we put in $y = \frac{1}{2}(1 - \varphi(G))|S|$, and the last inequality is by Taylor expansion of the function $\sqrt{1 + \varphi}$.

Small-Set Expansion of Graph Powers In this subsection, we prove Corollary 1.1 about small-set expansion of graph powers. Recall that the δ-small-set expansion is defined as

$$\phi_\delta(G) = \min_{S \subseteq V, |S| \leqslant \delta n} 1 - \frac{w(S, S)}{|S|}.$$

We define the δ-small-set combinatorial gap as

$$\varphi_\delta(G) = \min_{S \subseteq V, T \subseteq V, |S| = |T| \leqslant \delta n} 1 - \frac{w(S, T)}{|S|}.$$

The following is a simple relation between these two quantities.

LEMMA 2.4. *For any* $\delta \leqslant 1/2$*, we have* $\varphi_{\delta/2}(G) \geqslant \phi_\delta(G)/2$.

Proof. Suppose S, T are two subsets of size at most $\delta n/2$ that achieve $w(S, T) = (1 - \varphi_{\delta/2}(G))|S|$ and $|S| = |T|$. We argue that $S \cup T$ has small expansion, since

$$\begin{aligned}
w(S \cup T, S \cup T) &\geqslant 2w(S, T) - w(S \cap T, S \cap T) \\
&\geqslant 2(1 - \varphi_{\delta/2}(G))|S| - |S \cap T| \\
&= 2|S| - |S \cap T| - 2|S|\varphi_{\delta/2}(G) \\
&= |S \cup T| - 2|S|\varphi_{\delta/2}(G) \\
&\geqslant |S \cup T|(1 - 2\varphi_{\delta/2}(G)).
\end{aligned}$$

Since $|S \cup T| \leqslant 2|S| \leqslant \delta n$, this implies that $\phi_\delta(G) \leqslant 1 - w(S \cup T, S \cup T)/|S \cup T| \leqslant 2\varphi_{\delta/2}(G)$.

In [KL14], using the inequality (2.3), it is proved that for any lazy graph,

$$\phi_{\delta/2}(G^t) \geqslant \Omega(\min(\sqrt{t} \cdot \phi_\delta(G), 1)).$$

The assumption of laziness is only used for the inequality (2.3) to hold. Now, with Lemma 2.1, we can replace $\phi_\delta(G)$ by $\varphi_\delta(G)$ and prove that for any graph G,

$$\phi_{\delta/2}(G^t) \geqslant \Omega(\min(\sqrt{t} \cdot \varphi_\delta(G), 1)).$$

Combining with Lemma 2.4, it follows that for any graph G,

$$\phi_{\delta/4}(G^t) \geqslant \Omega(\min(\sqrt{t} \cdot \phi_\delta(G), 1)).$$

2.4 Vertex Expansion We will prove Theorem 1.3 in this subsection. As in Section 2.3, we will first prove faster convergence for random walks and then for evolving sets. Intuitively, larger vertex expansion will lead to faster mixing, as the probability coming into a set S is from many different vertices, and so it cannot be the case that all probability in S come from a small number of vertices with high probability. Our proof

idea is also to look at the bar chart in Figure 2.2, and translate the definition of robust vertex expansion into an upper bound of the area of the largest vertices.

For $S \subseteq V$, we consider the vector $d_S \in \mathbb{R}^n$ with the i-th entry being $d_S(i)$, and as before we assume that $d_S(1) \geqslant d_S(2) \geqslant \ldots \geqslant d_S(n)$. We use the notation $C(d_S, x)$ in (2.1) on d_S so that we can talk about the largest x values in the vector d_S (note that x could be non-integral).

In Theorem 1.3, unlike in Theorem 1.2, we need the additional assumption that the random walk matrix is lazy, such that $w(i, i) \geqslant 1/2$ for any $i \in V$. The main reason of this assumption is to have $d_S(i) \geqslant 1/2$ for $i \in S$ and $d_S(j) \leqslant 1/2$ for $j \notin S$, so that we can assume that

$$S = \{1, 2, \ldots, |S|\},$$

i.e. the vertices in S are the vertices with the largest values in d_S. Recall that we defined

$$\begin{aligned}
N_{1/2}(S) := \min\{|T| \mid T \subseteq V - S, w(S, T) \\
\geqslant \frac{1}{2}w(S, V - S)\}.
\end{aligned}$$

Since $S = \{1, \ldots, |S|\}$, we have for any integral x,

$$\begin{aligned}
\max_{T: T \subseteq V - S, |T| = x} w(S, T) &= \max_{T: T \subseteq V - S, |T| = x} \sum_{i \in T} d_S(i) \\
&= \sum_{i = |S| + 1}^{|S| + x} d_S(i) \\
&= C(d_S, |S| + x) - C(d_S, |S|).
\end{aligned}$$

In words, the set of vertices $T \subseteq V - S$ that maximize $w(S, T)$ are the vertices $\{|S| + 1, \ldots, |S| + |T|\}$ in the ordering defined by d_S. So, we can rewrite the definition of $N_{1/2}(S)$ as

$$\begin{aligned}
N_{1/2}(S) := \min\{x \mid C(d_S, |S| + x) - C(d_S, |S|) \\
\geqslant \frac{1}{2}w(S, V - S)\}.
\end{aligned}$$

Note that we allow x to be non-integral. This differs by at most one compared with the original definition in [KLM06], and will make our proofs much cleaner. The robust vertex expansion is defined as $\phi^V(S) := \min\{N_{1/2}(S)/|S|, 1\}$ and $\phi^V(G) := \min_{S: |S| \leqslant |V|/2} \phi^V(S)$ as before. Similarly, $\Psi(S)$ is defined as before using the new definition of $N_{1/2}(S)$. The following lemma translates the definition of $\phi^V(S)$ to a bound on the cumulative sum of the largest vertices in the bar chart.

