Testing Small Set Expansion in General Graphs*

Angsheng Li¹ and Pan Peng¹,²

¹State Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences
²Department of Computer Science, Technische Universität Dortmund

angsheng@iso.ac.cn pan peng@tu-dortmund.de

Abstract

We consider the problem of testing small set expansion in general graphs. A graph $G$ is a $(k, \phi)$-expander if every subset of volume at most $k$ has conductance at least $\phi$. Small set expansion has recently received significant attention due to its close connection to the unique games conjecture and the local graph partitioning algorithms.

Our main result is a two-sided error tester for small set expansion in a model which we call rotation map model that allows degree queries and neighbor & index queries to access $G$. The tester takes as input an $n$-vertex graph $G$, a volume bound $k$, an expansion bound $\phi$ and a distance parameter $\epsilon > 0$, and with probability at least $2/3$, accepts the graph if it is a $(k, \phi)$-expander and rejects the graph if it is $\epsilon$-far from any $(k^*, \phi^*)$-expander, where $k^* = \Theta(k\epsilon)$ and $\phi^* = \Theta((\log n)^2 \epsilon^{-2})$. The query complexity and running time of the tester is $\tilde{O}(\sqrt{n\phi^{-2}\epsilon^{-2}})$.

As a byproduct of our analysis for the two-sided error tester, we give a one-sided error tester for the adjacency list model, that allows degree queries and neighbor queries, by invoking a recent local algorithm for finding small sparse cuts.

1 Introduction

Graph property testing is an effective algorithmic paradigm to deal with real-world networks, the scale of which has become so large that it is even impractical to read the whole input. In the setting of testing a graph property $P$, we are given as input a graph $G$ and we want to design an algorithm (called tester) to distinguish the case that $G$ has property $P$ from the case that $G$ is “far from” the property $P$ with high success probability (say $2/3$). Here, the notion of being “far from” is parameterized by a distance parameter $\epsilon$. In most situations, a graph $G$ is said to be $\epsilon$-far from property $P$ if one has to modify at least an $\epsilon$ fraction of the representation (or edges) of $G$ to obtain a graph $G'$ with property $P$. We assume the input graph $G$ can be accessed through an oracle $O_G$ and the goal is to design property testers that make as few queries as possible to $O_G$.

Since the seminal work of Goldreich and Ron [GR02], many testers have been developed for different graph properties, such as $k$-colorability, bipartiteness, acyclicity, triangle-freeness and many others. Most of these testers apply only to the adjacency matrix model or the adjacency list model, depending on the types of queries the testers are allowed to perform to the oracle. The

---

*Both authors are partially supported by the Grand Project “Network Algorithms and Digital Information” of the Institute of Software, Chinese Academy of Sciences and by a 973 program 2014CB340302. The second author acknowledges the support of ERC grant No. 307696.
former model is most suitable for dense graphs and general characterizations on the testability of a property in this model has been given (e.g., [AFNS09]). The latter model is most suitable for sparse graphs, and several property testers using the techniques of local search or random walks are known, while it is not well understood what properties are testable in constant time in this model. Even less is known about testers, testability results or even models for general graphs (see recent surveys [Ron10, Gol10]).

In this paper, we focus on property testers for general graphs. We will consider three types of queries to the oracle of $G$: the degree query, the neighbor query and the neighbor & index query. For the degree query, when specified a vertex $v$, the oracle returns the degree of $v$; for the neighbor query, when specified a vertex $v$ and an index $i$, the oracle returns the $i$th neighbor of $v$; for the neighbor & index query, when specified a vertex $v$ and an index $i$, the oracle returns a pair $(u,j)$ such that $u$ is the $i$th neighbor of $v$ and $j$ is the index of $u$ as a neighbor of $v$. Accordingly, we consider two models: the rotation map model that allows degree queries and neighbor & index queries to the oracle [PP11]; the adjacency list model that allows degree queries and neighbor queries to the oracle [PR02]. The adjacency list model also applies to the bounded degree graphs with an additional restriction that a fixed upper-bound was assumed on the degrees [GR02].

We study the problem of testing small set expansion for general graphs. Given a graph $G = (V,E)$ with $n$ vertices and $m$ edges, and a set $S \subseteq V$, let the volume of $S$ be the sum of degree of vertices in $S$, that is, $\text{vol}(S) := \sum_{v \in S} \text{deg}_G(v)$, where $\text{deg}_G(v)$ denotes the degree of vertex $v$. Define the conductance of $S$ as $\phi(S) := \frac{\text{cut}(S)}{\text{vol}(S)}$, where $\text{cut}(S,V \setminus S)$ is the number of edges leaving $S$; and define the $k$-expansion profile of $G$ as $\phi(k) := \min_{S: \text{vol}(S) \leq k} \phi(S)$. A graph $G$ is called a $(k,\phi)$-expander if $\phi(k) \geq \phi$, that is, all the subsets in $G$ with volume at most $k$ have conductance at least $\phi$. We will refer to small set expander as $(k,\phi)$-expander and refer to small set expansion as $\phi(k)$.

Besides of the relation to the mixing time of random walks [LK99], small set expansion has been of much interest recently for its close connection to the unique games conjecture [RS10, ABS10], and the design of local graph partitioning algorithms in massive graphs [ST08, ACL06, AP09, OT12, KL12]. Approximation algorithms and spectral characterizations for the small set expansion problem have been studied [ABS10, LRTV12, LOT12, KL12, OT12, OW12]. It is natural to ask if one can efficiently (in sublinear time) test if a graph is a small set expander.

1.1 Our results

We give testers for small set expansion in the general graphs. We use the common definition of distance between graphs. More precisely, a graph $G$ with $m$ edges is said to be $\epsilon$-far from a $(k,\phi)$-expander if one has to modify at least $\epsilon m$ edges of $G$ so that it becomes a $(k,\phi)$-expander.

Our main result is a two-sided error tester in the rotation map model.

Theorem 1. Given as input an $n$-vertex graph, a volume bound $k$, a distance parameter $\epsilon$ and a conductance bound $\phi$, there exists an algorithm that with probability at least $2/3$, accepts any graph that is a $(k,\phi)$-expander, and rejects any graph that is $\epsilon$-far from any $(k^*,\phi^*)$-expander, where $k^* = \Theta(k\epsilon)$ and $\phi^* = \Theta(\frac{\phi^2}{\log n / \log k})$. The running time of the algorithm is $\tilde{O}(\sqrt{n\phi^{-2}\epsilon^{-2}})$.

Note that the running time of the tester matches the best known algorithms for testing the conductance of $G$ which corresponds to the case $k = m$ (see more discussions below).

As a byproduct of our analysis for the above two-sided error tester, we obtain a one-sided error tester (that accepts every $(k,\phi)$-expander) for the adjacency list model by invoking a local algorithm for finding small sparse cuts. We show the following result.
Theorem 2. Given as input an \( n \)-vertex graph, a volume bound \( k \), a conductance bound \( \phi \), and a distance parameter \( \varepsilon \), there exists an algorithm that always accepts any graph that is a \((k, \phi)\)-expander, and with probability at least \( 2/3 \) rejects any graph that is \( \varepsilon \)-far from any \((k^*, \phi^*)\)-expander, where \( k^* = O(k^{1-\xi}) \) and \( \phi^* = O(\xi \phi^2) \) for any \( 0 < \xi < 1 \). Furthermore, whenever it rejects a graph, it provides a certificate that the graph is not a \((k, \phi)\)-expander in the form of a set of volume at most \( k \) and expansion at most \( \phi \). The running time of the algorithm is \( \tilde{O}(k/\varepsilon \phi) \).

Note that \( \xi \) is not necessarily a constant, and the running time of the above algorithm is sublinear in \( n \) for \( k = O(\frac{n}{\log \Omega(1)n}) \) and constant \( \phi \).

