Gap Amplification for Small-Set Expansion via Random Walks*

Prasad Raghavendra and Tselil Schramm

University of California, Berkeley
{prasad,tschramm}@cs.berkeley.edu

Abstract

In this work, we achieve gap amplification for the Small-Set Expansion problem. Specifically, we show that an instance of the Small-Set Expansion Problem with completeness ϵ and soundness $\frac{1}{2}$ is at least as difficult as Small-Set Expansion with completeness ϵ and soundness $f(\epsilon)$, for any function $f(\epsilon)$ which grows faster than $\sqrt{\epsilon}$. We achieve this amplification via random walks – the output graph corresponds to taking random walks on the original graph. An interesting feature of our reduction is that unlike gap amplification via parallel repetition, the size of the instances (number of vertices) produced by the reduction remains the same.

1998 ACM Subject Classification F.2 Analysis of Algorithms and Problem Complexity

Keywords and phrases Gap amplification, Small-Set Expansion, random walks, graph products, Unique Games

Digital Object Identifier 10.4230/LIPIcs.APPROX-RANDOM.2014.381

1 Introduction

The small-set expansion problem refers to the problem of approximating the edge expansion of small sets in a graph. Formally, given a graph $G = (V, E)$ and a subset of vertices $S \subseteq V$ with $|S| \leq |V|/2$, the edge expansion of S is

$$\phi(S) = \frac{E(S, \bar{S})}{\mathrm{vol}(S)},$$

where $\mathrm{vol}(S)$ refers to the fraction of all edges of the graph that are incident on the subset S. The edge expansion of the graph G is given by $\phi_G = \min_{S \subseteq V, \mathrm{vol}(S) \leq 1/2} \phi(S)$. The problem of approximating the value of ϕ_G is the well-studied uniform sparsest cut problem [10, 4, 2].

In the small-set expansion problem, the goal is to approximate the edge expansion of the graph at a much finer granularity. Specifically, for $\delta > 0$ define the parameter $\phi_G(\delta)$ as follows:

$$\phi_G(\delta) = \min_{S \subseteq V, \mathrm{vol}(S) \leq \delta} \phi(S).$$

The problem of approximating $\phi_G(\delta)$ for all $\delta > 0$ is the small-set expansion problem.

The small-set expansion problem has received considerable attention in recent years due to its close connections to the unique games conjecture. To describe this connection, we will define a gap version of the problem.

* Prasad Raghavendra is supported by NSF Career Award and Alfred Sloan P. Fellowship. Tselil Schramm is supported by a Berkeley Chancellor's Fellowship and the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE 1106400.

17th Int'l Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX'14) / 18th Int'l Workshop on Randomization and Computation (RANDOM'14).
Editors: Klaus Jansen, José Rolim, Nikhil Devanur, and Cristopher Moore; pp. 381–391
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

▶ **Definition 1.** For constants $0 < s < c < 1$ and $\delta > 0$, the $SSE_\delta(c, s)$ problem is defined as follows: Given a graph $G = (V, E)$ distinguish between the following two cases:

- G has a set S with $\mathrm{vol}(S) \in [\delta/2, \delta]$ with expansion less than $1 - c$
- All sets S with $\mathrm{vol}(S) \le \delta$ in G have expansion at least $1 - s$.

We will omit the subscript δ and write $SSE(c, s)$ when we refer to the $SSE_\delta(c, s)$ problem for all constant $\delta > 0$.

Recent work by Raghavendra and Steurer [13] introduced the following hardness assumption and showed that it implies the unique games conjecture.

▶ **Hypothesis 2.** *For all $\epsilon > 0$, there exists $\delta > 0$ such that $SSE_\delta(1 - \epsilon, \epsilon)$ is NP-hard.*

▶ **Theorem 3** ([13])**.** *The small set expansion hypothesis implies the unique games conjecture.*

Moreover, the small set expansion hypothesis is shown to be equivalent to a variant of the Unique Games Conjecture wherein the input instance is promised to be a small-set expander [14]. Assuming the small-set expansion hypothesis, hardness results have been obtained for several problems including Balanced Separator, Minimum Linear Arrangement [14] and the problem of approximating vertex expansion [11].