LEMMA 2.5. *For any* $S \subseteq V$ *with* $|S| \leqslant |V|/2$*, we have*

$$C(d_S, (1 + \phi^V(S))|S|) \leqslant (1 - \frac{\phi(S)}{2})|S|.$$

Proof. Since $C(d_S, |S|+x)$ is continuous with respect to x, the minimum in the definition of $N_{1/2}(S)$ is attained when $C(d_S, |S| + x) - C(d_S, |S|) = \frac{1}{2} w(S, V - S)$. As $x = N_{1/2}(S) \leqslant \phi^V(S)|S|$, we have

$$C(d_S, (1 + \phi^V(S))|S|) - C(d_S, |S|) \leqslant \frac{1}{2} w(S, V - S).$$

Finally, since $S = \{1, \ldots, |S|\}$, we have $C(d_S, |S|) = w(S, S) = (1 - \phi(S))|S|$ and $w(S, V - S) = \phi(S) \cdot |S|$, we have

$$C(d_S, (1 + \phi^V(S))|S|) \leqslant (1 - \frac{\phi(S)}{2})|S|.$$

Using Lemma 2.5, we will prove a bound similar to that of Lemma 2.1. The two steps (Lemma 2.2 and Lemma 2.3) of proving Lemma 2.1 are integrated and streamlined in the proof of the following lemma.

LEMMA 2.6. *Assume $C(d_S, a|S|) \leqslant b|S|$ for any $|S| \leqslant n/2$ for some $a > 1$ and $b < 1$, then for any $p \in \mathbb{R}^n$, we have for any S with $|S| \leqslant n/2$,*

$$C(Ap, |S|) \leqslant \left(\frac{a - b}{a - b^2}\right) C(p, b|S|) + \left(\frac{b - b^2}{a - b^2}\right) C(p, \frac{a|S|}{b}).$$

Proof. Using the same concavity argument as in Lemma 2.2, for any threshold $t \in (0, 1)$ and k such that $d_S(k) > t \geqslant d_S(k + 1)$, we have for any S with $|S| \leqslant n/2$,

$$\begin{aligned}
&(Ap)(S) \\
&\leqslant (1 - t) \cdot C\left(p, \frac{1}{1-t}\Big(\sum_{i=0}^{k-1}(d_S(i) - d_S(i+1)) \cdot i \right. \\
&\qquad\qquad\qquad + (d_S(k) - t) \cdot k\Big)\Big) \\
&\quad + t \cdot C\left(p, \frac{1}{t}\Big((t - d_S(k+1)) \cdot k \right. \\
&\qquad\qquad\qquad\quad + \sum_{i=k+1}^{n}(d_S(i) - d_S(i+1)) \cdot i\Big)\right).
\end{aligned}$$

Let $x = \sum_{i=0}^{k-1}(d_S(i) - d_S(i+1)) \cdot i + (d_S(k) - t) \cdot k$ be the upper area above the threshold t. Then, it follows that

$$\begin{aligned}
C(p, |S|) &= \max_S (Ap)(S) \\
&\leqslant (1 - t) \cdot C(p, \frac{x}{1-t}) + t \cdot C(p, \frac{|S| - x}{t}).
\end{aligned}$$

It remains to prove an analog of Lemma 2.3 to bound the upper area x. We again consider two cases.

The first case is when $k \leqslant a|S|$, in which

$$\begin{aligned}
x &= \sum_{i=1}^{k}(d_S(i) - t) \leqslant \sum_{i=1}^{k}(1 - t) \cdot d_S(i) \\
&= (1 - t) \cdot C(d_S, a|S|) \leqslant (1 - t) \cdot b|S|,
\end{aligned}$$

where the last inequality uses the assumption that $C(d_S, a|S|) \leqslant b|S|$. The second case is when $k > a|S|$. Note that for any $i > a|S|$, $d_S(i) \leqslant d_S(a|S|) \leqslant b/a$, as otherwise the assumption $C(d_S, a|S|) \leqslant b|S|$ would be violated. Hence, for any $i > a|S|$,

$$d_S(i) - t \leqslant \left(\frac{b/a - t}{b/a}\right) \cdot d_S(i) = (1 - \frac{at}{b}) \cdot d_S(i),$$

and we get

$$\begin{aligned}
x &= \sum_{i=1}^{k}(d_S(i) - t) \\
&\leqslant C(d_S, a|S|) - a|S|t + \sum_{i=a|S|+1}^{k}(1 - \frac{at}{b}) \cdot d_S(i) \\
&= C(d_S, a|S|) - a|S|t + \big(|S| - C(d_S, a|S|)\big)(1 - \frac{at}{b}) \\
&\leqslant |S|(1 - \frac{at}{b}),
\end{aligned}$$

where the last inequality uses the assumption that $C(d_S, a|S|) \leqslant b|S|$ and some simple calculations. We balance the two upper bounds $b|S|(1 - t)$ and $|S|(1 - at/b)$ by choosing $t = (b - b^2)/(a - b^2)$, so that in both cases we have

$$(2.5) \qquad x \leqslant \frac{b(a - b)}{a - b^2}|S|.$$

Putting the choice of t and the bound on x back (and using concavity), we obtain the conclusion of the lemma.

Lemma 2.6 is a generalization of (2.3), and we can use it to derive a generalization of (2.2). Note that any probability distribution p satisfies the condition $C(p, x) \leqslant \frac{x}{n} + \sqrt{\min\{x, n - x\}}$ in the following lemma.

LEMMA 2.7. *Assuming $C(d_S, a|S|) \leqslant b|S|$ for any $|S| \leqslant n/2$ for some $a > 1$ and $b < 1$, then for any $p \in \mathbb{R}^n$ satisfying $C(p, x) \leqslant \frac{x}{n} + c\sqrt{\min\{x, n - x\}}$ for all $x \in [0, n]$ for some c, we have*

$$\begin{aligned}
C(A^t p, x) &\leqslant \frac{x}{n} + c\sqrt{\min\{x, n - x\}} \\
&\qquad \cdot (1 - \frac{(\sqrt{a} - \sqrt{b})(1 - \sqrt{b})}{\sqrt{a} + b})^t.
\end{aligned}$$