1.2 Our techniques

Our two-sided error tester in the rotation map model for general graphs is obtained by combining a (two-sided error) tester in the adjacency list model for bounded degree graphs and a new reduction that (virtually) transfers any graph into a bounded degree graph.

The tester in bounded degree model is very intuitive and simple: we sample a small number of vertices, and for each sampled vertex \( v \), we perform independently a number of random walks from \( v \) and calculate the number of collisions \( Z_v \) between the endpoints of these random walks. We accept the graph if and only if \( Z_v \) is small for every sampled vertex \( v \). This idea originates from the tester for expansion for bounded degree graphs. The main differences between our small set expansion tester and the previous expansion testers is the choice of parameters. However, the tester for bounded degree graphs does not apply to general graph, which may have an arbitrary large degree. For example, in a star graph the collision probability will be very large on the “central” vertex, however, the conductance of star graph is large and it is thus a small set expander. This implies that we cannot directly apply our tester for bounded degree graphs to general graphs. To solve this problem, we introduce a new graph product called non-uniform replacement product that turns every graph into a bounded degree graph, and at the mean time, the expansion profile of these two graphs do not differ too much.

1.3 Other related work

There is an interesting line of research on testing the special case of the \((k, \phi)\)-expander for \( k = m \), which is often abbreviated as \( \phi\)-expander. And \( \phi(m) \) is often abbreviated as the expansion (or conductance) of \( G \). Goldreich and Ron [GR00] have proposed an expansion tester for bounded degree graphs. The tester (with different setting parameters) has later been analyzed by Czumaj and Sohler [CS10], Nachmias and Shapira [NS11], and Kale and Seshadhri [KS11], and it is proven that the tester can distinguish \( d \)-regular \( \phi \)-expanders from graphs that are \( \varepsilon \)-far from any \( d \)-regular \( \Omega(\phi^2) \)-expanders. The running time of the tester is \( O(n^{0.5+\eta}(\varepsilon^{-1} \log n)^{\Omega(1)}) \) for any \( \eta > 0 \), which is almost optimal by a lower bound of \( \Omega(\sqrt{n}) \) given by Goldreich and Ron [GR02]. Li, Pan and Peng [LPP11] give an expansion tester for general graphs that matches the best known tester for bounded degree graphs. We note that when \( k = m \), our two-sided tester can be also guaranteed to test the conductance \( \phi(m) \) of \( G \) with the same running time and approximation performance (see the analysis in Section 4). In [LPP11], a product called non-uniform zig-zag product was proposed to transfer an arbitrary graph into a bounded degree graph. However, the analysis there is more involved and seems not generalize to the \( k \)-expansion profile for any \( k \leq m \) as considered here. Our analysis here is both simple and applicable to the broader case.

The techniques of random walks have also been used to test bipartiteness under different models [GR99, KKR04, CMOS11]. In particular, Kaufman et al. extend the bipartiteness tester in bounded degree graphs to general graphs [KKR04]. Ben-Eliezer et al. studied the strength of
different query types in the context of property testing in general graphs [BEKKR08]. The analysis for the expansion of the replacement product (and the zig-zag product) of two regular graphs are introduced in [RVW02, RV05, RTV06].

1.4 Organization of the paper

The rest of the paper is organized as follows. In Section 2 we give some basic definitions and introduce some tools for our analysis. Then we present a two-sided tester for $d$-regular graphs in Section 3. In Section 4, we introduce the non-uniform replacement product, describe the two-sided error tester for general graphs and prove Theorem 1. The one-sided error tester and its analysis are given in Section 5.

2 Preliminaries

Let $G = (V, E)$ be an undirected, unweighted graph with $|V| = n$ and $|E| = m$. Let $\deg_G(v)$ denote the degree of a vertex $v$. When it is clear from context, we will omit the subscript $G$. As mentioned in the introduction, we consider two different models for general graphs, the rotation map model and the adjacency list model. In the rotation map model, the graph is represented by its rotation map that for each vertex $u$ and an index $i \leq \deg(u)$, in the $(u,i)$th location of the representation the pair $(v,j)$ is stored such that $v$ is the $i$th neighbor of $u$ and $u$ is the $j$th neighbor of $v$. We are given an oracle access $O_G$ to the rotation map of $G$ and allowed to perform degree queries and neighbor queries to $O_G$. In the adjacency list model, the graph is represented by its adjacency list, which is also accessible through an oracle access $O_G$. Furthermore, degree and neighbor & index queries are allowed to perform to $O_G$. For a graph with maximum degree bounded by $d$, we assume that $d$ is a constant independent of $n$.

For a vertex subset $S \subseteq V$, let $e(S,V\setminus S)$ be the number of edges leaving $S$. Let $\vol(S) := \sum_{v \in S} \deg_G(v)$ and $\phi(S) := e(S,S)/\vol(S)$ be the volume and the conductance of $S$, respectively. Note that $\vol(G) := \vol(V) = 2|E|$. Define the $k$-expansion profile of $G$ as $\phi(k) := \min_{S, \vol(S) \leq k} \phi(S)$ in particular, $\phi(m)$ is often referred to the conductance (or expansion) of $G$ and we let $\phi(G) := \phi(m)$. A graph is called a $\phi$-expander if $\phi(G) \geq \phi$.

**Definition 1.** A graph $G$ is a $(k,\phi)$-expander if $\phi(k) \geq \phi$. Equivalently, $G$ is a $(k,\phi)$-expander if for every $S \subseteq V$ with volume $\vol(S) \leq k$ has conductance $\phi(S) \geq \phi$.

We have the following definition of graphs that are $\varepsilon$-far from $(k,\phi)$-expanders.

**Definition 2.** A graph $G$ is $\varepsilon$-far from a $(k,\phi)$-expander if one has to modify at least $\varepsilon m$ edges of $G$ to obtain a $(k,\phi)$-expander.

The main tool in our algorithm and analysis is the lazy random walk on $G$. In a lazy random walk, if we are currently at vertex $v$, then in the next step, we choose a random neighbor $u$ with probability $1/2\deg_G(v)$ and move to $u$. With the remaining probability $1/2$, we stay at $v$. Let $A, D$ denote the adjacency matrix and diagonal degree matrix of $G$, respectively.

We will use bold letters to denote row vectors. For any vector $p \in \mathbb{R}^V$, let $p(S) := \sum_{v \in S} p(v)$ and let $\|p\|_1 = \sum_{v \in V} |p(v)|, \|p\|_2 = \sqrt{\sum_{v \in V} p(v)^2}$ denote the $l_1, l_2$-norm of $p$, respectively. Let $\text{supp}(p)$ be the support of $p$. Let $1_S$ be the vector such that $1_S(v) = 1$ if $v \in S$ and $1_S(v) = 0$ otherwise. Let $1_v := 1_{\{v\}}$.

For a given graph $G$, let $A$ denote its adjacency matrix and let $D$ denote the diagonal matrix such that $D_{u,u} = \deg(u)$ for any $u$. Let $I$ denote the identity matrix. Then $W := (I + D^{-1}A)/2$
Lemma 1 (Cheeger’s inequality, [AM85, Alo86, SJ89]).

is probability transition matrix of the lazy random walk of $G$. Note that if $p_0$ is probability distribution on $V$, then $p_0 W^t$ denotes distribution of the endpoint of a length $t$ lazy random walks with initial distribution $p_0$. In particular, we let $p_v^t = 1_v W^t$ be the probability distribution of the endpoint of a walk of length $t$ starting from vertex $v$. Furthermore, we let $\|p_v^t\|_2^2$ denote the collision probability of such a walk.

For any lazy random walk matrix $W = \frac{I + D^{-1/2}A}{2}$, since $W = D^{1/2}(I + D^{-1/2}AD^{-1/2})D^{-1/2}$, we know that the eigenvalues of $W$ are the same as the eigenvalues of the symmetric matrix $I + D^{-1/2}AD^{-1/2}$. Therefore, all the eigenvalues are real. We let $\eta_i(W)$ denote the eigenvalues of $W$. It is well known that $0 \leq \eta_i(W) \leq 1$ for any $i \leq n$ (see eg. [OT12]).