In this work, we will be concerned with gap amplification for the small set expansion problem. Gap amplification refers to an efficient reduction that takes a weak hardness result for a problem Π with a small gap between the completeness and soundness and produces a strong hardness with a much larger gap. Formally, this is achieved via an efficient reduction from instances of problem Π to *harder* instances of the same problem Π. Gap amplification is a crucial step in proving hardness of approximation results. An important example of gap amplification is the parallel repetition of 2-prover 1-round games or Label Cover. Label cover is a constraint satisfaction problem which is the starting point for a large number of reductions in hardness of approximation [7]. Starting with the PCP theorem, one obtains a weak hardness for label cover with a gap of 1 vs $1 - \beta_0$ for some tiny absolute constant β_0 [3]. Almost all label-cover based hardness results rely on the much stronger 1 vs ϵ hardness for label cover obtained by gap amplification via the parallel repetition theorem of Raz [16]. More recently, there have been significant improvements and simplifications to the parallel repetition theorem [15, 8, 5].

It is unclear if parallel repetition could be used for gap amplification for small set expansion. Given a graph G, the parallel repetition of G would consist of the product graph G^R for some large constant R. Unfortunately, the product graph G^R can have small non-expanding sets even if G has no small non-expanding sets. For instance, if G has a balanced cut then G^R could have a non-expanding set of volume $\frac{1}{2^R}$.

In this work, we show that random walks can be used to achieve gap amplification for small set expansion. Specifically, given a graph G the gap amplification procedure constructs G^t on the same set of vertices as G, but with edges corresponding to t-step lazy random walks in G. Using this approach, we are able to achieve the following gap amplification.

▶ **Theorem 4.** *Let f be any function such that $\lim_{\epsilon \to 0} \frac{f(\epsilon)}{\sqrt{\epsilon}} \to \infty$. Then:*

If for all $\epsilon > 0$, $SSE'(1 - \epsilon, 1 - f(\epsilon))$ is NP-hard then for all $\eta > 0$, $SSE(1 - \eta, 1/2)$ is NP-hard.

We remark here that the result has some discrepancy in the set sizes between the original instance and the instance produced by the reduction. For this reason, the reduction has to start with a slightly different version of the Small set expansion problem SSE' (See Definition 10).

The above result nicely complements the gap amplification result for the closely related problem of Unique Games obtained via parallel repetition [15]. For the sake of completeness we state the result below.

▶ **Theorem 5** ([15]). *Let f be any function such that $\lim_{\epsilon \to 0} \frac{f(\epsilon)}{\sqrt{\epsilon}} \to \infty$. Then:*
If for all $\epsilon > 0$ if $UG(1-\epsilon, 1-f(\epsilon))$ is NP-hard then for all $\eta > 0$, $UG(1-\eta, 1/2)$ is NP-hard.

Note that the size of the instance produced by our reduction remains bounded by $O(n^2)$. In fact, the instance produced has the same number of vertices but possibly many more edges. This is in contrast to parallel repetition wherein the size of the instance grows exponentially in the number of repetitions used.

Technically, the proof of the result is very similar to an argument in the work of Arora, Barak and Steurer [1] to show that graphs with sufficiently high threshold rank cannot be small-set expanders (see Steurer's thesis [17] for an improved version of the result). The work of O'Donnell and Wright [12] recast these arguments using continuous-time random walks instead of lazy-random walks, yielding cleaner and more general proofs. In this work, we will reuse the proof technique and obtain upper and lower bounds for the expansion profile of lazy random walks (see Theorem 11). These upper and lower bounds immediately imply the desired gap amplification result for small-set expansion.

Subsequent to our work, Kwok and Lau [9] have obtained a stronger analysis of our gap amplification theorem, yielding almost tight bounds.

2 Preliminaries

Unless otherwise specified, we will be concerned with an undirected graph $G = (V, E)$ with n vertices and associated edge weights $w : E \to \mathbb{R}^+$. The degree of vertex i denoted by $d(i) = \sum_{(i,j)\in E} w(i,j)$. The volume of a set $S \subseteq V$ is defined to be $\mathrm{vol}(S) = \sum_{i\in S} d(i)$. Henceforth, we will assume that the total volume is 1, i.e., $\sum_{i\in V} d(i) = 1$. The adjacency matrix A of the graph G has entries $A_{ij} = w(i,j)$. The degree matrix D is a $n \times n$ diagonal matrix with $D_{ii} = d(i)$.

2.1 Expansion Profile

The expansion profile of a graph is defined as follows.

▶ **Definition 6.** For a graph G, define the expansion profile $\phi_G : \mathbb{R}^+ \to [0,1]$ as

$$\phi_G(\delta) = \min_{S\subseteq V, \mathrm{vol}(S)\leq\delta} \phi(S)$$

where $\phi(S) = \frac{E(S,\bar{S})}{\mathrm{vol}(S)}$.

2.2 Lazy Random Walks

The transition matrix for a lazy random walk on G is given by

$$M = \tfrac{1}{2}(I + D^{-1}A)\,.$$

The lazy random walk corresponds to staying at the same vertex with probability $\frac{1}{2}$, and moving to a random neighbor with probability $\frac{1}{2}$. We will let G^t denote the graph corresponding to the t-step lazy random walk. The adjacency matrix of G^t is given by DM^t.