Proof. We only consider the case that $x \leqslant n/2$ in the following; the case $x > n/2$ can be handled in the same manner. When $x \leqslant n/2$, we have $C(p, x) \leqslant \frac{x}{n} + c\sqrt{x}$. By Lemma 2.6,

$$
\begin{aligned}
C(Ap, x) &\leqslant \left(\frac{a-b}{a-b^2}\right) \cdot C(p, bx) + \left(\frac{b-b^2}{a-b^2}\right) \cdot C(p, \frac{ax}{b}) \\
&\leqslant \left(\frac{a-b}{a-b^2}\right) \cdot \left(\frac{x}{n} + c\sqrt{bx}\right) \\
&\quad + \left(\frac{b-b^2}{a-b^2}\right) \cdot \left(\frac{x}{n} + c\sqrt{\frac{ax}{b}}\right) \\
&= \frac{x}{n} + c\sqrt{x}\left(\left(\frac{a-b}{a-b^2}\right)\sqrt{b} + \left(\frac{b-b^2}{a-b^2}\right)\sqrt{\frac{a}{b}}\right) \\
&= \frac{x}{n} + \frac{c\sqrt{x}}{a-b^2}\left(\left((a-b) + (1-b)\sqrt{a}\right)\sqrt{b}\right) \\
&= \frac{x}{n} + \frac{c\sqrt{x}}{a-b^2}\left((\sqrt{a}-b)(1+\sqrt{a})\sqrt{b}\right) \\
&= \frac{x}{n} + c\sqrt{x}\left(\frac{(1+\sqrt{a})\sqrt{b}}{\sqrt{a}+b}\right).
\end{aligned}
$$

Now note that,

$$
\begin{aligned}
1 - \frac{(1+\sqrt{a})\sqrt{b}}{\sqrt{a}+b} &= \frac{\sqrt{a}+b-\sqrt{b}-\sqrt{ab}}{\sqrt{a}+b} \\
&= \frac{(\sqrt{a}-\sqrt{b})(1-\sqrt{b})}{\sqrt{a}+b}.
\end{aligned}
$$

The above argument shows that $C(Ap, x) \leqslant \frac{x}{n} + c\sqrt{x}(1 - (\sqrt{a}-\sqrt{b})(1-\sqrt{b})/(\sqrt{a}+b))$. Apply the same argument inductively (with different c in each iteration) gives the lemma. ∎

In particular, for any probability distribution p, we have the following generalization of (2.2):

$$
C(A^t p, x) \leqslant \frac{x}{n} + \sqrt{x} \cdot \left(1 - \frac{(\sqrt{a}-\sqrt{b})(1-\sqrt{b})}{\sqrt{a}+b}\right)^t.
$$

Evolving Sets We prove Theorem 1.3 about the gauge in the evolving set process.

Proof. By Claim 1, for any $t \in [0, 1]$, the lower area below threshold t is

$$
\begin{aligned}
t \cdot \mathbb{E}[|\tilde{S}| \mid U \leqslant t] &= \sum_{i \in V} \min\{t, d_S(i)\} \\
&= |S| - \sum_{i \in V} \max\{d_S(i) - t, 0\},
\end{aligned}
$$

and the upper area above threshold t is

$$
\begin{aligned}
(1-t) \cdot \mathbb{E}[|\tilde{S}| \mid U \geqslant t] &= \mathbb{E}[\tilde{S}] - t \cdot \mathbb{E}[|\tilde{S}| \mid U \leqslant t] \\
&= \sum_{i \in V} \max\{d_S(i) - t, 0\}.
\end{aligned}
$$

By the same argument in Lemma 2.6, using the assumption $C(d_S, a|S|) \leqslant b|S|$ and setting $t = (b-b^2)/(a-b^2)$, the upper area above threshold t is

$$
x := \sum_{i \in V} \max\{d_S(i) - t, 0\} \leqslant \frac{b(a-b)}{a-b^2}|S|
$$

as stated in (2.5). Therefore,

$$
\begin{aligned}
\psi(S) &= 1 - \mathbb{E}[\sqrt{|\tilde{S}|/|S|}] \\
&= 1 - t \cdot \mathbb{E}[\sqrt{|\tilde{S}|/|S|} \mid U \leqslant t] \\
&\quad - (1-t) \cdot \mathbb{E}[\sqrt{|\tilde{S}|/|S|} \mid U \geqslant t] \\
&\geqslant 1 - t\sqrt{\mathbb{E}[|\tilde{S}|/|S| \mid U \leqslant t]} \\
&\quad - (1-t)\sqrt{\mathbb{E}[|\tilde{S}|/|S| \mid U \geqslant t]} \\
&= 1 - t\sqrt{\frac{1}{|S|}\frac{|S|-x}{t}} - (1-t)\sqrt{\frac{1}{|S|}\frac{x}{1-t}} \\
&\geqslant 1 - t\sqrt{\frac{1}{t|S|}\left(|S| - \frac{b(a-b)}{a-b^2}|S|\right)} \\
&\quad - (1-t)\sqrt{\frac{1}{(1-t)|S|}\frac{b(a-b)}{a-b^2}|S|},
\end{aligned}
$$

where the first inequality is by the concavity of the square root function, and the second inequality is by the following fact: Suppose f is a concave function and $c_1 \geqslant c_2 \geqslant c_3 \geqslant c_4$ satisfy $tc_1 + (1-t)c_4 = tc_2 + (1-t)c_3$ for some $t \in [0, 1]$, then $tf(c_1) + (1-t)f(c_4) \leqslant tf(c_2) + (1-t)f(c_3)$. Note that

$$
\begin{aligned}
\frac{1}{t|S|}\left(|S| - \frac{b(a-b)}{a-b^2}|S|\right) &= \frac{a-b^2}{b-b^2}\left(1 - \frac{b(a-b)}{a-b^2}\right) \\
&= \frac{a-b^2-b(a-b)}{b-b^2} = \frac{a}{b},
\end{aligned}
$$

and

$$
\frac{1}{(1-t)|S|}\frac{b(a-b)}{a-b^2}|S| = \frac{a-b^2}{a-b}\frac{b(a-b)}{a-b^2} = b.
$$