2.1 Spectral characterization of expansion profile

For a graph $G$, let $L := I - D^{-1/2}AD^{-1/2}$ be the normalized Laplacian matrix of $G$. Let $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n \leq 2$ be eigenvalues of $L$. It is straightforward to verify that $\eta_i = 1 - \frac{\lambda_i}{2}$ for any $1 \leq i \leq n$, where $\eta_i$ is the $i$th largest eigenvalue of the lazy random walk matrix $W$ of $G$. We have the following lemma relating the expansion profile and the eigenvalues of $L$.

Lemma 1 (Cheeger’s inequality, [AM85, Alo86, SJ89]). For every graph $G$, we have $\frac{\lambda_2}{2} \leq \phi(G) \leq \sqrt{2\lambda_2}$.

Lemma 2 ([LOT12, LRTV12]). For every graph $G$, $h \in \mathbb{N}$ and any $\alpha > 0$, we have $\phi(\frac{(1+\alpha)2m}{h}) \leq O(\frac{1}{\sqrt{\lambda_h}} \sqrt{\lambda_h \log h})$.

Lemma 3 ([Ste10, OT12, OW12]). For every graph $G$, $h \in \mathbb{N}$ and any $\alpha > 0$, we have $\phi(\frac{2m}{h^{1+\alpha}}) \leq O(\sqrt{\lambda_h / \alpha} \log h n)$.

We remark that in some of references (eg. [LOT12]), the $k$-expansion profile is defined to the minimum conductance over all possible subsets of size at most $k$, rather than the volume measurement as defined here. However, their proofs imply that Lemma 2 and 3 also hold for our case.

2.2 A local algorithm for finding small non-expanding set

To give a one-sided error tester in general graphs, we need the following local algorithm for finding small sparse sets. Here, the local algorithm takes as input a vertex $v$ and only explores a small set of the vertices and edges that are “close” to $v$, if the volume $k$ of the target set is small. It only needs to perform degree queries and neighbor queries to the oracle of the input graph.

LocalSS$(G,v,T,\delta)$

1. Let $q_0 = 1_v$. For each time $0 \leq t \leq T$:
   (a) Define $\tilde{p}_t$ such that $\tilde{p}_t(u) = q_t(u)$ if $q_t(u) \geq \delta \deg(v)$ and $\tilde{p}_t(u) = 0$ if $q_t(u) < \delta \deg(v)$. Compute $q_{t+1} := \tilde{p}_t W$.
   (b) Let $s_t = |\text{supp}(\tilde{p}_t)|$. Order the vertices in $\text{supp}(\tilde{p}_t)$ so that $\frac{\tilde{p}_t(v_1)}{\deg(v_1)} \geq \frac{\tilde{p}_t(v_2)}{\deg(v_2)} \geq \cdots \geq \frac{\tilde{p}_t(v_{s_t})}{\deg(v_{s_t})}$.
      (c) For each $1 \leq i \leq s_t$, let $S_{i,t}$ be the first $i$ vertices in this ordering.
2. Output the subgraph $X$ with the smallest conductance among all the sets $\{S_{i,t}\}_{0 \leq t \leq T, 1 \leq i \leq s_t}$. 


The performance of the above algorithm is guaranteed in the following lemma, which follows by combining Proposition 8 in [OT12] and Theorem 2 in [KL12]. (More specifically, the first part of the lemma is Proposition 8 in [OT12] and the “Furthermore” part of the lemma follows from the proof of Theorem 2 in [KL12]. See also the paragraph “Independent Work” in [KL12].)

**Lemma 4.** Let \( G = (V, E) \) and \( t \geq 1 \). If \( S \subseteq V \) satisfies that \( \phi(S) \leq \psi \), then there exists a subset \( \hat{S} \subseteq S \) such that \( \text{vol} (\hat{S}) \geq \text{vol} (S) / 2 \), and for any \( v \in \hat{S} \), we have \( p_t^v (S) \geq c_1 (1 - \frac{3 \psi}{2})^t \) for some constant \( c_1 > 0 \). Furthermore, if \( \text{vol} (S) \leq k \), then the algorithm \( \text{LocalSS} \), with parameters \( G, v, T = O(\frac{\zeta \log k}{\psi}), \delta = O(\frac{k^{-1} - \zeta}{T}) \) for any \( \zeta > 0 \), will find a set \( X \) such that \( \text{vol} (X) \leq O(k^{1+\zeta}) \) and \( \phi(X) \leq O(\sqrt{\psi/\zeta}) \). The algorithm can be implemented in time \( \tilde{O}(k^{1+2\zeta} \psi^{-2}) \).

### 2.3 Properties of lazy random walks

Let \( t, r \geq 1 \). Consider \( r \) independent samples from the distribution \( p_t^v \). Note that such samples can be obtained by first independently performing \( r \) lazy random walks of length \( t \) from \( v \) and then taking the endpoints. For any vertex \( v \in V \), let \( Z_v \) denote the number of pairwise collisions among these samples. We have the following lemma that follows the first paragraph of the proof of Lemma 4.2 in [CS10] by setting \( \varepsilon = 1/2 \) there.

**Lemma 5.** If \( r \geq 16 \sqrt{n} \), then with probability at least \( 1 - \frac{16 \sqrt{n}}{r}, \frac{1}{2} \| p_t^v \|_2^2 \leq Z_v \leq \frac{3}{2} \| p_t^v \|_2^2 \).

### 3 A two-sided error tester for \( d \)-regular graphs

In this section, we propose an algorithm \( \text{SSETester2-Reg} \) to test small set expander for any \( d \)-regular graph \( G \) that may contain self-loops, where we assume that \( d \) is a constant and each self-loop contributes to the degree by 1. (Such self-loops are called half-loops in [Fri93, Fri08].) The algorithm is described as follows.

\[
\text{SSETester2-Reg}(G, s, r, \ell, \sigma)
\]

1. Repeat \( s \) times:
   
   (a) Select a vertex \( v \) uniformly at random from \( V \).
   
   (b) Perform \( r \) independent lazy random walks of length \( \ell \) starting from \( v \).
   
   (c) Let \( Z_v \) be the number of pairwise collisions among the endpoints of these \( r \) random walks.
   
   (d) If \( Z_v > \sigma \) then abort and output \text{reject}.

2. Output \text{accept}.

The above algorithm is almost the same as the testers for expansion in bounded degree graphs [GR00, CS10, KS11, NS10]. The main differences between our small set expansion tester above and the previous expansion testers is the choice of parameters.

The above algorithm can be easily applied for any \( d \)-bounded degree graph \( G \) in the adjacency list model, on which we perform the random walk as follows: at each vertex \( v \), with probability \( \deg_G(v)/2d \), we jump to a randomly chosen neighbor of \( v \), and with the remaining probability \( 1 - \frac{\deg_G(v)}{2d} \), we stay at \( v \). This random walk is equivalent to the lazy random walk on the virtually constructed \( d \)-regular graph \( G_{\text{reg}} \) that is obtained by adding an appropriate number of self-loops on
each vertex in $G$. Also note that to perform such a random walk, we only need to perform neighbor queries to the oracle of $G$. Such an idea has been used to give expansion testers for bounded degree graphs.

By our analysis below, we can show that the above algorithm is a property tester for small set expansion in bounded degree graphs (See Section 4, where we prove a stronger result). That is, with probability at least $2/3$, the algorithm accepts any $(k,\phi)$-expander graph $G$ with degree bounded $d$, and rejects any graph that is $\varepsilon$-far from $(k^*,\phi^*)$-expander with degree bounded by $d$, for some appropriately chosen $k^*,\phi^*$. We omit the formal statement here.

