An expansion tester for bounded degree graphs

Satyen Kale*
Yahoo! Research
Santa Clara, CA
skale@yahoo-inc.com

C. Seshadhri†
Sandia National Labs‡
Livermore, CA
scomand@sandia.gov

Abstract

We consider the problem of testing graph expansion (either vertex or edge) in the bounded degree model (Goldreich & Ron, ECCC 2000). We give a property tester that takes as input a graph with degree bound d, an expansion bound α, and a parameter $\varepsilon > 0$. The tester accepts the graph with high probability if its expansion is more than α, and rejects it with high probability if it is ε-far from any graph with expansion α' with degree bound d, where $\alpha' < \alpha$ is a function of α. For edge expansion, we obtain $\alpha' = \Omega(\frac{\alpha^2}{d})$, and for vertex expansion, we obtain $\alpha' = \Omega(\frac{\alpha^2}{d^2})$. In either case, the algorithm runs in time $\widetilde{O}(\frac{n^{(1+\mu)/2}d^2}{\varepsilon\alpha^2})$ for any fixed $\mu > 0$.

1 Introduction

With the presence of large data sets, reading the whole input may be a luxury. It becomes important to design algorithms that run in time that is *sublinear* in (or even independent of) the size of the input. Sublinear algorithms are often achieved by dealing with a relaxed version of the decision problem. In *property testing* [GGR98, RS96], we wish to accept inputs that satisfy some predetermined property, and reject those that are sufficiently "far" from having that property. There is usually a well-defined notion of the "distance" of an input to a given property. In recent times, many advances have been made on algorithms for testing a variety of combinatorial, algebraic, and geometric properties (see surveys [Fis01, Gol98, Ron01, Gol10, Ron10]). For property testing in graphs [GGR98], there has been a large amount of work for testing in *dense graphs*. Here, it is assumed that the graph is given as an adjacency matrix. There are very general results about classes of properties that can tested in time independent of the size of the graph ([GGR98, AFNS09, AS08]).

The problem of property testing for bounded degree graphs was first dealt with by Goldreich and Ron [GR02]. The input graph G is assumed to have a constant degree bound d. The graph G is represented by *adjacency lists* - for every vertex v, there is list of vertices (of size at most d) adjacent to v. This allows testing algorithms to perform walks in the graph G. Given a property $\mathcal{P}$ and positive $\varepsilon < 1$, the graph G is ε-far from having $\mathcal{P}$ if G has to be modified at more than εnd edges for it to have property $\mathcal{P}$. Note that this includes both edge additions and deletions, and we want to keep the degree bound constant (usually, we require that the degree bound d is preserved). In this model, the first result about classes of testable properties was obtained by Goldreich and Ron [GR02]. They showed that a large class of connectivity properties could be tested in time

*This work was done when the author was at Princeton University and Microsoft Research, Redmond.

†This work was done when the author was at Princeton University.

‡Employee of Sandia National Laboratories. Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

essentially independent in the size of G. Czumaj, Shapira, and Sohler [CSS09] showed testability results for classes of graphs that do not contain expanders. Recently, Benjamini, Schramm, and Shapira [BSS10] showed a general result for testing minor-freeness properties.

For some properties, it was observed that testers could be constructed by performing random walks. The first such property tester was given by Goldreich and Ron [GR99], for the property of bipartiteness. Their tester made $\widetilde{O}(\sqrt{n})$ queries to the graph. In later work, Goldreich and Ron [GR00] posed the question of testing *expansion*. Given positive parameters $\lambda, \varepsilon < 1$, they constructed a $O(n^{0.5+\mu})$-time algorithm (for any constant $\mu > 0$). This algorithm provably accepted every graph G whose second largest eigenvalue $\lambda(G)$ is less than λ, and it was *conjectured* to reject every graph that is ε-far from having second eigenvalue less than λ' (here λ' could be much larger than λ, but $\lambda' \leq \lambda^{\Omega(1)}$). The running time is essentially tight (in n), since it has been proven that a property tester for expansion requires $\Omega(\sqrt{n})$ queries [GR02].

In the adjacency list model, the basic operation that we possess is that of walking in the graph. By performing a constant number of constant length walks, various connectivity properties were shown to be testable [GR02]. Yet there are properties for which such an approach would not work, because of non-constant lower bounds [GR02]. Random walks seem like a very natural operation to perform, and this relates to the expansion of the graph. One of the major parts of the analysis of the algorithm of [GR99] for bipartiteness deals with the expansion properties of the graph. This immediately raises the question of whether random walks can be used to test expansion. The problem of designing a property tester for expansion remained open for more than 6 years, until recently, when Czumaj and Sohler [CS07] provided a tester for *vertex expansion*. We describe this problem more formally below.