Hence,

$$
\psi(S) \geqslant 1 - t\sqrt{\frac{a}{b}} - (1-t)\sqrt{b} \geqslant \frac{(\sqrt{a}-\sqrt{b})(1-\sqrt{b})}{\sqrt{a}+b},
$$

where the last inequality follows from the calculations in Lemma 2.7 (starting from the third line in the first block of calculations). We put $a = 1 + \phi^V(S)$ and $b = 1 - \frac{\phi(S)}{2}$. Note that since $\phi^V(S) \leqslant 1$ by definition, we have $\sqrt{a} \geqslant 1 + \phi^V(S)/3$. This is the only place

we need to assume $\phi^V(S) \leqslant 1$. On the other hand $\sqrt{b} \leqslant 1 - \phi(S)/2$. Therefore we have

$$\psi(S) \geqslant \frac{(\frac{1}{3}\phi^V(S) + \frac{1}{2}\phi(S))\frac{1}{2}\phi(S)}{3}$$
$$\geqslant \frac{\phi^V(S)\phi(S)}{18} = \frac{\Psi(S)}{18}.$$

Local Graph Partitioning We obtain Corollary 1.3 about the performance of the evolving set algorithm in [AP09, OT12]. In Lemma 5.2 of [OT12], Oveis Gharan and Trevisan actually showed that the (volume biased) evolving set process will return a set S with $\psi(S) = O(\phi(S^*)/\epsilon)$, and they used the fact that $\psi(S) = \Omega(\phi(S)^2)$ to get Theorem 1.1. Now, with Theorem 1.3, we can replace $\phi(S)^2$ by $\Psi(S)$ and obtain Corollary 1.3.

2.5 General Graphs Our results generalize to non-regular undirected graphs, with appropriate changes in various definitions.

Expansion: For a general undirected graph G, we use $\mathrm{vol}(S) := \sum_{i \in S} \deg(i)$ to denote the volume of a subset S. It is the non-regular analog of the size $|S|$. The conductance of a set $S \subseteq V$ and the conductance of the graph are defined as

$$\phi(S) := \frac{w(S, V - S)}{\mathrm{vol}(S)},$$

and

$$\phi(G) := \min_{S: \mathrm{vol}(S) \leqslant \mathrm{vol}(V)/2} \phi(S).$$

Ideally, the analog of the combinatorial gap would be

$$\varphi(G) := \min_{S,T: \mathrm{vol}(S) = \mathrm{vol}(T) \leqslant \mathrm{vol}(V)/2} 1 - \frac{w(S,T)}{\mathrm{vol}(S)}.$$

However, this definition may not say much since it can happen that any two different subsets have different volume. In order to handle this situation, we revise the definition and allow S and T to be fractional. Let $\vec{d}$ be the degree vector of G. We define the combinatorial gap as

$$\varphi(G) := \min_{\substack{\chi_S \in [0,1]^V, \chi_T \in [0,1]^V \\ \langle \chi_S, \vec{d}\rangle = \langle \chi_T, \vec{d}\rangle \leqslant \mathrm{vol}(V)/2}} 1 - \frac{\langle \chi_S, A\chi_T \rangle}{\langle \chi_S, \vec{d}\rangle}.$$

Lovász Simonovits curve: In general graphs, the Lovász Simonovits curve $C(p) : \mathrm{vol}(V) \to \mathbb{R}$ is defined as
$$C(p, x) := \max_{c \in [0,1]^n : \langle c, \vec{d}\rangle = x} \langle c, p \rangle.$$

Suppose the vertices are sorted so that $p(1) \geq p(2) \geq \cdots \geq p(n)$. The extreme points of the curve $C(p)$ is $\sum_{j=1}^{i} \deg(j)$ for $j = 0, \ldots n$.

Bar chart: We sort the vertices so that $d_S(i)/\deg(i)$ is decreasing. We should view the bar chart so that each bar has width $\deg(i)$ and height $d_S(i)/\deg(i)$. So the total width is $\mathrm{vol}(V)$. Same as before, we put a threshold $1/2$ (or choosing another threshold $t \in [0, 1]$ in the proof for vertex expansion) and consider the upper area x and lower area y, and show that

$$(Ap)(S) \leqslant \frac{1}{2}(C(p, 2x) + C(p, 2y)).$$

With this figure in mind, the proofs for the extended results are essentially the same as the original proofs.

Vertex expansion: The robust vertex expansion is defined as follows. Let $q(i) = d_S(i)/\deg(i)$, and

$$N_{1/2}(S) := \min\{x \mid C(q, \mathrm{vol}(S) + x) - C(q, \mathrm{vol}(S))$$
$$\geq \frac{w(S, \bar{S})}{2}\}.$$

Then $\phi^V(S) := \min\{N_{1/2}(S)/\mathrm{vol}(S), 1\}$, and $\phi^V(G) := \min_{S: \mathrm{vol}(S) \leqslant \mathrm{vol}(V)/2} \phi^V(S)$.

Restating the Theorems for General Graphs In the following, we restate our results on general graphs without proofs. Lemma 2.1 becomes

LEMMA 2.8. *For any extreme point $x \leqslant \mathrm{vol}(V)/2$,*

$$C(Ap, x) \leqslant \frac{1}{2}(C(p, x(1 - \varphi(G))) + C(p, x(1 + \varphi(G)))).$$

Lemma 2.2 becomes

LEMMA 2.9. *For any subset $S \subseteq V$ with $\mathrm{vol}(S) \leqslant \mathrm{vol}(V)/2$, let*

$$x = \sum_{i \in V} \deg(i) \cdot \max\{\frac{d_S(i)}{\deg(i)} - \frac{1}{2}, 0\},$$

then

$$(Ap)(S) \leqslant \frac{1}{2}(C(p, 2x), C(p, 2(\mathrm{vol}(V) - x))),$$

for any random walk matrix A and any vector $p \in \mathbb{R}^n$.

Lemma 2.3 becomes

LEMMA 2.10. *For any subset $S \subseteq V$ with $\mathrm{vol}(S) \leqslant \mathrm{vol}(V)/2$, let $x = \sum_{i \in V} \deg(i) \max\{d_S(i)/\deg(i) - 1/2, 0\}$, then*

$$x \leqslant \frac{1}{2}(1 - \varphi(G))\mathrm{vol}(S).$$

Lemma 2.6 becomes

LEMMA 2.11. *Let* $q(i) := d_S(i)/\deg(i)$. *Assume* $C(q, a\operatorname{vol}(S)) \leqslant b\operatorname{vol}(S)$ *for any* S *with* $\operatorname{vol}(S) \leqslant \operatorname{vol}(V)/2$ *for some* $a > 1$ *and* $b < 1$, *then for any* $p \in \mathbb{R}^n$, *we have*

$$C(p, \operatorname{vol}(S))$$
$$\leqslant (\frac{a-b}{a-b^2}) \cdot C(p, b\operatorname{vol}(S)) + (\frac{b-b^2}{a-b^2}) \cdot C(p, \frac{a\operatorname{vol}(S)}{b}).$$

With these definitions and lemmas in place, Theorem 1.2 and Theorem 1.3 hold as stated in the introduction.