We recall a few standard facts about lazy random walks here.

▶ **Fact 7.** *If G is a graph with adjacency matrix A, then G's lazy random walk operator $M = \frac{1}{2}(I + D^{-1}A)$ has the property that $\|D^{1/2}Mv\|_2^2 = v^T DM^2 v$ for any vector v.*

Proof. We use the fact that $M = \frac{1}{2}D^{-1/2}(I + D^{-1/2}AD^{-1/2}))D^{1/2}$:

$$\begin{aligned}\|D^{1/2}Mv\|_2^2 &= \tfrac{1}{4}v^T M^T DMv \\ &= \tfrac{1}{4}v^T D^{1/2}(I + D^{-1/2}AD^{-1/2})D^{-1/2}DD^{-1/2}(I + D^{-1/2}AD^{-1/2})D^{1/2}v \\ &= \tfrac{1}{4}v^T D^{1/2}(I + D^{-1/2}AD^{-1/2})^2 D^{1/2}v \\ &= v^T DM^2 v,\end{aligned}$$

as desired. ◀

▶ **Fact 8.** *If G is a graph with adjacency matrix A, then for the lazy random walk operator $M = \frac{1}{2}(I + D^{-1}A)$, we have*

$$\|D^{1/2}v\|_2^2 = v^T Dv \geq v^T DMv \geq v^T DM^2 v = \|D^{1/2}Mv\|_2^2.$$

Proof. Since the eigenvalues λ_i of $D^{-1/2}AD^{-1/2}$ are between $[-1, 1]$, the eigenvalues of $M' = \frac{1}{2}(I + D^{-1/2}AD^{-1/2})$ are $\mu_i = \frac{1}{2}(1 + \lambda_i)$, and so $\mu_i \in [0, 1]$. Let $D^{1/2}v = \sum \alpha_i u_i$ be the decomposition of $D^{1/2}v$ in terms of the eigenvectors of M'. Then we have

$$D^{1/2}Mv = M'D^{1/2}v = \sum \alpha_i \mu_i u_i,$$

and so $v^T Dv = \sum \alpha_i^2$, $v^T DMv = \sum \alpha_i^2 \mu_i$, and $v^T DM^2 v = \sum \alpha_i^2 \mu_i^2$. Since $\mu_i \in [0, 1]$, we have $v^T Dv \geq v^T DMv \geq v^T DM^2 v$, as desired. ◀

▶ **Fact 9.** *For the lazy random walk operator $M = \frac{1}{2}(I + D^{-1}A)$ and any vector $v \in \mathbb{R}^V$, $v \geq 0$ we have*

$$\|Dv\|_1 = \|DMv\|_1.$$

Proof. Let $v \in \mathbb{R}^V$. We have

$$\|DMv\|_1 = 1^T D(\tfrac{I+D^{-1}A}{2})v \;=\; \tfrac{1}{2}((1^T D)v + (1^T A)v) \;=\; \|Dv\|_1,$$

where the last inequality follows because $1^T D = 1^T A$. ◀

2.3 Small-Set Expansion Problem

The formal statement of the SSE' problem is as follows.

▶ **Definition 10.** For constants $0 < s < c < 1$ and $\delta > 0$, the Small-Set Expansion problem $SSE'_\delta(c, s)$ is defined as follows: Given a graph $G = (V, E)$, distinguish between the following two cases:

- G contains a set S such that $\text{vol}(S) \in [\delta/2, \delta]$ and $\phi(S) \leq 1 - c$
- All sets S with $\text{vol}(S) \leq \mathbf{8}\delta$ in G have expansion $\phi(S) \geq 1 - s$.

The key difference from $SSE_\delta(c, s)$ is that the soundness is slightly stronger in that even sets of size 8δ have expansion at least $1 - s$.

2.4 Organization

In Section 3, we will obtain upper and lower bounds (Theorem 11) for expansion profile of lazy random walks. Subsequently, we use these bounds to conclude the main result of the paper in Section 4. In Appendix A, we give a reduction that establishes the equivalence of the search versions of two different notions of Small-Set Expansion. Finally, we also present a reduction from SSE on irregular graphs to SSE on regular graphs in Appendix B.

3 Expansion Profile of Lazy Random Walks

Let $G = (V, E)$ be a graph with adjacency matrix A, and diagonal degree matrix D. The transition matrix for a lazy random walk on G is $M = \frac{1}{2}(I + D^{-1}A) = \frac{1}{2}D^{-1/2}(I + D^{-1/2}AD^{-1/2})D^{1/2}$.