We are given an input graph $G = (V, E)$ on n vertices with degree bound d. Assume that d is a sufficiently large constant. Given a cut $(S, \bar{S})$ (where $\bar{S} = V \setminus S$) in the graph, let $E(S, \bar{S})$ be the number of edges crossing the cut. The edge expansion of the cut is $\frac{E(S,\bar{S})}{\min\{|S|,|\bar{S}|\}}$. The edge expansion of the graph is the minimum edge expansion of any cut in the graph. The vertex expansion of the cut is $\frac{|\partial S|}{|S|}$, where ∂S is the set of vertices in $\bar{S}$ that are adjacent to vertices in S. The vertex expansion of the graph is the minimum vertex expansion of any cut in the graph.

Hereafter, when we use the term "graph", we are only concerned with graphs having degree bound d. We are interested in designing a property tester for expansion (either edge or vertex). The graph is represented by an adjacency list, so we have constant time access to the neighbors of any vertex. Given parameters, $\alpha > 0$ and $\varepsilon > 0$, we want to accept to all graphs with expansion greater than α, and reject all graphs that are ε-far from having expansion less than $\alpha' < \alpha$ (where α' is some function of α). This means that G has to be changed at least εnd edges (either removing or adding, keeping the degree bound d) to make the expansion at least α'.

1.1 Our results

The problem of testing vertex expansion was first discussed by Czumaj and Sohler [CS07]. Their algorithm was based on that of Goldreich and Ron [GR00], and they used combinatorial techniques to prove the correctness of their algorithm. Their tester runs in time $O(\alpha^{-2}\varepsilon^{-3}d^2\sqrt{n}\ln(n/\varepsilon))$ and has parameter $\alpha' = \Theta(\frac{\alpha^2}{d^2 \log n})$.

Independently, using the same algorithm but via algebraic proof techniques, we gave an analysis [KS07] that allowed us to remove the dependence of n in α', and we obtain $\alpha' = \Theta(\frac{\alpha^2}{d^2})$ for vertex expansion and $\alpha' = \Theta(\frac{\alpha^2}{d})$ for edge expansion. This improvement in α' is significant since in most algorithmic applications of expanders, we need the graph to have *constant* expansion, and our property tester allows us to distinguish graphs that have constant expansion from those that are far from having (a smaller) constant expansion.

However, in the initial unpublished version of this paper, which appeared as a technical report on ECCC [KS07], we prove that the tester rejects graphs that are ε-far from any graph of expansion α' with degree bound $2d$, rather than degree bound d. In this version of the paper, in addition to our previous results, we also show how a small modification to our earlier techniques improves the degree bound to d. We recently found out that independently, the degree bound improvement was also obtained by Nachmias and Schapira [NS07] using a combination of our techniques and those of Czumaj and Sohler.

To describe our results, we set up some preliminaries. Consider the following slight modification of the standard random walk on the graph. Starting from any vertex, the probability of choosing any outgoing edge is $1/2d$, and with the remaining probability, the random walk stays at the current vertex. Thus, for a vertex of degree $d' \leq d$, the probability of staying at this vertex (a self-loop) is $1 - d'/2d \geq 1/2$. This walk is symmetric and reversible; therefore, its stationary distribution is uniform over the entire graph. Consider a cut $(S, \bar{S})$ with $|S| \leq n/2$. The *conductance* of this cut is the probability that, starting from the stationary distribution, the random walk leaves the set S in one step, conditioned on the event that the starting state is in S. For our chain, the conductance thus becomes $E(S, \bar{S})/2d|S|$, which is just the expansion of the cut divided by $2d$. The conductance of the graph, Φ_G, is the minimum conductance of any cut in the graph.

Our goal is to design a property tester for graph conductance. The tester is given two parameters Φ and ε. The tester must (with high probability[1]) accept if $\Phi_G > \Phi$ and reject if G is ε-far from having $\Phi_G > c\Phi^2$ (for some absolute constant $c > 0$). Our tester is almost identical to the one described in [GR00]. Now we present our main result:

Theorem 1.1 *Given any conductance parameter Φ, and any constant $\mu > 0$, there exists a constant[2] c and an algorithm with the following properties. The algorithm runs in time $O(\frac{n^{(1+\mu)/2}\log(n)\log(1/\varepsilon)}{\varepsilon\Phi^2})$ and with high probability, accepts any graph with degree bound d whose conductance is at least Φ, and rejects any graph that is ε-far from any graph of conductance at least $c\Phi^2$ with degree bound d.*

Remark: In Theorem 1.1, even though we have specified μ to be a constant, the theorem still goes through even if μ were a function of n, though naturally the conductance bound degrades. For instance, if $\mu = 1/\log(n)$, then the running time of our algorithm matches that of [CS07], but the conductance bound becomes $\Omega(\Phi^2/\log(n))$.