4 A two-sided error tester for general graphs

In this section, we extends the previous tester in bounded degree model to general graph model. To do this, we introduce a new graph product called non-uniform replacement product to transfer any graph $G$ into a $d$-regular graph.

4.1 Non-uniform replacement product

Let $G = (V,E)$ and let $\mathcal{H} = \{H_u\}_{u \in V}$ be a set of $|V|$ graphs. $\mathcal{H}$ is called a proper $d$-regular graph family of $G$ if for each $u \in V$, $H_u$ is a $d$-regular graph (with possible parallel edges or self-loops) with $\deg_G(u)$ vertices. For any graph $G$ and its proper $d$-regular graph family $\mathcal{H}$, the non-uniform replacement product of $G$ and $\mathcal{H}$, denoted by $G \circledast \mathcal{H}$, is defined as follows.

- For each vertex $u$ with degree $D$, the graph $G \circledast \mathcal{H}$ contains a copy of a $H_u$.
- For any edge $(u,v) \in E(G)$ such that $v$ is the $i$th neighbor of $u$ and $u$ is the $j$th neighbor of $v$, we place $d$ parallel edges between the $i$th vertex in $H_u$ and the $j$th vertex in $H_v$.

Note that $G \circledast \mathcal{H}$ is a $2d$-regular graph with $2|E|$ vertices. We will use $(u,i)$ to index the vertices in $G \circledast \mathcal{H}$.

4.2 Description of the tester

Let $\mathcal{H}$ be a proper $d$-regular graph family for $G$ such that for each $v$, $\eta_2(W_{H_v}) \leq 1 - \delta$ for some constant $\delta$. It is well known that such a family can be explicitly constructed (see e.g., [RVW02, ILW06]). Our tester for general graph simply returns the result of the tester $\text{SSETester2-Reg}$ on input graph $G \circledast \mathcal{H}$.

\begin{verbatim}
\text{SSETester2-General}(G,s,r,\ell,\sigma)
\begin{itemize}
  \item Return the result of \text{SSETester2-Reg}(G \circledast \mathcal{H},s,r,\ell,\sigma)
\end{itemize}
\end{verbatim}

We note that the product graph $G' = G \circledast \mathcal{H}$ is only virtually constructed when we are applying $\text{SSETester2-Reg}$ with $G'$. By our construction, the lazy random walk on $G'$ can be performed by performing as follows: if we are currently at a vertex $(u,i)$, then with probability $1/2$, we stay at $(u,i)$; with probability $1/4$, we jump to a randomly chosen neighbor $(u,j)$ in $H_u$; with the remaining probability $1/4$, we jump to vertex $(v,j)$ such that $v$ is the $i$th neighbor of $u$ and $u$ is the $j$th neighbor of $v$ in $G$. Note that only in the last case, we need to perform neighbor index queries to the oracle of $G$. In addition, to sample uniformly at random a vertex from $G'$ as required in $\text{SSETester2-Reg}$, we assume that we know the number of edges $m$ of $G$. Then we can first sample uniformly at random an edge from $G$ and then choose a random endpoint of this edge.
4.3 Analysis of SSETester2-General

In this section, we state two lemmas that will be used in our analysis of SSETester2-General. The proofs are deferred in Section 1.4.

The first lemma formally characterize the intuition that if all the graphs in $\mathcal{H}$ are expanders, then the (spectral) expansion profile of $G'$ will not differ too much from the expansion profile of $G$.

Lemma 6. Let $G = (V,E)$ and $d \geq 3$. Let $\mathcal{H}$ be a proper $d$-regular graph family of $G$, and let $G' = G \oplus \mathcal{H}$. We have that

- If $S \subset V(G)$ is subset with $\phi(S) \leq \phi$ then the set $S := \{(u,i) \in V(G')|u \in S, 1 \leq i \leq \deg_G(u)\}$ satisfies that $|S'| = |S|$ and $\phi_G(S') \leq \phi/2$.
- If for any set $S \subset V(G)$ with $\text{vol}(S) \leq k$, $\phi(S) \geq \phi$ and for any $u$, $\eta_2(W_{H_u}) \leq 1 - \delta$ for some $\delta > 0$, then for any $\alpha > 0$, $\eta_{\ell}(W_{G'}) \leq 1 - O(\delta^2 \phi^2 (\log 2m/k)^{-1})$, and $\eta_{\ell}(W_{G'}) \leq 1 - O(\alpha \phi^2 \log(n)(2m/k))$. Furthermore, when $k = m$, we have $\eta_2(W_{G'}) \leq 1 - O(\delta^2 \phi^2)$.

Our next lemma shows that if a graph $G$ is far from $(k, \phi)$-expander, then there exist disjoint subsets such that each of them is of small size and small conductance, and the total volume of these sets are large. This lemma is also useful for us to design and analyze a one-sided tester for small set expansion (see Section 3).

Lemma 7. Let $c_2$ be some constant and let $\phi^* \leq \frac{1}{20c_2}$. If a graph $G$ is $\varepsilon$-far from $(k^*, \phi^*)$-expander, then there exists disjoint subsets $S_1, \ldots, S_q \subset V$ such that $\text{vol}(S_1 \cup \cdots \cup S_q) \geq \frac{c_2}{20} \text{vol}(S)$, and for each $i \leq q$, $\text{vol}(S_i) \leq 2k^*$, $\phi(S_i) < 11c_2 \phi^*$.

Now we use the above two lemmas to prove Theorem 1. We prove the completeness and soundness separately. We set $k^* = \Theta(k\varepsilon)$ and $\phi^* = \Theta(\frac{\phi^2}{(\log(n)(\log k))})$. We set $s = \Theta(1/\varepsilon)$, $r = \Omega(s\sqrt{n})$, $\ell = O(\frac{(\log n)(\log k)}{\phi^2})$ and $\sigma = \bigg(\frac{r}{\ell}\bigg)^{60} \frac{60}{\ell k}$ in the algorithm SSETester2-General.

Lemma 8. If $G$ is a $(k, \phi)$-expander, then the algorithm SSETester2-General accepts $G$ with probability at least $2/3$.

Proof. Recall that in our tester, we have chosen a proper $d$-regular graph family $\mathcal{H}$ such that for each $v$, $\eta_2(W_{H_v}) \leq 1 - \delta$ for some constant $\delta$. Let $\eta_i$ denote the $i$th largest eigenvalue of $W_{G'}$ for any $1 \leq i \leq 2m$. Since $G$ is a $(k,\phi)$-expander, then by Lemma 6, for the virtual graph $G' = G \oplus \mathcal{H}$ and its lazy random walk $W_{G'}$, we have for any $\alpha > 0$, $\eta_{2m/k} \leq 1 - O(\alpha \phi^2 \log(n)(2m/k))$. (Note that compared with Lemma 6, we omit $\delta^2$ in this inequality since we have assumed that $\delta$ is a constant.) In particular, if we let $\alpha = \frac{\log(2m/k)}{\phi^2}$, then $\eta_{2m/k} \leq 1 - O(\phi^2 \frac{1}{\log n})$.

Note that for $\ell \geq 1$, the trace of matrix $W_{G'}^{2t}$, denoted $\text{Tr}(W_{G'}^{2t})$, satisfies that $\text{Tr}(W_{G'}^{2t}) = \sum_{i=1}^{2m} \eta_i^{2t} \leq \sum_{i=1}^{2m/k} \eta_i^{2t} + 2m \cdot \eta_{2m/k}^{2t} \leq 4m/k + 2m(1 - O(\phi^2 \frac{1}{\log n}))^{2t}$. By setting $t = \ell$, we have that $\text{Tr}(W_{G'}^{2t}) \leq \frac{8m}{k}$.