For every $t \in \mathbb{N}$, let G^t denote the graph corresponding to the t-step lazy random walk whose adjacency matrix is given by DM^t. We will prove the following theorem about the expansion profile of G^t.

▶ **Theorem 11.** *For all $t \in \mathbb{N}$ and $\eta, \delta \in (0, 1]$, if G^t denotes the graph corresponding to the t-step lazy random walk in a graph $G = (V, E)$ then,*

$$\min\left(1 - \left(1 - \frac{\phi_G^2(\frac{4\delta}{\eta})}{32}\right)^t, 1 - \eta\right) \le \phi_{G^t}(\delta) \le \frac{t}{2} \cdot \phi_G(\delta)\,.$$

We will split the proof of the above theorem in to two parts: Lemma 12 and Lemma 13

▶ **Lemma 12.** *For every subset $S \subseteq V$,*

$$\phi_{G^t}(S) \le \frac{t}{2} \cdot \phi_G(S),$$

and therefore $\phi_{G^t}(\delta) \le \frac{t}{2} \cdot \phi_G(\delta)$.

Proof. Fix a subset $S \subset V$. From [6], we have that the probability $p(t)$ that a lazy random walk stays entirely in S for t steps is bounded below by

$$p(t) \ge \left(1 - \frac{1}{2}\phi(S)\right)^t.$$

Now, the expansion of S in G^t is the probability of leaving the set on the tth step of the random walk, which is at most $1 - p(t)$. Hence,

$$\phi_{G^t}(S) \le 1 - p(t) \le 1 - \left(1 - \frac{1}{2}\phi(S)\right)^t \le \frac{t}{2}\phi(S),$$

as desired. The result immediately follows for all sets of volume $\le \delta$. ◀

▶ **Lemma 13.** *For all t, η,*

$$\phi_{G^t}(\delta) \ge \min\left(1 - \left(1 - \frac{\phi_G^2(\frac{4\delta}{\eta})}{32}\right)^t, 1 - \eta\right).$$

We prove this lemma by contradiction, by showing that if the expansion in the final graph is not large enough then there exists a vector with bounded Rayleigh quotient with respect to the original graph, from which we can extract a non-expanding set. The intuition is that the expansion of a set in the final graph DM^t corresponds to the neighborhood of the random walk after t steps, and if the neighborhood is not large enough after t steps, there must be at least one step (or application of M) during which it did not grow.

Proof. Suppose by way of contradiction that this is not the case. Let $\beta = \phi_G(\frac{4\delta}{\eta})$ and let $\delta' = \frac{4\delta}{\eta}$. Further, let $\hat{\beta} = \frac{1}{2}\beta$.

Let S be a set of volume at most $\delta \cdot \mathrm{vol}(V)$ such that

$$\phi_{G^t}(S) \leq \min\left(1 - \left(1 - \frac{\hat{\beta}^2}{8}\right)^t, 1 - \eta\right). \tag{1}$$

Let $v_0 = 1_S$ be the vector corresponding to the indicator function of the set S. Define $v_i = M^i v_0$, and for the diagonal degree matrix D of A, define $w_i = D^{1/2} v_i$. Note that $\|w_0\|_2^2 = \mathrm{vol}(S)$, and $\|Dv_0\|_1 = \mathrm{vol}(S)$. By Fact 9 we also have $\|Dv_i\|_1 = \mathrm{vol}(S)$ for all i.

We first lower-bound $\|w_{\frac{t}{2}}\|_2$. By definition of expansion,

$$\phi_{G^t}(S) = 1 - \frac{v_0^T D M^t v_0}{v_0^T D v_0}$$

which by Fact 7 implies that $\|D^{1/2} M^{\frac{t}{2}} v_0\|_2^2 = \mathrm{vol}(S)(1 - \phi_{G^t}(S))$. Now, using (1) we get

$$\|w_{\frac{t}{2}}\|_2^2 \ = \ \|D^{1/2} M^{\frac{t}{2}} v_0\|_2^2 = \mathrm{vol}(S)(1 - \phi_{G^t}(S)) \ \geq \ \mathrm{vol}(S) \cdot \max\left(\eta, (1 - \frac{1}{8}\hat{\beta}^2)^t\right) \tag{2}$$

By Fact 8, we have $\|w_i\|_2 \geq \|w_{i+1}\|_2 \geq 0$ for all i, and (2) holds for all $i \leq \frac{t}{2}$.