The following easy relations hold:

$$\text{edge expansion} \;=\; \text{conductance} \times 2d,$$

$$\frac{\text{vertex expansion}}{2} \;\geq\; \text{conductance} \;\geq\; \frac{\text{vertex expansion}}{2d}$$

Using these relations, we immediately obtain property testers for vertex and edge expansion for a given expansion parameter α by running the property tester for conductance with parameter $\Phi = \alpha/2d$, and we get the following corollary to Theorem 1.1:

Corollary 1.2 *Given any expansion parameter α, and any constant $\mu > 0$, there is an algorithm which runs in time $O(\frac{d^2 n^{(1+\mu)/2}\log(n)\log(1/\varepsilon)}{\varepsilon\alpha^2})$ and with high probability, accepts any graph with degree bound d whose expansion is at least α, and rejects any graph that is ε-far from any graph of expansion at least α' with degree bound d. For edge expansion, $\alpha' = \Omega(\frac{\alpha^2}{d})$, and for vertex expansion, $\alpha' = \Omega(\frac{\alpha^2}{d^2})$.*

[1] Henceforth, "with high probability" means with probability at least 2/3.
[2] We can set $c = \mu/400$.

Goldreich and Ron's formulation of the problem [GR00] asks for a property testing algorithm that given a parameter $\lambda < 1$, accepts any graph with second largest eigenvalue (of the transition matrix of the random walk given above) less than λ, and rejects any graph that is ε-far from having second largest eigenvalue less than λ', for some $\lambda' \leq \lambda^{\Omega(1)}$. Given a graph G, the following well known inequality (see [Sin93]) states that the second largest eigenvalue $\lambda(G)$ satisfies

$$1 - \Phi_G \ \leq \ \lambda(G) \ \leq \ 1 - \Phi_G^2/2.$$

Now, if we assume that λ is a constant less than 1, then we obtain a property tester in the Goldreich-Ron formulation, for $\lambda' = (1 - c^2(1-\lambda)^4/2) \leq \lambda^{\Omega(1)}$, since $\lambda \leq 1 - \Omega(1)$. Here, c is the constant from Theorem 1.1. We run our property tester for conductance with parameter $\Phi = 1 - \lambda$. For any graph G with $\lambda(G) \leq \lambda$, we have $\Phi_G \geq \Phi$, so the tester accepts G. Any graph G with $\Phi_G \geq c\Phi^2$ has $\lambda(G) \leq \lambda'$, so the tester rejects any graph that is ε-far from having $\lambda(G) \leq \lambda'$. Thus, we have the following corollary to Theorem 1.1:

Corollary 1.3 *Given any constant $\lambda < 1$, and any constant $\mu > 0$, there exists $\lambda' < 1$ such that $\lambda' = 1 - \Omega((1-\lambda)^4)$ and an algorithm the following properties. The algorithm runs in time $O(\frac{n^{(1+\mu)/2}\log(n)\log(1/\varepsilon)}{\varepsilon(1-\lambda)^2})$ and with high probability, accepts any graph with degree bound d with $\lambda(G) \leq \lambda$, and rejects any graph that is ε-far from having $\lambda(G) \leq \lambda'$ with degree bound d.*

2 Description of the Property Tester

We first define a procedure called VERTEX TESTER which will be used by the expansion tester.

VERTEX TESTER
Input: Vertex $v \in V$.
Parameters: $\ell = (2\ln n)/\Phi^2$ and $m = 8n^{(1+\mu)/2}$.

1. Perform m random walks of length ℓ from v.
2. Let A be the number of pairwise collisions between the endpoints of these walks.
3. The quantity $A/\binom{m}{2}$ is the *estimate* of the vertex tester. If $A/\binom{m}{2} \geq (1 + 2n^{-\mu/4})/n$, then output **Reject**, else output **Accept**.

Now, we define the Conductance Tester.

CONDUCTANCE TESTER
Input: Graph $G = (V, E)$.
Parameters: $t = \Omega(\varepsilon^{-1})$ and $N = \Omega(\log(\varepsilon^{-1}))$.

1. Choose a set S of t random vertices in V.
2. For each vertex $v \in S$:
   (a) Run VERTEX TESTER on v for N trials.
   (b) If a majority of the trials output **Reject**, then the CONDUCTANCE TESTER aborts and outputs **Reject**.
3. Output **Accept**.

3 Proof of Theorem 1.1

Before we give the details of the proof, we give a high level exposition of the ideas. We characterize vertices of the graph as *strong* or *weak* (this was already implicit in the ideas of [GR00]). Random walks of length ℓ starting from strong vertices mix very rapidly, while those from weak vertices do not. We expect the vertex tester to accept strong vertices and reject weak ones.

One of the main differences from the result of Czumaj-Sohler is that we have a very strict definition of strong vertices. We need the mixing from strong vertices to be *very rapid*, and this is what allows us to remove the dependence of n from α'. In the main technical contribution of this paper, we prove that a bad conductance cut will contain a sufficiently large number of weak vertices. We get very strong quantitative bounds using algebraic techniques to analyze the random walks starting from inside the bad cut. We then show that if there are very few weak vertices in G (and therefore, the tester will probably accept the graph), there is a patch-up procedure that can add εnd edges to boost the expansion to α' and preserves the degree bound. This completes the proof.