On the other hand, $\text{Tr}(W_{G'}^{2t}) = \sum_{(v,j) \in V(G')} \|1_{(v,j)} W_{G'}^t \|_2^2$. Thus, the average value of $\|1_{(v,j)} W_{G'}^t \|_2$ over all $2m$ possible vertices $(v,j)$ is at most $\frac{4}{k}$. Furthermore, if we let $U := \{(v,j)|\|1_{(v,j)} W_{G'}^t \|_2 < \frac{40}{k}\}$, then by Markov’s inequality, $|U| \geq (1 - \varepsilon/100m)$. Therefore, the probability that all the sampled vertices are in $U$ is at least $1 - \varepsilon/10^s \geq 5/6$, where $s = \Theta(1/\varepsilon)$ is the sample size.

Now we assume that all the sampled vertices are in $U$. By Lemma 6 and the definition of $U$, we know that for each sampled vertex $(v,j)$, $Z_{(v,j)} \leq \frac{4}{k^2} \frac{40}{k} = \sigma$ holds with probability at least $1 - \frac{16/5}{k^2} \geq 1 - \frac{1}{10s}$, where the last inequality follows from our choice that $r = O(\sqrt{ns})$. Then with
probability at least $1 - \frac{1}{10^6} \cdot s \geq \frac{5}{6}$, for all sampled vertices $(v, j)$, $Z_{(v, j)} \leq \sigma$, and thus the tester will accept $G$.

Overall, the probability that the tester will accept $G$ is $\frac{5}{6} \cdot \frac{5}{6} \geq \frac{2}{3}$.

\begin{lemma}
If $G$ is $\varepsilon$-far from $(k^*, \phi^*)$-expander, then the algorithm $\text{SSETester2-General}$ rejects $G$ with probability at least $2/3$.
\end{lemma}

\begin{proof}
Let $S_1, \ldots, S_q$ be the sets with properties in Lemma 8. By the first item of Lemma 8, for each $i \leq q$, there exists a set $S_i' \subseteq V(G')$ satisfying that $|S_i'| = \text{vol}(S_i)$, and $\phi_{G'}(S_i') \leq 11c_2\phi^*/2$. Furthermore, by the fact that $S_i$ are disjoint and the definition of $S_i'$, these sets $S_i'$ are disjoint.

Now we apply Lemma 8 with $G = G'$, $S = S_i'$ and $t = \ell$, $\psi = 11c_2\phi^*/2$ to find $\hat{S}_i' \subseteq S_i'$ such that $|\hat{S}_i'| \geq \tilde{S}_i'$ and for each vertex $(v, j) \in \hat{S}_i'$, if we let $p = 1_{(v, j)}W_{G'}^\ell$, then $\|p\|^2 \geq \sum_{(u, i) \in \hat{S}_i'}p_{(u, i)}^2 \geq \frac{\sum_{(u, i) \in \hat{S}_i'}p_{(u, i)}}{|\hat{S}_i'|} \geq \frac{\ell}{\ell} \left(1 - \frac{33c_2\phi^*}{4}\right)^\ell \geq \frac{121}{27k}$, where we used the Cauchy-Schwarz inequality, the fact that $|\hat{S}_i'| \leq k^*$ and our choice of parameters.

Thus, if we have sampled some vertex $(v, j) \in \hat{S}_i'$ for some $i \leq q$, then the collision probability of the corresponding random walk will be at least $\frac{121}{27k}$. Then by Lemma 8, with probability at least $1 - \frac{16\sqrt{n}}{r} > 5/6$, $Z_{(v, j)} \geq \frac{1}{2} \left(\frac{n}{2}\right) \frac{121}{27k} > \sigma$, where the last inequality follows by our assumption on $\xi$, and then the tester will reject the graph $G$.

Now we show that with probability at least $5/6$, we will sample some vertex $(v, j) \in \hat{S}_i'$ for some $i \leq q$. Note that since $\text{vol}(S_1 \cup \cdots \cup S_q) \geq \frac{2m}{\alpha}$, then $|\tilde{S}_1 \cup \cdots \cup \tilde{S}_q| \geq \frac{1}{2}|S_1 \cup \cdots \cup S_q| \geq \frac{2m}{6\alpha}$. Since we sampled $\Theta(1/\varepsilon)$ vertices, and the total number of vertices in $G'$ is $2m$, we can guarantee that the probability that the algorithm will sample out a vertex from $\tilde{S}_1 \cup \cdots \cup \tilde{S}_q$ is at least $5/6$.

Therefore, the overall probability that the tester will reject $G$ is at least $\frac{5}{6} \cdot \frac{5}{6} > \frac{2}{3}$.

Finally, the query complexity and the running time of our algorithm is $O(rs\ell) = \frac{\sqrt{n(\log n)(\ln k)}}{\epsilon^2 \phi^4}$. This completes the proof of Theorem 1.

\begin{remark}
1) In the proof above, we only used the inequality that $\eta(2m/k)^{1+\alpha}(W_{G'}) \leq 1 - O(\alpha\delta^2\phi^2 \log m(2m/k))$ from Lemma 8. One can also use the other inequality that $\eta(1+\varepsilon/2m)(W_{G'})^3 \leq 1 - O(\delta^2\phi^6 \log m(2m/k) - 1)$ to prove the performance of our tester by choosing different parameters. In addition, the “Furthermore” part of Lemma 8 can be used to give stronger upper bound on the collision probability, which combined with the lower bound of collision probability given in [KS11] can also be used to design conductance tester in rotation map model with the same running time and approximation guarantee as in [LPP11]. We omit these proofs here.

2) We note that Lemma 8 also applies to the bounded degree graphs in the adjacency list model, where a $d$-degree bounded graph $G$ is $\varepsilon$-far from any small set expander if one needs to modify at least $\varepsilon dn$ edges to get a small set expander with maximum degree $d$. (See the proof of Lemma 7 in the next section, where the modification algorithm for a graph preserves the degree for every vertex). Therefore, our analysis above for Theorem 1 implies that the algorithm $\text{SSETester2-Reg}$ in Section 3 with appropriately chosen parameters can be directly used to test the small set expansion of a bounded degree graph $G$ in the adjacency list model, as we only need to perform degree and neighbor queries to the oracle of $G$.

3) One may try to analyse combinatorially the expansion of small set in $G' = G\overline{H}$, and indeed, following the same analysis as in the proof of Theorem 1.3 in [ASS08], we can show that if $\phi_k(G) \geq \phi$, then $\phi_k(G') \geq \Omega(\phi^2)$. However, such a result only allows us to give a weaker tester that distinguishes $(k, \phi)$-expander graphs from any graph that is $\varepsilon$-far from $(k^*, \phi^*)$-expander graph with $\phi^* = O(\phi^4)$, rather than the bound $\phi^* = O(\phi^2)$ as given here.

\end{remark}
4.4 Proofs of Lemma 6 and 7

In this section, we prove Lemma 6 and 7. We first give a useful lemma for the proof of Lemma 6. Let $J_n$ denote the $n \times n$ matrix with all elements equal to $\frac{1}{n}$. Recall that $\eta_i(W)$ is the $i$th largest eigenvalue of matrix $W$.

**Lemma 10.** Let $H$ be a d-regular graph on $n$ vertices and let $W$ be its lazy random walk matrix. If $\eta_2(W) \leq 1 - \delta$ for some $0 \leq \delta < 1$, then $W = \delta J_n + (1 - \delta) B$, where $\eta_1(B) \leq 1$.