3 Limitations

We prove Theorem 1.4 that provides a hard small-set expansion instance for the evolving set process studied in [AP09, OT12, AOPT16]. As mentioned in the introduction, it will be a noisy hypercube H over alphabet size $k = 1/\delta$. Formally, H is a graph on k^d vertices, representing all strings of length d over alphabet $[k]$. For two vertices x, y, the edge weight $w(x, y)$ is set to be the probability to go from vertex x to vertex y in one step of a random walk, where each symbol of x is independently rerandomized with probability ϵ: For each $i \in [d]$, with probability $1 - \epsilon$, set $y_i = x_i$, otherwise y_i is sampled uniformly at random from $[k]$. Note that H is 1-regular. It is easy to see that H has a small sparse cut.

CLAIM 2. *There is a set* S *with expansion at most* ϵ *and* $|S| = \delta n$ *where* $\delta = 1/k$.

Proof. Indeed, the coordinate cut $S = \{x \in [k]^d \mid x_1 = 0\}$ has size δn and expansion ϵ.

We will show that all the sets explored by the evolving set process have expansion close to one. First, we argue that the evolving set process will only explore the Hamming balls of the noisy hypercube in Lemma 3.1. Then, we will show that the expansion of all Hamming balls of size $O(\delta n)$ is close to one in Lemma 3.2.

The evolving set process starts from a singleton set on H. By symmetry, we may assume this set is $\{0^d\}$. We now show that the evolving set process only explores sets that are Hamming balls $B(r)$ (around 0^d), where $B(r)$ denotes all strings of Hamming weight at most r:

$$B(r) := \{x \in [k]^d \mid |x| \leqslant r\}$$

where

$$|x| := |\{i \in [d] \mid x_i \neq 0\}|.$$

Indeed, the initial set $\{0^d\}$ is the Hamming ball $B(0)$. The following lemma implies that, if the current set is a Hamming ball $B(r)$, then so is the next set,

and thus by induction the evolving set process will only explore Hamming balls. Intuitively the lemma says that if x has less Hamming weight than y, then x is closer to $B(r)$.

LEMMA 3.1. *Suppose* $\epsilon \leqslant 1/2$. *For any* $r \geqslant 0$, *any* $x, y \in [k]^d$, *if* $|x| \leqslant |y|$, *then*

$$w(x, B(r)) \geqslant w(y, B(r)).$$

(It follows that if S *is a Hamming ball, then* $y \in \tilde{S}$ *implies that* $x \in \tilde{S}$, *and thus* $\tilde{S}$ *is also a Hamming ball.)*

Proof. Note that $w(x, z)$ depends only on the Hamming distance $|x - z|$ (coordinate-wise subtraction modulo k). We first show via a symmetry argument that

$$(3.1) \quad w(x, B(r)) = w(y, B(r)) \text{ whenever } |x| = |y|.$$

To this end, we will construct a permutation π on $[k]^d$ that (i) preserves Hamming distances: $|\pi(a) - \pi(b)| = |a - b|$ for all $a, b \in [k]^d$, (ii) $\pi(x) = y$, and (iii) $\pi(0^d) = 0^d$. Assuming this permutation exists, we get that $z \in B(r)$ if and only if $\pi(z) \in B(r)$, since $|\pi(z) - 0| = |\pi(z) - \pi(0)| = |z - 0|$. Also, $|x - z| = |\pi(x) - \pi(z)| = |y - \pi(z)|$, thus

$$w(x, z) = w(y, \pi(z)).$$

Summing this equality over all $z \in B(r)$, we get (3.1).

We now construct such a permutation π. Take any bijection σ on $[d]$ that maps $I := \{i \mid x_i \neq 0\}$ onto $\{i \mid y_i \neq 0\}$. For $i \in I$, let τ_i be the permutation on $[k]$ that simply swaps x_i and $y_{\sigma(i)}$. Then we define $\pi(a) = b$ where $b_i = a_i$ if $i \notin I$, and $b_i = \tau_i(a_{\sigma(i)})$ otherwise. It is easy to verify that π has all the required properties.

We now deal with the general case $|x| < |y|$. It suffices to prove the lemma assuming $|y| = |x| + 1$. By (3.1), we may assume that x is the indicator vector on a subset $S = [c]$ for some c, and y is the indicator vector on $[c+1]$, so that y differs from x only at position c. Picking a random neighbor Z of x is equivalent to picking $W = x - Z$ and setting $Z = x + W$, so the lemma is equivalent to $\mathbb{P}_W[x + W \in B(r)] \geqslant \mathbb{P}_W[y + W \in B(r)]$. In fact, we will show this inequality conditioned on all values of W_i except $i = c$. Let $w_{-c} \in [k]^{[d] \setminus \{c\}}$ be a fixing of all those values. We will show

$$\mathbb{P}_{W_c}[x + W \in B(r) \mid W_{-c} = w_{-c}]$$
$$(3.2) \quad \geqslant \mathbb{P}_{W_c}[y + W \in B(r) \mid W_{-c} = w_{-c}].$$

There are two cases. If $|x_{-c} + w_{-c}| \neq r$, then $x + W$ and $y + W$ are both in $B(r)$ or both outside of $B(r)$, and therefore (3.2) holds as an equality. In the remaining case, the left hand side of (3.2) is at least $\mathbb{P}[W_c = 0] \geqslant 1 - \epsilon$, while the right hand side is at most $\mathbb{P}[W_c \neq 0] \leqslant \epsilon$, so the inequality follows by our assumption that $\epsilon \leqslant 1/2$.