3.1 Preliminaries

Let us fix some notation. The probability of reaching u by performing a random walk of length l from v is $p^l_{v,u}$. Denote the (row) vector of probabilities $p^l_{v,u}$ by $\vec{p}_v^{\,l}$. The *collision probability* for random walks of length l starting from v is denoted by $\gamma_l(v)$ - this is the probability that two independent random walks of length l starting from v will end at the same vertex. It is easy to see that $\gamma_l(v) = \|\vec{p}_v^{\,l}\| = \sum_u (p^l_{v,u})^2$ (henceforth, we use $\|\cdot\|$ to denote the $\mathcal{L}_2$ norm). Let $\vec{1}$ denote the all 1's vector. The norm of the discrepancy from the stationary distribution will be denoted by $\Delta_l(v)$:

$$\Delta_l(v)^2 = \|\vec{p}_v^{\,l} - \vec{1}/n\|^2 = \sum_{u \in V} (p^l_{v,u} - 1/n)^2 = \sum_{u \in V} (p^l_{v,u})^2 - 1/n = \gamma_l(v) - 1/n.$$

Since l will usually be equal to ℓ, in that case we drop the subscripts (or superscripts). The relationship between $\Delta(v)$ and $\gamma(v)$ is central to the functioning of the tester. The parameter $\Delta(v)$ is a measure of how well a random walk from v mixes. The parameter $\gamma(v)$ can be estimated in sublinear time, and by its relationship with $\Delta(v)$, allows us to test mixing of random walks in sublinear time. The following is Lemma 1 of [GR00] (the value α in their lemma corresponds to $\mu/2$):

Lemma 3.1 *The estimate of $\gamma(v)$, viz. $A/\binom{m}{2}$, provided by the* VERTEX TESTER *lies outside the range $[(1 - n^{-\mu/4}/2)\gamma(v), (1 + n^{-\mu/4}/2)\gamma(v)]$ with probability $< 1/3$.*

For clarity of notation, we set $\sigma = n^{-\mu/4}$. We now have the following corollary:

Corollary 3.2 *Consider any set of vertices S chosen by* CONDUCTANCE TESTER. *The following holds with probability at least 5/6. For every $v \in S$, the following is true. If $\gamma(v) < (1 + \sigma)/n$, then the majority of the N trials of* VERTEX TESTER *run on v return* **Accept**. *If $\gamma(v) > (1 + 6\sigma)/n$, then the majority of the N trials of* VERTEX TESTER *run on v return* **Reject**.

Proof: Consider a fixed $v \in S$. Suppose $\gamma(v) < (1 + \sigma)/n$. Then by Lemma 3.1, the estimate provided by VERTEX TESTER is less than $(1 + 2\sigma)/n$ with probability at least 2/3. Hence VERTEX TESTER accepts with probability at least 2/3. By a Chernoff bound, a majority of the runs of VERTEX TESTER accept with probability $> 1 - \varepsilon^{-2}$ (since $N = \Omega(\log(\varepsilon^{-1}))$). If $\gamma(v) > (1 + 6\sigma)/n$, then the estimate of VERTEX TESTER is more than $(1+2\sigma)/n$ with probability at least $2/3$. Analogously, a majority of the VERTEX TESTER trials reject with probability $> 1-\varepsilon^{-2}$. By taking a union bound over the error probabilities for all $v \in S$, with probability at least $5/6$, the above is true for all $v \in S$. □

We are now ready to analyze the correctness of our tester. First, we show the easy part. Let M denote the transition matrix of the random walk. The top eigenvector of M is $\vec{1}$. We will also need the matrix $L = I - M$, which is the (normalized) Laplacian (I denotes the identity matrix). The eigenvalues of L are of the form $(1-\lambda)$, where λ is an eigenvalue of M.

Lemma 3.3 *If $\Phi_G \geq \Phi$, then the* CONDUCTANCE TESTER *accepts with probability at least* $2/3$.

Proof: Let λ_G be the second largest eigenvalue of M. It is well known (see, e.g., [Sin93]) that $\lambda_G \leq 1 - \Phi_G^2/2 \leq 1 - \Phi^2/2$. Thus, we have for any $v \in V$, if $\vec{e}_v$ denotes the row vector which is 1 on coordinate v and zero elsewhere,

$$\begin{aligned} \|\vec{p}_v - \vec{1}/n\|^2 &= \|(\vec{e}_v - \vec{1}/n)M^\ell\|^2 \\ &\leq \|\vec{e}_v - \vec{1}/n\|^2 \lambda_G^{2\ell} \\ &\leq (1-\Phi^2/2)^{4\Phi^{-2}\ln n} \\ &\leq 1/n^2. \end{aligned}$$

The second inequality follows because $\vec{e}_v - \vec{1}/n$ is orthogonal to the top eigenvector $\vec{1}$. As a result, $\Delta(v)^2 \leq 1/n^2$, and $\gamma(v) < (1+\sigma)/n$ for all $v \in V$. By Corollary 3.2, the tester accepts with probability at least $2/3$. □

Our main technical theorem is the following. Call a vertex v *weak* if $\gamma(v) > (1+6\sigma)/n$, all others will be called *strong*.

Theorem 3.4 *If there are less than $\frac{1}{25}\varepsilon n$ weak vertices, then $\varepsilon n d$ edges can be added or removed to make the conductance $\Omega(\Phi^2)$, while ensuring that all degrees are at most d.*

The remaining sections are devoted to proving this. We will show how this theorem allows us to prove the main result.