**Proof.** Define $B := \frac{W \cdot J_n}{\delta}$. Since $H$ is $d$-regular, $W$ is symmetric and thus we can find orthonormal eigenvectors $v_1, \ldots, v_n$ of $W$ corresponding eigenvalues $\eta_1(W), \ldots, \eta_n(W)$ such that $\{v_i\}_{i=1}^n$ form an orthonormal basis of $\mathbb{R}^V$. Then by the spectral decomposition theorem, $W = \sum_{i=1}^n \eta_i(W)v_i^T v_i$. Noting that $\eta_1(W) = 1$ and $v_1 = (\frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}})$, we have that $B = \frac{(1-\delta)v_1^Tv_1 + \sum_{i=2}^n \eta_i(W)v_i^Tv_i}{\delta}$. Now for any nonzero vector $x$, we can write it as $x = \sum_{i=1}^n a_i v_i$, and thus $\frac{x^T B x}{x^T x} = \frac{\alpha_1^2 + \sum_{i=2}^n \alpha_i^2 \eta_i(W)}{\sum_{i=1}^n \alpha_i^2} \leq \frac{\sum_{i=1}^n \alpha_i^2}{\sum_{i=1}^n \alpha_i^2} = 1$, where the last inequality follows from the fact that for any $i \geq 2$, $\eta_i \leq \eta_2$ and the precondition that $\eta_2(W) \leq 1 - \delta$. Therefore, $\eta_1(B) = \max_{x \in \mathbb{R}^V, x \neq 0} \frac{x^T B x}{x^T x} \leq 1$. \hfill $\square$

Let $R_G$ be the permutation matrix corresponding the rotation map of $G$. That is, $R_G$ is an $\text{vol}(G) \times \text{vol}(G)$ matrix such that for each row indexed $(u, i)$, only in the column indexed $(v, j)$ the entry is 1, and in any other column, the entry is 0, where $v$ is the $i$th neighbor of $u$ and $u$ is the $j$th neighbor of $i$. For each $u \in V$, let $W_{H_u}$ be the lazy random walk matrix of $H_u$ and let $W_G$ be the block diagonal matrix with each block $W_{H_u}$. Note that by our construction, the lazy random walk matrix $W_{G'}$ of $G'$ satisfies that $W_{G'} = \frac{1}{2} (I + R_G) + W_H$. Now we are ready to prove Lemma 6.

**Proof of Lemma 6.** The first part of the lemma is straightforward. By the definition of $S'$, $S'$ is the set consisting of all vertices in $H_u$ for any $u \in H$. Thus, $|S'| = \sum_{u \in S} \deg_{G_u}(u) = \text{vol}(S)$. Furthermore, since $\phi(S) \leq \phi$, then $\epsilon(S, V \setminus S) \leq \phi \text{vol}(S)$. By our construction of $G'$, the number of edges between $S'$ and $V(G') \setminus S'$ is $d \cdot e(S, V \setminus S) \leq d \cdot \phi \text{vol}(S) = \phi \text{vol}_{G'}(S')/2$, which gives that $\phi_{G'}(S') \leq \phi/2$.

Now we prove the second part of the lemma. For each $u \in V$, let $W_{H_u}$ denote the lazy random walk matrix of $H_u$. Then by Lemma 6 and the assumption that $\eta_2(W_{H_u}) \leq 1 - \delta$ for every $u$, we have $W_{H_u} = \delta J_{\deg_{G_u}(u)} + (1 - \delta) B_u$, where $\eta_1(B_u) \leq 1$. Let $B_H$ (resp., $J_H$) be the block diagonal matrix with each block $B_u$ (resp., $J_{\deg_{G_u}(u)}$). Therefore $W_H = \delta J_H + (1 - \delta) B_H$, and $W_{G'} = \frac{1}{2} (I + R_G) + W_H = \frac{1}{2} (I + \frac{R_G}{2} + \delta J_H + (1 - \delta) B_H)$. Furthermore, $W_{G'} = \frac{1}{2} (I + R_G) + \theta J_H + (1 - \delta) B_H$. Then we expand all terms to get that $W_{G'}^3 = (1 - \frac{\delta^2}{8} B + \frac{\delta^2}{8}) J_H (I + R_G) J_H$. By Weyl’s inequality [Tao12], we have that for any $j \leq 2m$, $\eta_j(W_{G'}^3) \leq (1 - \frac{\delta^2}{8}) \eta_j(B) + \frac{\delta^2}{8} \eta_j(P) \leq 1 - \frac{\delta^2}{8} + \frac{\delta^2}{8} \eta_j(P)$.

Now if we can show that for any $\alpha > 0$, $\eta_{\frac{\alpha k}{2m} + 2m}(P) \leq 1 - O(\alpha^6 \phi^2 \log(\frac{2m}{k})^{-1})$, and $\eta_{\frac{2m}{k} + 2m}(P) \leq 1 - O(\alpha^6 \phi^2 \log(\frac{2m}{k})^{-1})$, and $\eta_{\frac{2m}{k} + 2m}(W_{G'}) \leq 1 - O(\alpha^6 \phi^2 \log(\frac{2m}{k})^{-1})$, and $\eta_{\frac{2m}{k} + 2m}(W_{G'}) \leq 1 - O(\alpha^2 \phi^2 \log(\frac{2m}{k})^{-1})$. The lemma then follows by noting that $\eta_j(W_{G'}) = \eta_j(\text{vol}_{G'})$.

In the remaining of the proof, we bound the eigenvalues of $P$. We need the following two claims.

**Claim 3.** For any $(u, i), (v, j) \in V(G')$, $P_{(u, i), (v, j)} = \frac{1}{\deg_{G'}(v)} \left( (D_{G'}^{-1} A_G)(u, v) + I(u, v) \right) = W_{G'}(u, v) / \deg_{G'}(v)$.

**Proof.** Since $P = J_H R_G J_H$, then $P$ can be seen as the random walk matrix on $V(G')$ that does the following from $(u, i)$: first chooses a random number $k_1$ from set $\{\deg_{G_u}(u)\} := \{1, \ldots, \deg_{G_u}(u)\}$, and then 1) with half probability it stays at $(u, k_1)$, and then chooses a random number $k_2$ from
that for any \( \alpha > 0 \), there is a one-to-one correspondence between the nonzero eigenvalues of \( P \) and the nonzero eigenvalues of \( D_G^{-1}P \).  

Proof. On one hand, let \( x \in \mathbb{R}^{V(G')} \) be an eigenvector of \( P \) with eigenvalue \( \eta \neq 0 \). Then for any \( (v, j) \), 
\[
\eta x(v, j) = \sum_{(u, i)} x(u, i) M_G(u, v) \sum_{(u, i)} x(u, i) W_G(u, v) \sum_{(v, j)} x(v, j),
\]
which is independent of \( j \). Since \( \eta \neq 0 \), this means that for any \( v \) and \( 1 \leq j_1, j_2 \leq \deg_G(v) \), \( x(v, j_1) = x(v, j_2) \). Furthermore, if we let \( y_v = x(v, j) \), then by the above calculation, for any \( v \in V(G) \), 
\[
y_MG(v) = \sum u \deg_G(u) y_u W_G(u, v) = \sum u y_u M_G(u, v) = \eta y_v.
\]
which gives that \( y \) is the eigenvector of \( M_G \) with eigenvalue \( \eta \).

On the other hand, let \( y \in \mathbb{R}^{V(G)} \) be an eigenvector of \( M_G \) with eigenvalue \( \eta \). Then we define for each \( u \) and \( 1 \leq i \leq \deg(u) \), \( x_u(i) = y_u \). Then for any \( (v, j) \in V(G') \), \( xP(v, j) = \sum_{(u, i)} x(u, i) M_G(u, v) = \sum u y_u M_G(u, v) = \eta y_v = \eta x(v, j) \), which means that \( x \) is an eigenvector of \( P \) corresponding to eigenvalue \( \eta \).

Now note that the eigenvalues of \( W_G \) are the same as the eigenvalues of \( W_G \) since \( W_G = D_G^{-1}M_G D_G^{-1} \), and both the eigenvalues of \( P \) and \( M_G \) are non-negative. These two facts combined imply that for any \( i \leq n \), \( \eta_i(P) \leq \eta_i(W_G) = 1 - \frac{\lambda_i}{4} \), where \( \lambda_i \) is the \( i \)th smallest eigenvalue of the Laplacian matrix \( L \) of \( G \). Finally, by the fact that \( \phi_G(k) \geq \phi \) and Lemma 11, we get that for any \( \alpha > 0 \), \( \eta_{(1+\epsilon)2m}(P) \leq 1 - O(\alpha \phi^2 (\log 2m/k)^{-1}) \), and \( \eta(2m/k)^{1+\epsilon}(P) \leq 1 - O(\alpha \phi^2 \log_k 2m/k) \).