We now show that any small Hamming ball has large expansion. The same result appears earlier in [CMN14] in a different context. We give a proof below, filling in some details missing in [CMN14]. The main idea is to show via central limit theorem that Hamming balls have similar expansion as in the Gaussian noise graph, which can be analyzed using hypercontractive inequality. Similar connection was commonly used to show integrality gap examples for convex relaxations, and here it is used for the first time to show limitations for random walks based algorithms (see the end of this section for the subtle difference between the two settings).

LEMMA 3.2. *For any $\epsilon, \eta > 0$, there exists $\delta = \delta_{\epsilon,\eta}$ (independent of k) such that for any sufficiently large $d \geqslant d_{\epsilon,\delta,\eta}$, all Hamming balls of size $\leqslant \delta n$ has expansion $1 - O(\eta)$.*

Proof. We will analyze the expansion of a Hamming ball $B(r)$ by relating $B(r)$ to halfspaces $A_{r'} := \{x \in \mathbb{R} \mid x \leqslant r'\}$ in Gaussian probability space.

Consider drawing a random edge (x, y) from H according to its weight, and we would like to analyze the probability that both vertices are in $B(r)$. The event $x \in B(r)$ is the same as $|x| = \sum_{i \in [d]} \mathbb{1}(x_i \neq 0) \leqslant r$. Since $|x|$ is a sum of independent random variables and each summand has bounded third moment, by Berry–Esseen central limit theorem, for large d, the sum is closely approximated by a Gaussian random variable g with the same mean and variance as $|x|$. That is, for all large enough $d \geqslant d_{\epsilon,\delta,\eta}$,

$$(3.3) \qquad \mathbb{P}[|x| \leqslant r] \approx_{\delta'} \mathbb{P}[g \leqslant r] \text{ for all } r \in \mathbb{R}$$

for some δ' depending on ϵ, δ, η to be specified later. Here we write $C \approx_{\delta'} D$ to mean $|C - D| \leqslant \delta'$.

Moreover, multivariate central limit theorem (e.g. [Saz68]) implies that the event $(|x| \leqslant r) \wedge (|y| \leqslant r)$ has roughly the same probability as the event $(g \leqslant r) \wedge (h \leqslant r)$, where the bivariate Gaussian (g, h) has the same mean and covariance as $(|x|, |y|)$. That is, for large enough d, it holds that for all $r \in \mathbb{R}$,

$$(3.4) \ \ Pr[(|x| \leqslant r) \wedge (|y| \leqslant r)] \approx_{\delta'} Pr[(g \leqslant r) \wedge (h \leqslant r)].$$

We note that the following calculations do not depend on the dimension d other than the CLT approximation errors (as we are not concerned about the graph size $n = k^d$), so we can choose a very large d at the end to make the CLT approximation errors δ' to be arbitrarily small for the proof to go through.

Shift g and h to have zero mean and renormalize them to have unit variance. We get $g' = (g -$

$\mathbb{E}[g]) / \sqrt{\mathrm{Var}[g]}$ from g and similarly h' from h. Then g' and h' have covariance

$$\mathrm{Cov}(g', h') = \frac{\mathrm{Cov}(|x|, |y|)}{\sqrt{\mathrm{Var}[|x|] \, \mathrm{Var}[|y|]}}$$
$$= \frac{\mathrm{Cov}(\mathbb{1}(x_1 \neq 0), \mathbb{1}(y_1 \neq 0))}{\sqrt{\mathrm{Var}[\mathbb{1}(x_1 \neq 0)] \, \mathrm{Var}[\mathbb{1}(y_1 \neq 0)]}}.$$

We have $\mathrm{Cov}(\mathbb{1}(x_1 \neq 0), \mathbb{1}(y_1 \neq 0)) = (1 - \epsilon) \, \mathrm{Var}[\mathbb{1}(x_1 \neq 0)]$, so $\mathrm{Cov}(g', h') = 1 - \epsilon$. Therefore,

$$\mathbb{P}[(g \leqslant r) \wedge (h \leqslant r)] = \mathbb{P}[(g' \leqslant r') \wedge (h' \leqslant r')],$$

where r' is chosen so that $\mathbb{P}[g' \leqslant r'] = \mathbb{P}[g \leqslant r]$.

For standard Gaussians g' and h' with covariance $1 - \epsilon$, we claim that

$$(3.5) \qquad \mathbb{P}[h' \leqslant r' \mid g' \leqslant r'] \leqslant \mathbb{P}[g' \leqslant r']^{\epsilon/2}.$$

This inequality follows from Gaussian hypercontractive inequality (e.g. [ODo14, Section 11.1])

$$\underset{(1-\epsilon) \text{ correlated } g', h'}{\mathbb{E}} [\mathbb{1}_{\leqslant r'}(g') \mathbb{1}_{\leqslant r'}(h')] \leqslant \|\mathbb{1}_{\leqslant r'}\|_{2-\epsilon}^2,$$

where $\mathbb{1}_{\leqslant r'}(g') := \mathbb{1}(g' \leqslant r')$ is the indicator function for the halfspace $A_{r'}$, and the fact that

$$\|\mathbb{1}_{\leqslant r'}\|_{2-\epsilon} = \mathbb{E}[\mathbb{1}_{\leqslant r'}(g')^{2-\epsilon}]^{1/(2-\epsilon)}$$
$$= \mathbb{E}[\mathbb{1}_{\leqslant r'}(g')]^{1/(2-\epsilon)} = \mathbb{P}[g' \leqslant r']^{1/(2-\epsilon)}.$$

Set $\delta := \eta^{2/\epsilon}$. Note that for r' small enough so that $\mathbb{P}[g' \leqslant r'] \leqslant \delta$, then (3.5) implies that the halfspace $A_{r'}$ has expansion

$$\mathbb{P}[h' > r \mid g' \leqslant r'] = 1 - \mathbb{P}[h' \leqslant r' \mid g' \leqslant r']$$
$$\geqslant 1 - \mathbb{P}[g' \leqslant r']^{\epsilon/2} \geqslant 1 - \delta^{\epsilon/2} = 1 - \eta.$$