Proof: (of Theorem 1.1) If G has conductance Φ, then Lemma 3.3 completes the proof. We now show that if G is ε-far from having conductance $\Omega(\Phi^2)$, then the tester rejects with high probability. Actually, we argue the contrapositive: if the tester does *not* reject with high probability, then G is ε-close to having conductance $\Omega(\Phi^2)$. Suppose there are more than $\frac{1}{25}\varepsilon n$ weak vertices. Then with probability at least $5/6$, the random sample S chosen by the CONDUCTANCE TESTER has a weak vertex, since the sample has $\Omega(\varepsilon^{-1})$ random vertices. By Corollary 3.2, CONDUCTANCE TESTER rejects G with high probability.

Hence, if the tester does not reject with high probability, then there are at most $\frac{1}{25}\varepsilon n$ weak vertices. By Theorem 3.4, G can be modified at $\varepsilon n d$ edges make the conductance $\Omega(\Phi^2)$. □

3.2 Algebraic Lemmas

We now state and prove the key algebraic lemmas connecting bad conductance cuts to bad mixing. The quantitative bounds given here are the main tool used to prove that if the graph G has few weak vertices, then G is close to being an expander. Technically, it is Lemma 3.6 that is our main lemma. For the sake of clarity, we state a slightly simpler version below (Lemma 3.5) and prove it first. The proof of this lemma contains most of the technical ideas. We will use certain arguments made in this proof to deal with Lemma 3.6.

Lemma 3.5 *Consider a set $S \subset V$ of size $s \leq n/2$ such that the cut $(S, \overline{S})$ has conductance less than δ. Then, for any integer $l > 0$, there exists a vertex $v \in S$ such that $\Delta_l(v) > (2\sqrt{s})^{-1}(1-4\delta)^l$.*

Proof: Denote the size of S by s ($s \leq n/2$). Let us consider the starting distribution $\vec{p}$ where:

$$p_v = \begin{cases} 1/s & v \in S \\ 0 & v \notin S \end{cases}$$

Let $\vec{u} = \vec{p} - \vec{1}/n$. Note that $\vec{u}M^l = \vec{p}M^l - \vec{1}/n$. Let $1 = \lambda_1 \geq \lambda_2 \cdots \geq \lambda_n > 0$ be the eigenvalues of M and $\vec{f_1}, \vec{f_2}, \cdots, \vec{f_n}$ be the corresponding orthogonal unit eigenvectors. Note that $\vec{f_1} = \vec{1}/\sqrt{n}$. We represent $\vec{u}$ in the orthonormal basis formed by the eigenvectors of M as $\vec{u} = \sum_i \alpha_i \vec{f_i}$. (As an aside note that $\alpha_1 = 0$, since $\vec{u} \cdot \vec{1} = 0$.)

$$\begin{aligned} \sum_i \alpha_i^2 &= \|\vec{u}\|_2^2 \\ &= s\left(\frac{1}{s} - \frac{1}{n}\right)^2 + \frac{n-s}{n^2} \\ &= \frac{1}{s} - \frac{1}{n}. \end{aligned}$$

Taking the Rayleigh quotient with the Laplacian L:

$$\begin{aligned} \vec{u}^\top L \vec{u} &= \vec{u}^\top I \vec{u} - \vec{u}^\top M \vec{u} \\ &= \|\vec{u}\|_2^2 - \sum_i \alpha_i^2 \lambda_i. \end{aligned}$$

By the properties of the graph Laplacian (refer to Equation 1.1 in Chapter 1 of [Chu92]),

$$\vec{u}^\top L \vec{u} = \sum_{i<j} M_{ij}(u_i - u_j)^2$$

We now bound the right hand side. By construction of u, only the edges $\{i,j\} \in E(S, \bar{S})$ contribute to the above sum. Each edge contributes $1/(2ds^2)$. Since the conductance of the cut $(S, \bar{S})$ is less than δ, $|E(S, \bar{S})| \leq 2\delta ds$. Hence,

$$\vec{u}^\top L \vec{u} < 2\delta ds \times \frac{1}{2d} \times \frac{1}{s^2} = \frac{\delta}{s}.$$

Putting the above together:

$$\begin{aligned} \sum_i \alpha_i^2 \lambda_i &> \left(\frac{1}{s} - \frac{1}{n}\right) - \frac{\delta}{s} \\ &= \frac{1-\delta}{s} - \frac{1}{n}. \end{aligned}$$

If $\lambda_i > (1-4\delta)$, call it *heavy*. Let H be the index set of heavy eigenvalues, and $\bar{H}$ be the index set of the rest. Since $\sum_i \alpha_i^2 \lambda_i$ is large, we expect many of the α_i corresponding to heavy eigenvalues to be large. This would ensure that the starting distribution $\vec{p}$ will not mix rapidly. We have

$$\sum_{i \in H} \alpha_i^2 \lambda_i + \sum_{i \in \bar{H}} \alpha_i^2 \lambda_i \;>\; \frac{1-\delta}{s} - \frac{1}{n}.$$