The “Furthermore” part of the lemma follows from the above analysis and the Cheeger’s inequality given in Lemma 11.

Now we prove Lemma 11. We will use the following lemma that is implied in the proof of Lemma 8, 9 and 10 in [LPP11]. The corresponding lemma for bounded degree graphs are given in [KS11] and [CS10].

Lemma 11 ([LPP11]). Let \( G = (V, E) \) and let \( c_2 \) be some constant. If there exists a set \( A \subseteq V \) such that \( \text{vol}(A) \leq \frac{\varepsilon \text{vol}(G)}{20} \), and the subgraph \( G[V \setminus A] \) is a \( \phi^* \)-expander, then there exists an algorithm that modifies at most \( \varepsilon m \) edges to get a \( c_2 \phi^* \)-expander \( G' = (V, E') \) such that for each \( v \in V \), \( \deg_{G'}(v) \leq \deg_G(v) \).

The following result is a direct corollary of the above lemma. That is, we can use the same proof and modification algorithm of Lemma 11 to show that if \( G[V \setminus A] \) is a \( (k^*, \phi^*) \)-expander for some set \( A \) with small volume, then \( G \) is not \( \varepsilon \)-far from \( (k^*, c_2 \phi^*) \)-expander. Actually, the proof in [LPP11] studies the expansion of all possible sets of volume at most \( \frac{\text{vol}(G)}{2} \), here we only need to consider sets of volume at most \( k^* \leq \frac{\text{vol}(G)}{2} \).

Corollary 1. If there is a set \( A \subseteq V \) with \( \text{vol}(A) \leq \frac{\varepsilon \text{vol}(G)}{10} \) such that \( G[V \setminus A] \) is a \( (k^*, \phi^*) \)-expander, then \( G \) is not \( \varepsilon \)-far from \( (k^*, c_2 \phi^*) \)-expander. Furthermore, if the maximum degree of \( G \) is bounded by some constant \( d \), then \( G \) is not \( \varepsilon \)-far from to a \( (k^*, c_2 \phi^*) \)-expander with maximum degree at most \( d \).

For a set \( S \subset V \) and \( T \subset S \), we use \( \text{vol}_S(T) \) and \( \phi_S(T) \) to denote the volume and conductance of \( T \) measured in the induced subgraph \( G[S] \). If \( S = V \), we drop the subscript of \( \text{vol}_S(T), \phi_S(T) \). We let \( \epsilon(S) \) denote the number of edges in \( S \).
Proof of Lemma 7. We perform the following algorithm on \( G \). Let \( A_0 \) be the empty set and let \( V_0 := V \). For each \( i \geq 1 \), if \( \text{vol}(\bigcup_{j<i} A_j) \leq \frac{2m}{10} \), then we apply Corollary 4 with \( A = \bigcup_{j<i} A_j \) to find a subset \( A_i \subseteq V_{i-1} \) such that \( \text{vol}_G[V_{i-1} \setminus (A_i)] \leq k^* \) and \( \phi_{G[V_{i-1} \setminus (A_i)]} < 2c_2 \phi^* \), then we remove \( A_i \) from \( V_{i-1} \) and let \( V_i := V_{i-1} \setminus A_i \). By Corollary 4 we can repeat this process until at some time \( s \), \( \text{vol}(A_1 \cup \cdots \cup A_s) \geq \frac{2m}{10} \). We let \( P = A_1 \cup \cdots \cup A_s \).

Note that \( \sum_{i=1}^s e(A_i, V \setminus A_i) \leq 2 \sum_{i=1}^s e(A_i, V_{i-1} \setminus A_i) \leq 2 \sum_{i=1}^s \text{vol}_{V_{i-1}}(V_i)c_2 \phi^* \leq 2c_2 \text{vol}(P)\phi^* \).

Now we call an index \( i \) bad, if \( \text{vol}_{V_{i-1}}(A_i) < (1 – 10c_2 \phi^*)\text{vol}(A_i) \), and good otherwise, for each \( 1 \leq i \leq s \). Note that for a bad index \( i \) and the corresponding set \( A_i \), we have \( e(A_i, V \setminus A_i) = \text{vol}(A_i) – 2e(A_i) \geq \text{vol}(A_i) – \text{vol}_{V_{i-1}}(A_i) > 10c_2 \phi^*\text{vol}(A_i) \), where the last inequality follows by our definition of bad indices. Therefore, \( \sum_{i: \text{bad}} \text{vol}(A_i) < \frac{1}{10c_2 \phi^*} \sum_{i: \text{bad}} e(A_i, V \setminus A_i) < \frac{1}{3} \text{vol}(P) \). This means that \( \sum_{i: \text{good}} \text{vol}(A_i) \geq (1 – \frac{1}{3}) \text{vol}(P) \geq \frac{2m}{15} \), and for each good \( i \), \( \text{vol}(A_i) \leq \frac{1 – 10c_2 \phi^*}{1 – 10c_2 \phi^*} \text{vol}_{V_{i-1}}(A_i) \leq 2k^* \) by our assumption that \( \phi^* \leq \frac{20c_2}{k^*} \).

Furthermore, by definition, \( \phi_{V_{i-1}}(A_i) = \frac{e(A_i, V_i \setminus A_i)}{\text{vol}_{V_{i-1}}(A_i)} = \frac{\text{vol}_{V_{i-1}}(A_i) – 2e(A_i)}{\text{vol}_{V_{i-1}}(A_i)} \leq 2c_2 \phi^* \), we have that \( 2e(A_i) \geq (1 – 2c_2 \phi^*)\text{vol}_{V_{i-1}}(A_i) \geq (1 – 2c_2 \phi^*)(1 – 10c_2 \phi^*)\text{vol}(A_i) \geq (1 – 11c_2 \phi^*)\text{vol}(A_i) \), which gives that \( \phi(A_i) = \frac{\text{vol}(A_i) – 2e(A_i)}{\text{vol}(A_i)} \leq 11c_2 \phi^* \). The lemma follows by specifying \( S_j \) to be sets \( A_i \) with good indices \( i \).

\( \square \)

5 A tester with one-sided error for adjacency list model

Now we present our property testing algorithm \( \text{SSETester}_1 \) with one-sided error for small set expansion. This tester applies to the adjacency list model.

\[ \text{SSETester}_1(G, s, T, \delta) \]

1. Repeat \( s \) times:

   (a) Pick a random vertex \( v \) with probability proportional to its degree.

   (b) If \( \text{LocalSS}(G, v, T, \delta) \) finds a set \( X \) with volume at most \( k \) and conductance at most \( \phi \), then abort and output \( \text{reject} \).

2. Output \( \text{accept} \).

Note that in the above, to pick a random vertex with probability proportional to its degree, we again assume that we know the number of edges \( m \) of \( G \). Then we can first sample uniformly at random an edge from \( G \) and then choose a random endpoint of the edge.