We use (3.3) and (3.4) to translate this expansion result from Gaussian space to the noisy hypercube. Let $\delta'' \leqslant \delta$ be a constant depending on ϵ, δ, η to be specified later. Any Hamming ball of H of size at least $\delta'' n$ corresponds to a halfspace of roughly the same Gaussian measure via (3.3), and has roughly the same noise stability via (3.4). Choosing the CLT approximation error $\delta' := \delta'' \eta$, we can ensure that all Hamming balls $B(r)$ of size between $\delta'' n$ and δn have expansion $\geqslant 1 - \eta$. Indeed,

$$\mathbb{P}[x \text{ and } y \in B(r)] \leqslant \mathbb{P}[g' \text{ and } h' \in A_{r'}] + \delta'$$

by (3.4) and

$$\delta \geqslant \mathbb{P}[x \in B(r)] \geqslant \mathbb{P}[g' \in A_{r'}]/2$$

by (3.3) and our assumption that $\mathbb{P}[x \in B(r)] \geqslant \delta''$, so

$$
\begin{aligned}
1 - \phi(B(r)) &= \frac{\mathbb{P}[x \text{ and } y \in B(r)]}{\mathbb{P}[x \in B(r)]} \\
&\leqslant 2 \frac{\mathbb{P}[g' \text{ and } h' \in A_{r'}] + \delta'}{\mathbb{P}[g' \in A_{r'}]}.
\end{aligned}
$$

We now analyze the right hand side. We have

$$
\frac{\mathbb{P}[g' \text{ and } h' \in A_{r'}]}{\mathbb{P}[g' \in A_{r'}]} \leqslant \mathbb{P}[g' \in A_{r'}]^{\epsilon/2} \leqslant O(\eta)
$$

by (3.5) and the fact that $\mathbb{P}[g' \in A_{r'}] \leqslant 2\mathbb{P}[x \in B(r)] \leqslant 2\delta$. Also $\delta'/\mathbb{P}[g' \in A_{r'}] \leqslant 2\delta'/\mathbb{P}[x \in B(r)] \leqslant 2\delta'/\delta'' = 2\eta$. Therefore $\phi(B(r)) \geqslant 1 - O(\eta)$, for those $B(r)$ of size between $\delta''n$ and δn, as required.

To deal with Hamming balls of size smaller than $\delta''n$, we simply apply the hypercontractive inequality on H directly. For any subset B on H (not necessarily a Hamming ball), we have

$$
\mathbb{P}[y \in B \mid x \in B] \leqslant \mathbb{P}[x \leqslant B]^{c\epsilon/\log(1/\delta)}
$$

for some $c > 0$. The exponent $c\epsilon/\log(1/\delta)$ is from [Wol07]. Taking $\delta'' := \eta^{\log(1/\delta)/c\epsilon}$, we see that $\phi(B) \geqslant 1 - \eta$ whenever $|B| \leqslant \delta''n$.

The key point of Lemma 3.2 is that everything is independent of k. Therefore, given any ϵ, we just need to set $k \geqslant 1/\delta$ so that H has a set of expansion ϵ and size $\leqslant \delta n$ (by Claim 2), while the evolving set process only explores Hamming balls (by Lemma 3.1) and all Hamming balls of size $\leqslant \delta_{\epsilon,\eta}n$ have expansion $\geqslant 1 - \epsilon$ (by Lemma 3.2). This proves Theorem 1.4 that the evolving set process fails on the k-ary ϵ-noisy hypercube with probability one.

Random Walks, Personal Pagerank, and Heat Kernels The random walk local graph partitioning algorithm [ST13, ABS10, KL12] works by computing the vector $p_t := A^t \chi_v$ for every vertex v for $1 \leqslant t \leqslant O(\log n)$, sorting the vertices so that $p_t(1) \geqslant p_t(2) \geqslant \ldots \geqslant p_t(n)$, and trying all the level sets $\{1, \ldots, j\}$ for $1 \leqslant j \leqslant n$. Using the same k-ary noisy hypercube example, it is not difficult to see from Lemma 3.1 that all the level sets that the algorithm explored are Hamming balls, and thus the random walk algorithm will also fail to disprove the small-set expansion hypothesis.

The same argument also applies to the personal pagerank algorithm [ACL06, ZLM13] and the heat kernel algorithm, which work by computing some related vectors and trying all the level sets. We note that the vectors used by these algorithms are just convex combinations of the random walk vectors $A^t \chi_v$ for different t,

and therefore all the level sets are Hamming balls, and hence these algorithms also fail for the same reason.

We believe that this example exposes the limitations of all known local graph partitioning algorithms, and can be used as a basis to prove further lower bounds. An interesting question is to study whether the analysis of the $O(\sqrt{\phi(S)\log(|S|)})$-approximation of the evolving set algorithm in Theorem 1.1 is tight when $\epsilon = 1/\log(|S|)$.

Differences to integrality gap lower bounds Our lower bound instances are related to SDP integrality gap instances for small-set expansion and unique games, with subtle conceptual differences.

The integrality gap instances in [KV15, KS09] are based on Khot–Vishnoi instances (folded noisy Boolean hypercubes), and their analysis ultimately relies on the hypercontractive inequality on the Boolean hypercube. Directly using a similar hypercontractive inequality on the large alphabet noisy hypercube will not give us the bound we want, because our graphs have small sparse cuts. In other words, Khot–Vishnoi rules out all small sparse cuts with the help of folding, while we cannot fold so that we can plant small sparse cuts. This conceptual difference explains why our instance and analysis have been overlooked.

Acknowledgement

References

[Alo86] N. Alon. *Eigenvalues and expanders*. Combinatorica, 6, 83–96, 1986.

[AM85] N. Alon, V. Milman. λ_1, *isoperimetric inequalities for graphs, and superconcentrators*. Journal of Combinatorial Theory, Series B, 38(1), 73–88, 1985.

[ACL06] R. Andersen, F.R.K. Chung, K.J. Lang. *Local graph partitioning using PageRank vectors*. In Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS), 475–486, 2006.

[AOPT16] R. Andersen, S. Oveis Gharan, Y. Peres, L. Trevisan. *Almost optimal local graph clustering using evolving sets.* Journal of the ACM 63(2), Article No. 15, 2016.

[AP09] R. Andersen, Y. Peres. *Finding sparse cuts locally using evolving sets.* In Proceedings of the 41st Annual ACM Symposium on Theory of Computing (STOC), 235–244, 2009.

[ABS10] S. Arora, B. Barak, D. Steurer. *Subexponential algorithms for unique games and related problems.* In Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS), 563–572, 2010.