We now assert that there must be some i for which

$$\frac{\|w_{i+1}\|_2^2}{\|w_i\|_2^2} > 1 - \tfrac{1}{4}\hat{\beta}^2.$$

To see this, consider the product of all such terms for $i < \frac{t}{2}$. Some algebraic simplification shows that

$$\prod_{i=0}^{\frac{t}{2}-1} \frac{\|w_{i+1}\|_2^2}{\|w_i\|_2^2} \ = \ \frac{\|w_{\frac{t}{2}}\|_2^2}{\|w_0\|_2^2} \ > \ \frac{(1 - \frac{1}{8}\hat{\beta}^2)^t \cdot \mathrm{vol}(S)}{\mathrm{vol}(S)} \ = \ \left(1 - \tfrac{1}{8}\hat{\beta}^2\right)^t,$$

where the second-to-last inequality follows from (2). Thus for some $i < \frac{t}{2}$ we have

$$\frac{\|w_{i+1}\|_2^2}{\|w_i\|_2^2} \ > \ \left((1 - \frac{1}{8}\hat{\beta}^2)^t\right)^{\frac{2}{t}} \ > \ 1 - \frac{1}{4}\hat{\beta}^2.$$

Then let w_i be the vector corresponding to the first i for which $\|w_{i+1}\|_2^2 \geq (1 - \frac{1}{4}\hat{\beta}^2)\|w_i\|_2^2$.

Since w_{i+1} is obtained from v_i via one step of a lazy random walk and a normalization, we can bound the Rayleigh quotient of v_i with respect to the Laplacian of $DM = \frac{1}{2}(D + A)$:

$$\frac{v_i^T D(I - M) v_i}{v_i^T D v_i} = 1 - \frac{v_i^T D M v_i}{v_i^T D v_i},$$

by Fact 8,

$$\leq 1 - \frac{v_i^T D M^2 v_i}{v_i^T D v_i}$$

and by Fact 7,

$$\begin{aligned} &= 1 - \frac{\|w_{i+1}\|_2^2}{\|w_i\|_2^2} \\ &\leq \frac{1}{4}\hat{\beta}^2. \end{aligned} \tag{3}$$

We now truncate the vector v_i, then run Cheeger's algorithm on the truncated vector in order to find a non-expanding small set, and thus obtain a contradiction. Let $\theta = \frac{\eta}{4}$. We take the truncated vector

$$z_i(j) = \begin{cases} v_i(j) - \theta & v_i(j) \geq \theta \\ 0 & \text{otherwise} \end{cases} .$$

By Fact 9, Dv_i has L_1 mass $\mathrm{vol}(S)$. Thus, the total volume of the set S_z of vertices with nonzero support in z_i is

$$\mathrm{vol}(S_z) \;=\; \sum_{v_i(j)>\theta} d(j) \;\leq\; \sum_{v_i(j)>\theta} \frac{1}{\theta} d(j) v_i(j) \;\leq\; \frac{1}{\theta} \cdot \|Dv_i\|_1 = \frac{4\,\mathrm{vol}(S)}{\eta}$$

Hence any subset of S_z has volume at most $\frac{4\,\mathrm{vol}(S)}{\eta}$.

For the vector v_i, we know that $\|Dv_i\|_1 = \mathrm{vol}(S)$. Moreover using (2),

$$\|D^{1/2}v_i\|_2^2 = \|w_i\|_2^2 \geq \|w_{t/2}\|_2^2 \geq \eta\,\mathrm{vol}(S)\,.$$

Applying Lemma 14 to v_i and z_i to conclude,

$$\frac{z_i^T D(I-M) z_i}{z_i^T D z_i} \leq 2 \frac{v_i^T D(I-M) v_i}{v_i^T D v_i}\,.$$

Using (3), this implies the following bound on the Rayleigh quotient of z_i,

$$\frac{z_i^T D(I-M) z_i}{z_i^T D z_i} \;\leq\; \frac{1}{2}\hat{\beta}^2\,.$$

Thus, when we run Cheeger's algorithm on z_i, we get a set of volume at most $\frac{4\,\mathrm{vol}(S)}{\eta}$ and of expansion less than $\hat{\beta}$ in DM, and therefore less than β in G. Since $\beta = \phi_G(\frac{4\delta}{\eta})$, this is a contradiction. This completes the proof of Lemma 13. ◀

The following lemma, which gives an upper bound on the Rayleigh quotient of a truncated vector, is a slight generalization of Lemma 3.4 of [1].