Proof of Theorem 3. In the algorithm \( \text{SSETester}_1 \), we set \( s = \Theta(1/\varepsilon) \), \( T = O(\frac{\log k}{\phi^*}) \), and \( \delta = O(\frac{k^{-1+\xi/2}}{T}) \). It is obvious that for any input graph \( G \) that is a \((k, \phi)\)-expander, \( \text{SSETester}_1 \) cannot output \( \text{reject} \). Thus, we only need to consider the case that \( G \) is \( \varepsilon \)-far from \((k^*, \phi^*)\)-expander, where \( k^* = O(k^{1-\xi}) \) and \( \phi^* = O(\xi^2 \phi^2) \) for any \( 0 < \xi < 1/2 \). In this case, there exists disjoint subsets \( S_1, \ldots, S_q \subseteq V \) with properties in Lemma 6. Now for each \( i \leq q \), by applying Lemma 6 with \( S = S_i, \psi = 11c_2 \phi^*, k = 2k^* \) and \( \zeta = \xi/2 \), we know that there exists \( \tilde{S}_i \subseteq S_i \) such that \( \text{vol}(\tilde{S}_i) \geq \frac{1}{2} \text{vol}(S_i), \) and that for each \( v \in \tilde{S}_i \), the algorithm \( \text{LocalSS} \) with parameters \( G, v, T = O(\frac{\xi \log k^*}{\phi^*}) = O(\frac{\log k}{\phi^*}) \delta = O(\frac{k^* - 1+\xi/2}{T}) = O(\frac{k^{1+\xi/2}}{T}) \), will find a set \( X \) such that \( \text{vol}(X) \leq O((k^*)^{1+\xi/2}) = O(k^{1-\xi/2}) < k \) and \( \phi(X) \leq O(\sqrt{\psi/\xi}) < \phi \) by our choice of \( k^* \) and \( \phi^* \). Finally, noting that \( \text{vol}(\tilde{S}_1 \cup \cdots \cup \tilde{S}_q) \geq \frac{1}{2} \text{vol}(S_1 \cup \cdots \cup S_q) \geq \frac{2m}{30} \), and our sample size is \( s = \Theta(1/\varepsilon) \), we can guarantee
that with probability at least $2/3$, the algorithm will find a small non-expanding set and thus reject the graph $G$. Finally, the running time of the algorithm is $\tilde{O}(\frac{k}{\varepsilon \varphi^4})$.

References

[Alo86] N. Alon. Eigenvalues and expanders. *Combinatorica*, 6(2):83–96, 1986.

[AFNS09] N. Alon, E. Fischer, I. Newman, and A. Shapira. A combinatorial characterization of the testable graph properties: it’s all about regularity. *SIAM Journal on Computing*, 39(1):143–167, 2009.

[AM85] N. Alon and V. Milman. $\lambda_1$, isoperimetric inequalities for graphs, and superconcentrators. *Journal of Combinatorial Theory, Series B*, 38(1):73–88, 1985.

[ASS08] N. Alon, O. Schwartz, and A. Shapira. An elementary construction of constant-degree expanders. *Comb. Probab. Comput.*, 17(3):319–327, May 2008.

[ACL06] R. Andersen, F. Chung, and K. Lang. Local graph partitioning using pagerank vectors. In *Symposium on Computer Science*, 2006.

[AP09] R. Andersen and Y. Peres. Finding sparse cuts locally using evolving sets. In *Symposium on Theory of Computing*, STOC ’09, 2009.

[ABS10] S. Arora, B. Barak, and D. Steurer. Subexponential algorithms for unique games and related problems. In *Foundations of Computer Science (FOCS)*, pages 563–572, 2010.

[BEKKR08] I. Ben-Eliezer, T. Kaufman, M. Krivelevich, and D. Ron. Comparing the strength of query types in property testing: the case of testing k-colorability. In *Proceedings of the nineteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 1213–1222, 2008.

[CMOS11] A. Czumaj, M. Monemizadeh, K. Onak, and C. Sohler. Planar graphs: Random walks and bipartiteness testing. In *Foundations of Computer Science (FOCS), 2011 IEEE 52nd Annual Symposium on*, pages 423–432. IEEE, 2011.

[CS10] A. Czumaj and C. Sohler. Testing expansion in bounded-degree graphs. *Combinatorics, Probability and Computing*, 19(5-6):693–709, 2010.

[Fri93] J. Friedman. Some geometric aspects of graphs and their eigenfunctions. *Duke Mathematical Journal*, 69(3):487–525, 1993.

[Fri08] J. Friedman. A proof of Alon’s second eigenvalue conjecture and related problems. *Memoirs of the American Mathematical Society (AMS)*, 195(910), 2008.

[Gol10] O. Goldreich. *Property testing: current research and surveys*, volume 6390. Springer-Verlag New York Inc, 2010.

[GR99] O. Goldreich and D. Ron. A sublinear bipartiteness tester for bounded degree graphs. *Combinatorica*, 19(3):335–373, 1999.

[GR00] O. Goldreich and D. Ron. On testing expansion in bounded-degree graphs. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 7, 2000.
[GR02] O. Goldreich and D. Ron. Property testing in bounded degree graphs. Algorithmica, 32(2):302–343, 2002.

[HLW06] S. Hoory, N. Linial, and A. Wigderson. Expander graphs and their applications. Bulletin of the American Mathematical Society, 43(4):439–561, 2006.

[KS11] S. Kale and C. Seshadhri. An expansion tester for bounded degree graphs. SIAM J. Comput., 40(3):709–720, 2011.

[KKR04] T. Kaufman, M. Krivelevich, and D. Ron. Tight bounds for testing bipartiteness in general graphs. SIAM Journal on computing, 33(6):1441–1483, 2004.

[KL12] T. C. Kwook and L. C. Lau. Finding small sparse cuts by random walk. In APPROX-RANDOM, pages 615–626, 2012.

[LOT12] J.R. Lee, S. Oveis Gharan, and L. Trevisan. Multi-way spectral partitioning and higher-order cheeger inequalities. ACM symposium on Theory of computing, 2012.

[LPP11] A. Li, Y. Pan, and P. Peng. Testing conductance in general graphs. In Electronic Colloquium on Computational Complexity (ECCC), volume 18, page 101, 2011.

[LRTV12] A. Louis, P. Raghavendra, P. Tetali, and S. Vempala. Many sparse cuts via higher eigenvalues. ACM symposium on Theory of computing, 2012.

[LK99] L. Lovász and R. Kannan. Faster mixing via average conductance. In ACM symposium on Theory of computing, pages 282–287, 1999.

[NS10] A. Nachmias and A. Shapira. Testing the expansion of a graph. Information and Computation, 208(4):309–314, 2010.

[OW12] R. O’Donnell and D. Witmer. Improved small-set expansion from higher eigenvalues. Arxiv preprint arXiv:1204.4688, 2012.

[OT12] S. Oveis Gharan and L. Trevisan. Approximating the expansion profile and almost optimal local graph clustering. In Foundations of Computer Science, 2012.

[PR02] M. Parnas and D. Ron. Testing the diameter of graphs. Random Structures & Algorithms, 20(2):165–183, 2002.

[RS10] P. Raghavendra and D. Steurer. Graph expansion and the unique games conjecture. In ACM symposium on Theory of computing, pages 755–764, 2010.

[RTV06] O. Reingold, L. Trevisan, and S. Vadhan. Pseudorandom walks on regular digraphs and the rl vs. l problem. In Proceedings of the thirty-eighth annual ACM symposium on Theory of computing, pages 457–466. ACM, 2006.

[RVW02] O. Reingold, S. Vadhan, and A. Wigderson. Entropy waves, the zig-zag graph product, and new constant-degree expanders. Annals of Mathematics, 155:157–187, 2002.

[Ron10] D. Ron. Algorithmic and analysis techniques in property testing. Foundations and Trends in Theoretical Computer Science, 5(2):73–205, 2010.

[RV05] E. Rozenman and S. Vadhan. Derandomized squaring of graphs. In Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques, pages 436–447. Springer, 2005.
[SJ89] A. Sinclair and M. Jerrum. Approximate counting, uniform generation and rapidly mixing Markov chains. *Information and Computation*, 82(1):93–133, 1989.

[ST08] D.A. Spielman and S.H. Teng. A local clustering algorithm for massive graphs and its application to nearly-linear time graph partitioning. *arXiv:0809.3232*, 2008.

[Ste10] D. Steurer. On the complexity of unique games and graph expansion. *PhD diss.*, *Princeton University*, 2010.

[Tao12] T. Tao. *Topics in random matrix theory*, volume 132. American Mathematical Soc., 2012.