Setting $x = \sum_{i \in H} \alpha_i^2$:

$$x + (\sum_i \alpha_i^2 - x)(1-4\delta) \;>\; \frac{1-\delta}{s} - \frac{1}{n}.$$

We therefore get:

$$\begin{aligned} 4\delta x + \left(\frac{1}{s} - \frac{1}{n}\right)(1-4\delta) &> \frac{1-\delta}{s} - \frac{1}{n} \\ \therefore\; x &> \frac{3}{4s} - \frac{1}{n} \\ &\geq \frac{1}{4s}. \qquad \because n \geq 2s \end{aligned} \tag{1}$$

Note that $\vec{u}M^l = \sum_i \alpha_i \lambda_i^l \vec{f_i}$. Thus,

$$\begin{aligned} \|\vec{u}M^l\|_2^2 &= \sum_i \alpha_i^2 \lambda_i^{2l} \\ &\geq \sum_{i \in H} \alpha_i^2 \lambda_i^{2l} \\ &> \frac{1}{4s}(1-4\delta)^{2l}. \end{aligned}$$

So, $\|\vec{u}M^l\|_2 > \frac{1}{2\sqrt{s}}(1-4\delta)^l$. Note that $\vec{u} = \frac{1}{s}\sum_{v \in S}(\vec{e}_v - \frac{\vec{1}}{n})$, and hence $\vec{u}M^l = \frac{1}{s}\sum_{v \in S}(\vec{e}_v M^l - \frac{\vec{1}}{n})$. Now, $\vec{e}_v M^l - \frac{\vec{1}}{n}$ is the discrepancy vector of the probability distribution of the random walk starting from v after l steps. Thus, by Jensen's inequality, we conclude that

$$\frac{1}{s}\sum_{v \in S} \Delta_l(v) \;\geq\; \|\vec{u}M^l\| \;>\; \frac{1}{2\sqrt{s}}(1-4\delta)^l.$$

Hence, there is some $v \in S$ for which $\Delta_l(v) > (2\sqrt{s})^{-1}(1-4\delta)^l$. □

The following is the main lemma that will be applied later. As mentioned earlier, this proof will follow many of the arguments used above.

Lemma 3.6 *Consider sets $T \subseteq S \subseteq V$ such that the cut $(S, \bar{S})$ has conductance less than δ. Let $|T| = (1-\theta)|S|$. Assume $0 < \theta \leq \frac{1}{8}$. Then, for any integer $l > 0$, there exists a vertex $v \in T$ such that $\Delta_l(v) > \frac{(1-2\sqrt{2\theta})}{2\sqrt{s}}(1-4\delta)^l$.*

Proof: Let $\vec{u}_S$ (resp., $\vec{u}_T$) be the uniform distribution over S (resp., T) minus $\frac{\vec{1}}{n}$. Let s and t be the sizes of S and T resp. Let $\vec{u}_S = \sum_i \alpha_i \vec{f_i}$ and $\vec{u}_T = \sum_i \beta_i \vec{f_i}$ be representation of $\vec{u}_S$ and $\vec{u}_T$ in the basis $\{\vec{f_1}, \ldots, \vec{f_n}\}$, the unit eigenvectors of M.

Since the conductance of S is less than δ, by applying inequality (1) from Lemma 3.5, we have that

$$\sum_{i \in H} \alpha_i^2 > \frac{1}{4s}.$$

We have

$$\|\vec{u}_S - \vec{u}_T\|^2 \;=\; \frac{1}{t} - \frac{1}{s} \;=\; \frac{\theta}{(1-\theta)s} \;\leq\; \frac{2\theta}{s}.$$

Furthermore,

$$\|\vec{u}_S - \vec{u}_T\|^2 \;=\; \sum_i (\alpha_i - \beta_i)^2 \;\geq\; \sum_{i \in H} (\alpha_i - \beta_i)^2.$$

Using the triangle inequality $\|\vec{a} - \vec{b}\| \geq \|\vec{a}\| - \|\vec{b}\|$, we get that

$$\sum_{i \in H} \beta_i^2 \;\geq\; \left[\sqrt{\sum_{i \in H} \alpha_i^2} - \sqrt{\sum_{i \in H} (\alpha_i - \beta_i)^2}\right]^2 \;>\; \left[\frac{1}{2\sqrt{s}} - \frac{\sqrt{2\theta}}{\sqrt{s}}\right]^2 \geq \frac{(1 - 2\sqrt{2\theta})^2}{4s}.$$

Finally, reasoning as in Lemma 3.5, we get that $\|\vec{u}_T M^l\| > \frac{(1-2\sqrt{2\theta})}{2\sqrt{s}}(1-4\delta)^l$, and thus, by Jensen's inequality, there is a $v \in T$ such that $\Delta_l(v) > \frac{(1-2\sqrt{2\theta})}{2\sqrt{s}}(1-4\delta)^l$. □

This lemma immediately yields the following corollary:

Corollary 3.7 *Consider a set $S \subseteq V$ such that the cut $(S, \bar{S})$ has conductance less than δ. For positive $\theta \leq \frac{1}{8}$ and any integer $l > 0$, there exist at least $\theta|S|$ vertices $v \in S$ such that $\Delta_l(v) > \frac{(1-2\sqrt{2\theta})}{2\sqrt{s}}(1-4\delta)^l$.*

Using the above lemmas, we can now show that G looks almost like an expander.