▶ **Lemma 14.** *Let $x \in \mathbb{R}^V$ be non-negative, let L be the weighted Laplacian of a graph $G = (V, E)$ with weights $w(i,j)$ and degree matrix D. Suppose that*

$$4\theta\|Dx\|_1 \leq \|D^{1/2}x\|_2^2 \tag{4}$$

Then for the threshold vector y defined by

$$y(i) = \begin{cases} x(i) - \theta & x(i) > \theta \\ 0 & \textit{otherwise} \end{cases} ,$$

we have

$$\frac{y^T L y}{y^T D y} \;\leq\; 2 \cdot \frac{x^T L x}{x^T D x}.$$

Proof. First, we show $y^T L y \leq x^T L x$.

$$\begin{aligned} y^T L y &= \sum_{(i,j)\in E} w(i,j)(y(i)-y(j))^2 \\ &= \sum_{\substack{(i,j)\in E \\ y(i),y(j)\geq 0}} w(i,j)(x(i)-x(j))^2 + \sum_{\substack{(i,j)\in E \\ y(i)\geq 0, y(j)=0}} w(i,j)(x(i)-\theta)^2 \\ &\leq \sum_{(i,j)\in E} w(i,j)(x(i)-x(j))^2 \\ &= x^T L x, \end{aligned}$$

where the second-to-last inequality follows from the fact that if $y(i) = 0$, then $x(i) \leq \theta$.

Now, we show that $y^T Dy \geq \frac{1}{2} x^T Dx$. First, we note that $d(i)y(i)^2 \geq d(i)x(i)^2 - 2\theta d(i)x(i)$ for all k. Thus,

$$\begin{aligned}
\sum_{i \in V} d(i)y(i)^2 &\geq \sum_{i \in V} d(i)x(i)^2 - 2\theta d(i)x(i) \\
&= \left(\sum_{i \in V} d(i)x(i)^2\right) - 2\theta \left(\sum_{i \in V} d(i)x(i)\right) \\
&\geq \frac{1}{2} \sum_{i \in V} d(i)x(i)^2 .
\end{aligned}$$

Where the the last inequality follows by assumption (4).

Thus, we have

$$\frac{y^T Ly}{y^T Dy} \leq 2 \cdot \frac{x^T Lx}{x^T Dx},$$

as desired. ◀

4 Gap Amplification

In this section, we will prove Theorem 4 which we restate here for convenience.

▶ **Theorem 15** (Restatement of Theorem 4)**.** *Let f be any function such that $\lim_{\epsilon \to 0} \frac{f(\epsilon)}{\sqrt{\epsilon}} \to \infty$. Then:*

If for all $\epsilon > 0$, $SSE'(1 - \epsilon, 1 - f(\epsilon))$ is NP-hard then for all $\eta > 0$ $SSE(1 - \eta, \frac{1}{2})$ is NP-hard.

Proof. Fix ϵ small enough so that $\frac{64\epsilon}{f(\epsilon)^2} \leq \eta$. There exists such an ϵ since $\lim_{\epsilon \to 0} \frac{f(\epsilon)}{\sqrt{\epsilon}} \to \infty$. Fix $t = \frac{64}{f(\epsilon)^2}$.

Given an instance G of $SSE'(1 - \epsilon, 1 - f(\epsilon))$, the reduction just outputs the graph G^t obtained via t-step lazy random walks on G. Since the adjacency matrix of G' can be calculated with $\log t$ matrix multiplications, this reduction clearly runs in time $O(n^3 \log t)$.

Completeness. If there exists a set of S with $\mathrm{vol}(S) \in [\delta/2, \delta]$ and $\phi_G(S) \leq \epsilon$ then by Lemma 12 the same set S satisfies,

$$\phi_{G^t}(S) \leq \frac{t}{2} \phi_G(S) = \Theta\left(\frac{\epsilon}{f(\epsilon)^2}\right) \leq \eta \,.$$

Soundness. If $\phi_G(8\delta) \geq f(\epsilon)$ then by applying Lemma 13

$$\phi_{G^t}(\delta) \geq \min\left(1 - \left(1 - \frac{1}{32} f(\epsilon)^2\right)^t, 1/2\right) \geq \frac{1}{2} \,.$$

◀

References

1 Sanjeev Arora, Boaz Barak, and David Steurer. Subexponential algorithms for unique games and related problems. In *FOCS*, pages 563–572, 2010.

2 Sanjeev Arora, James R. Lee, and Assaf Naor. Euclidean distortion and the sparsest cut. *Journal of American Mathematical Society*, 21:1–21, 2008.

3 Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *JACM: Journal of the ACM*, 45, 1998.

4 Sanjeev Arora, Satish Rao, and Umesh Vazirani. Expander flows, geometric embeddings and graph partitioning. In *Proceedings of the thirty-sixth annual ACM Symposium on Theory of Computing (STOC-04)*, pages 222–231, New York, June 13–15 2004. ACM Press.