[BBHKSZ12] B. Barak, F.G.S.L. Brandao, A.W. Harrow, J.A. Kelner, D. Steurer, Y. Zhou. *Hypercontractivity, sum-of-squares proofs, and their applications.* In Proceedings of the 44th Annual ACM Symposium on Theory of Computing (STOC), 307–326, 2012.

[BGHMRS12] B. Barak, P. Gopalan, J. Hastad, R. Meka, P. Raghavendra, D. Steurer. *Making the long code shorter.* In Proceedings of the 53rd Annual IEEE Symposium on Foundations of Computer Science (FOCS), 370–379, 2012.

[BKS14] B. Barak, J.A. Kelner, D. Steurer. *Rounding sum-of-squares relaxations.* In Proceedings of the 46th Annaul ACM Symposium on Theory of Computing (STOC), 31–40, 2014.

[BL06] Y. Bilu, N. Linial. *Lifts, discrepancy and nearly optimal spectral gap.* Combinatorics 26(5), 495–519, 2006.

[CMN14] S.O. Chan, E. Mossel, J. Neeman. *On extracting common random bits from correlated sources on large alphabets.* IEEE Transactions on Information Theory 60(3), 1630–1637, 2014.

[Kho10] S. Khot *On the Unique Games Conjecture (Invited Survey).* IEEE Conference on Computation Complexity, 99–121, 2010

[KKMO07] S. Khot, G. Kindler, E. Mossel, R. O'Donnell. *Optimal inapproximability results for MAX-CUT and other 2-variable CSPs?* SIAM Journal on Computing 37(1), 319–357, 2007.

[KLM06] R. Kannan, L. Lovász, R. Montenegro. *Blocking conductance and mixing in random walks.* Combinatorics, Probability and Computing 15(4), 541–570, 2006.

[KL12] T.C. Kwok, L.C. Lau. *Finding small sparse cuts by random walk.* In Proceedings of the 16th International Workshop on Randomization and Computation (RANDOM), 615–626, 2012.

[KL14] T.C. Kwok, L.C. Lau. *Lower bounds on expansions of graph powers.* In Proceedings of the 17th Annual International Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX), 313–324, 2014.

[KLLOT13] T.C. Kwok, L.C. Lau, Y.T. Lee, S. Oveis Gharan, L. Trevisan. *Improved Cheeger's inequality: Analysis of spectral partitioning algorithms through higher order spectral gap.* In Proceedings of the 45th

Annual Symposium on Theory of Computing (STOC), 11–20, 2013.

[KLL16] T.C. Kwok, L.C. Lau, Y.T. Lee. *Improved Cheeger's inequality and analysis of local graph partitioning using vertex expansion and expansion profile.* In Proceedings of the 27th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), 1848–1861, 2016.

[KV15] S. Khot, N. Vishnoi. *The Unique Games Conjecture, integrality gap for cut problems and embeddability of negative-type metrics into ℓ_1.* Journal of Association of Computing Machinery 62(1), 8:1–8:39, 2015

[KS09] S. Khot, R. Saket. *SDP intergality gaps with local ℓ_1-embeddability.* In Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science (FOCS), 565–574, 2009.

[LPW08] D.A. Levin, Y. Peres, E.L. Wilmer. *Markov chains and mixing times.* American Mathematical Society, 2008.

[LK99] L. Lovász, R. Kannan. *Faster mixing via average conductance.* In Proceedings of the 31st Annual ACM Symposium on Theory of Computing (STOC), 282–287, 1999.

[LS90] L. Lovász, M. Simonovits. *The mixing time of Markov chains, an isoperimetric inequality, and computing the volume.* In Proceedings of the 31st Annual IEEE Symposium on Foundations of Computer Science (FOCS), 346–354, 1990.

[MT06] R. Montenegro, P. Tetali. *Mathematical aspects of mixing times in Markov chains.* Foundations and Trends in Theoretical Computer Science, Now Publishers, 2006.

[MOO10] E. Mossel, R. O'Donnell, K. Oleszkiewicz. *Noise stability of functions with low influences: Invariance and optimality.* Annals of Mathematics, 171(1), 2010

[MP05] B. Morris, Y. Peres. *Evolving sets, mixing and heat kernel bounds.* Probability Theory and Related Fields 133(2), 245–266, 2005.

[ODo14] R. O'Donnell. *Analysis of Boolean Functions.* Cambridge University Press, 2014.

[Ove13] S. Oveis Gharan. *New rounding techniques for the design and analysis of approximation algorithms.* PhD thesis, Stanford University, 2013.

[OT12] S. Oveis Gharan, L. Trevisan. *Approximating the expansion profile and almost optimal local graph clustering.* In Proceedings of the 53rd Annual IEEE Symposium on Foundations of Computer Science (FOCS), 187–196, 2012.

[RS14] P. Raghavendra, T. Schramm. *Gap amplication for small-set expansion via random walks.* In Proceedings of the 17th Annual International Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX), 381–391, 2014.

[RS10] P. Raghavendra, D. Steurer. *Graph expansion and the unique games conjecture.* In Proceedings of the 42nd Annual ACM Symposium on Theory of Computing (STOC), 755–764, 2010.

[RS09] P. Raghavendra, D. Steurer. *Integrality gaps for*

strong SDP relaxations of Unique Games In Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science (FOCS), 575–585, 2009.

[Saz68] V. V. Sazonov. *On the Multi-Dimensional Central Limit Theorem.* The Indian Journal of Statistics, Series A, Vol. 30, No. 2, 181–204, 1968.

[ST13] D.A. Spielman, S.-H. Teng. *A local clustering algorithm for massive graphs and its applications to nearly-linear time graph partitioning.* SIAM Journal on Computing 42(1), 1–26, 2013.

[Tre12] L. Trevisan. *Max cut and the smallest eigenvalue.* SIAM Journal on Computing 41(6), 1769–1786, 2012.

[Wol07] P. Wolff. *Hypercontractivity of simple random variables.* Studia Math 180, 219–236, 2007.

[ZLM13] Z.A. Zhu, S. Lattanzi, V. Mirrokni. *Local graph clustering beyond Cheeger's inequality.* In Proceedings of the 30th International Conference on Machine Learning (ICML), 396–404, 2013.