5 Irit Dinur and David Steurer. Analytical approach to parallel repetition. *CoRR*, abs/1305.1979, 2013.

6 Shayan Oveis Gharan and Luca Trevisan. Approximating the expansion profile and almost optimal local graph clustering. In *FOCS*, pages 187–196, 2012.

7 Johann Håstad. Some optimal inapproximability results. *Journal of the ACM*, 48(4):798–859, 2001.

8 Holenstein. Parallel repetition: Simplifications and the no-signaling case. In *STOC: ACM Symposium on Theory of Computing (STOC)*, 2007.

9 Tsz Chiu Kwok and Lap Chi Lau. Personal communication, 2014.

10 Frank Thomson Leighton and Satish Rao. Multicommodity max-flow min-cut theorems and their use in designing approximation algorithms. *J. ACM*, 46(6):787–832, 1999.

11 Anand Louis, Prasad Raghavendra, and Santosh Vempala. The complexity of approximating vertex expansion. *CoRR*, abs/1304.3139, 2013.

12 Ryan O'Donnell and David Witmer. Markov chain methods for small-set expansion. *Arxiv arXiv:1204.4688*, 2012.

13 Prasad Raghavendra and David Steurer. Graph expansion and the unique games conjecture. In *STOC*, pages 755–764, 2010.

14 Prasad Raghavendra, David Steurer, and Madhur Tulsiani. Reductions between expansion problems. In *IEEE Conference on Computational Complexity*, pages 64–73, 2012.

15 Anup Rao. Parallel repetition in projection games and a concentration bound. In Richard E. Ladner and Cynthia Dwork, editors, *Proceedings of the 40th Annual ACM Symposium on Theory of Computing (STOC'08)*, pages 1–10. ACM, 2008.

16 Ran Raz. A parallel repetition theorem. *SIAM Journal on Computing*, 27(3):763–803, June 1998.

17 David Steurer. *On the Complexity of Unique Games and Graph Expansion.* PhD thesis, Princeton University, 2010.

A Equivalence of Two Notions of the Small-Set Expansion Problem

There is a slightly different version of the Small-Set expansion decision problem that differs from Definition 1 in the soundness case.

▶ **Definition 16.** For constants $0 < s < c < 1$, and $\delta > 0$, the Small-Set Expansion problem $SSE^{=}_{\delta}(c, s)$ is defined as follows: Given a graph $G = (V, E)$ with $\mathrm{vol}(V) = N$, distinguish between the following two cases:

- G has a set of volume in the range $[\frac{1}{2}\delta N, \delta N]$ with expansion less than $1 - c$
- All sets in G of volume in the range $[\frac{1}{4}\delta N, \delta N]$ have expansion at least $1 - s$.

Clearly $SSE^{=}_{\delta}(c,s)$ is a harder decision problem than $SSE_{\delta}(c,s)$ since the soundness assumption is weaker. There is no known reduction from $SSE^{=}_{\delta}(c,s)$ to $SSE_{\delta}(c,s)$ that establishes the equivalence of the two versions. Here we observe that the search versions of these two problems are equivalent.

▶ **Proposition 17.** *For all $\delta_0, c, s > 0$ A search algorithm for $SSE_{\delta}(2c-1,s)$ for $\delta \in [\delta_0/2, \delta_0]$ gives a search algorithm for $SSE^{=}_{\delta}(c,s)$ in the range $\delta \in [\delta_0/2, \delta_0]$.*

Proof. Suppose we are given an algorithm A that finds a set S' of volume at most δN and expansion less than $1-s$ whenever there exists a set S with $\mathrm{vol}(S) \in [\frac{1}{4}\delta N, \delta N]$ and $\Phi(S) \leq 2-2c$. We construct a set $S \subseteq V$ such that $\mathrm{vol}(S) \in [\frac{1}{4}\delta N, \delta N]$ and $\phi(S) < 1-s$. We proceed iteratively, as follows.

We start with an empty initial set, S_{out}, and with the full graph, $G_0 = G$. If $\mathrm{vol}(S_{out}) \in [\frac{1}{4}\delta N, \delta N]$, we terminate and return S_{out}. Otherwise, at the ith step, we apply A to G_{i-1} to obtain a set S_i of expansion less than $1-s$. If $\mathrm{vol}(S_i) \in [\frac{1}{4}\delta N, \delta N]$ return S_i, otherwise add the vertices in S_i to S_{out}. We then set $G_i = G_{i-1} \setminus S_i$. If no such set can be found, then we terminate and return no.

Clearly, this algorithm terminates and runs in polynomial time. Suppose S' is a non-expanding set with $\mathrm{vol}(S') \in [\frac{1}{2}\delta N, \delta N]$. As long as S_{out} has volume smaller than $\frac{1}{4}\delta N$, $S' - S_{out}$ will have volume at least $\mathrm{vol}(S')/2$ and has expansion at most $2\phi(S') \leq 2-2c$. Hence by the assumption about algorithm A, it will return a set S_i of expansion at most $1-s$. The check of the volume of S_i ensures that S_{out} will never go from below the allowable volume range to above in a single step. Finally if S_i was never returned for any step i, the union of all the sets S_i has expansion at most $1-s$ and volume in the range $[\delta N/4, \delta N]$. ◀

B Reduction from Irregular Graphs to Regular Graphs

In this section, we present a reduction from small set expansion on irregular graphs to small set expansion on regular graphs. Specifically, we prove the following theorem.

▶ **Theorem 18.** *There exists an absolute constant C such that for all $\gamma, \beta \in (0,1)$ there is a polynomial time reduction from $SSE_{\delta}(1-\gamma, 1-\beta)$ on a irregular graph $G = (V,E)$ to $SSE_{\delta}(1-\gamma, 1-{}^{\beta}/_{C})$ on a 4-regular graph $G' = (V', E')$*

Proof. The reduction is as follows: we replace each vertex $v \in V$ with a 3-regular expander A_v on $\deg(v)$ vertices. Using standard constructions of 3-regular expanders, we can assume that the graphs A_v have edge expansion at least $\kappa = 0.01$. Now, for each edge $(v,w) \in E$, we add an edge between a particular vertex in A_v and A_w. The resulting graph on the expanders is G', with $V' = \cup_{v\in V} A_v$. Note that G' is d-regular, and that $|V'| = \sum_{v\in V} \deg(v) = \mathrm{vol}(V)$, as desired.

For the completeness, we note that if a set $S \subset V$ with volume at most $\delta|V|$ has $\phi_G(S) < \gamma$, then the set $S' = \cup_{v\in S} A_v$ has the same number of edges leaving the set as S, and the number of vertices in the set is equal to $\mathrm{vol}(S)$. Thus, $\phi_{G'}(S') < \gamma/4$, as desired.

For soundness, suppose there is a set $S' \subset V'$ with $|S'| \leq \delta|V'|$ and $\phi_{G'}(S') < \beta$. Then we can partition S' into sets corresponding to each A_v; let $B_v = S' \cap A_v$. Then consider the set

$$S^* = \cup_{|B_v| \geq \frac{1}{2}|A_v|} A_v,$$

the set of A_v that overlap with S' by at least half. We will argue that S^* has expansion at most $\frac{10}{\kappa}\beta$ in G'. First, by definition of expansion we have

$$\beta \geq \phi_{G'}(S') = \frac{\sum_v E(B_v, \bar{S}')}{4\sum_{v\in V}|B_v|} = \frac{\sum_v E(B_v, A_v \setminus B_v) + E(B_v, \bar{S}' \setminus A_v)}{4\sum_{v\in V}|B_v|},$$

where we distinguish between boundary edges of S' inside and outside of the A_v. In particular, we have

$$4\beta \sum_{v\in V}|B_v| \geq \sum_{v\in V} E(B_v, A_v \setminus B_v).$$

Now, we bound from below the number of boundary edges within A_v. Since A_v is an expander with expansion κ, we have

$$E(B_v,\ A_v \setminus B_v) \geq \kappa \cdot \min(|B_v|, |A_v \setminus B_v|).$$

Hence we will have,

$$S'\Delta S^* = \sum_v \min(|B_v|, |A_v \setminus B_v|) \leq \frac{1}{\kappa}\sum_v E(B_v, A_v \setminus B_v) \leq \frac{4\beta}{\kappa}\sum_{v\in V}|B_v| = \frac{4\beta}{\kappa}|S'|\,.$$

Since G' is a 4-regular graph, we can upper bound the expansion of S^* by

$$\phi_{G'}(S^*) \leq \frac{E[S', \bar{S}'] + 4|S'\Delta S^*|}{4|S'| - 4|S'\Delta S^*|} \leq \frac{4\beta|S'| + 16\beta/\kappa|S'|}{4|S'| - 16\beta/\kappa|S'|} \leq \frac{\beta\,(1 + 4/\kappa)}{1 - 4\beta/\kappa}\,.$$

Thus, in G the set $S = \{v \mid A_v \in S^*\}$ has expansion at most $\frac{10}{\kappa}\beta$, and $\mathrm{vol}(S) \in [\frac{1}{2}\delta\,\mathrm{vol}(V), 2\delta\,\mathrm{vol}(V)]$, as desired. ◀