Lemma 3.8 *Suppose there are more than $\frac{1}{25}\varepsilon n$ weak vertices. Then there is a partition of the graph G into two pieces, A and $\bar{A} := V \setminus A$, with the following properties:*

1. *$|A| \leq \frac{2}{5}\varepsilon n$.*
2. *Any cut in the induced subgraph on $\bar{A}$ has conductance $\Omega(\Phi^2)$.*

Proof: We use a recursive partitioning technique: start out with $A = \{\}$. Let $\bar{A} = V \setminus A$. If there is a cut $(S, \bar{S})$ in $\bar{A}$ with $|S| \leq |\bar{A}|/2$ with conductance less than $c\Phi^2$, then we set $A := A \cup S$, and continue as long as $|A| \leq n/2$. Here, c is a small constant to be chosen later.

We claim that the final set A has the required properties. If $|A| > \frac{2}{5}\varepsilon n$, then consider the cut $(A, \bar{A})$ in G. Suppose A was formed by the (disjoint) union of $A_1, A_2, \cdots$. We can bound

$$|E(A, \bar{A})| \leq \sum_i |E(A_i, \bar{A}_i)| \leq c\Phi^2 \times 2d \sum_i |A_i| \leq c\Phi^2(2d|A|)$$

Hence, the set A has conductance at most $c\Phi^2$. Now, Corollary 3.7 implies (with $\theta = 1/10$) that there are at least $\frac{1}{10}|A| > \frac{1}{25}\varepsilon n$ vertices in A such that for all such vertices v, and for $b = \frac{(1-2\sqrt{1/5})}{\sqrt{2}}$, we have

$$\Delta_\ell(v) \;>\; \frac{b}{\sqrt{n}}(1 - 4c\Phi^2)^\ell \;>\; \sqrt{6\sigma/n}$$

for a suitably small choice of c in terms of μ (say, $c = \mu/200$ suffices).

Thus, for all such vertices v, we have $\gamma_\ell(v) = \Delta_\ell(v)^2 + 1/n > (1 + 6\sigma)/n$, which implies that all such vertices are weak, a contradiction since there are only $\frac{1}{25}\varepsilon n$ weak vertices.

Since $|A| \le \frac{2}{5}\varepsilon n < n/2$, when the recursive partitioning procedure terminates, any cut in the induced subgraph on $\bar{A}$ has conductance $\Omega(\Phi^2)$. □

3.3 Getting an expander

Armed with the partitioning algorithm of Lemma 3.8, we are ready to present the patch-up algorithm, which changes the graph in εnd edges and raises its conductance to $\Omega(\Phi^2)$. Note that we do not perform this patch-up algorithm as part of our tester. It is merely used to show that G is close to an expander. The trivial patch-up algorithm would just add d random edges to every vertex in A. This would only add at most εnd edges and make the conductance $\Omega(\Phi^2)$. The drawback is that the degree bound will not be preserved. We have to be more careful to ensure that we can find a graph G' ε-close to G which is an expander *and* has a degree bound of d.

PATCH-UP ALGORITHM

1. Partition the graph into two pieces A and $\bar{A}$ with the properties given in Lemma 3.8.
2. Remove all edges incident on vertices in A.
3. For each vertex $u \in A$, repeat the following process until the degree of u becomes $d-1$ or d: choose a vertex $v \in \bar{A}$ at random. If the current degree of v is less than d, add the edge $\{u, v\}$. Otherwise, if there is an edge $\{v, w\}$ such that $w \in \bar{A}$, remove $\{v, w\}$, and add the edges $\{u, v\}$ and $\{u, w\}$ (call these newly added edges "paired"). Otherwise, re-sample the vertex v from $\bar{A}$, and repeat.

To implement Step 3, we need to ensure that the set of vertices in $\bar{A}$ with degree less than d or having an edge to another vertex in $\bar{A}$ is non-empty. In fact, we can show a stronger fact:

Lemma 3.9 *Suppose there are more than $\frac{1}{25}\varepsilon n$ weak vertices. At any stage in the patch-up algorithm, there are at least $\frac{1}{4}|\bar{A}| \ge \frac{1}{4}(1 - 2\varepsilon/5)n$ vertices in $\bar{A}$ with degree less than d or having an edge to another vertex in $\bar{A}$.*

Proof: Let $X \subseteq \bar{A}$ be the set of vertices of degree at most $d/2$ before starting the third step, and let $Y := \bar{A} \setminus X$. Now we have two cases:

1. $|X| \ge \frac{1}{2}|\bar{A}|$: We add at most $\frac{2}{5}\varepsilon nd$ edges, since $|A| \le \frac{2}{5}\varepsilon n$. At most half the vertices in X can have their degree increased to d, since $\frac{2}{5}\varepsilon nd \le \frac{1}{2}|X| \cdot \frac{d}{2}$, since $|X| \ge \frac{1}{2}(1 - 2\varepsilon/5)n$. Here, we assume that $\varepsilon \le 1/4$. Thus, at any stage we have at least $\frac{1}{4}|\bar{A}|$ vertices with degree less than d.
2. $|Y| \ge \frac{1}{2}|\bar{A}|$: we remove at most $\frac{1}{5}\varepsilon nd$ edges from the subgraph induced by $\bar{A}$. At most half of the vertices in Y can have their (induced) degrees reduced to 0, $\frac{1}{5}\varepsilon nd \le \frac{1}{2}|Y| \cdot \frac{d}{2}$, since $|Y| \ge \frac{1}{2}(1 - 2\varepsilon/5)n$. Again, we assume that $\varepsilon \le 1/4$. Thus, at any stage we have at least $\frac{1}{4}|\bar{A}|$ vertices with at least one edge to some other vertex in $\bar{A}$.

Now, we prove our main theorem:

Proof: (of Theorem 3.4) We run the patch-up algorithm on the given graph. It is easy to see that at the end of the algorithm, every vertex has degree bounded by d. Also, the total number of edges deleted is at most $\frac{2}{5}\varepsilon nd + \frac{1}{5}\varepsilon nd$, and the number of edges added is at most $\frac{2}{5}\varepsilon nd$. Thus the total number of edges changed is at most εnd.

Fix a set S of vertices. We will argue that with high probability over the randomness of the patch-up algorithm, S has high conductance. Taking a union bound over all (appropriate) sets, we will show that with non-zero probability, the graph output by the patch-up algorithm has high conductance. By the probabilistic method, this completes the proof.

Let set S be such that $|S| \leq n/2$. Let $S_A = S \cap A$, and $S_{\bar{A}} = S \cap \bar{A}$. Let $m := |S|$. We have two cases now:

1. $|S_{\bar{A}}| \geq m/2$: Suppose that $|S_{\bar{A}}| \leq |\bar{A} \setminus S_{\bar{A}}|$. In this case, note that in the subgraph of original graph induced by $\bar{A}$, the set $S_{\bar{A}}$ had conductance at least $c\Phi^2$. Hence the cut $(S_{\bar{A}}, \bar{A} \setminus S_{\bar{A}})$ had at least $2c\Phi^2|S_{\bar{A}}|d \geq c\Phi^2 md$ edges crossing it. Now, suppose $|S_{\bar{A}}| > |\bar{A} \setminus S_{\bar{A}}|$. By Lemma 3.8, $|\bar{A}| \geq (1 - 2\varepsilon/5)n$. Since $|S| \leq n/2$, $|\bar{A} \setminus S_{\bar{A}}| \geq n/3 \geq 2m/3$ (for sufficiently small ε). In the subgraph of the original graph induced by $\bar{A}$, $\bar{A} \setminus S_{\bar{A}}$ has conductance at least $c\Phi^2$. Arguing as earlier, the cut $(S_{\bar{A}}, \bar{A} \setminus S_{\bar{A}})$ had at least $2c\Phi^2|\bar{A} \setminus S_{\bar{A}}|d \geq c\Phi^2 md$ edges crossing it.

   For any edge $\{v, w\}$ that was in the cut $(S_{\bar{A}}, \bar{A} \setminus S_{\bar{A}})$ and was removed by the construction, we added two new edges $\{u, v\}$ and $\{u, w\}$ for some $u \in A$. Regardless of whether $u \in S_A$ or $u \notin S_A$, one of the two edges $\{u, v\}$ and $\{u, w\}$ crosses the cut $(S, \bar{S})$. Thus, at least $c\Phi^2 md$ edges cross the cut $(S, \bar{S})$, and hence it has conductance at least $\frac{c}{2}\Phi^2$. Note that this is independent of the randomness of the patch-up algorithm.

2. $|S_{\bar{A}}| \leq m/2$: We will focus on the number of edges that the patch-up algorithm adds to $S_{\bar{A}}$. For each vertex $u \in S_A$, some edges are randomly added to u. A vertex v is chosen uniformly at random, and if the adjacency conditions are satisfied, then the edge $\{u, v\}$ is added. This is repeated at least $d/2$ times. By Lemma 3.9, at any stage, there are at least $\frac{1}{4}(1-2\varepsilon/5)n \geq n/5$ vertices that can be connected to. Of these, at most m are in S. Therefore, whenever a (valid) vertex v is chosen, with probability at least $1 - 5m/n$, an edge is added to the cut $(S, \bar{S})$. This is done at least $d/2$ times for each vertex $u \in S_A$, and we have $|S_A| > m/2$. In other words, there are $md/4$ independent random edge additions, and for each, with probability least $1 - 5m/n$, the edge crosses the cut.

   It will be convenient deal with the random variable X, the number of these edges that do *not* cross the cut. This is just the sum of independent $0 - 1$ random variables. We have $\mathbf{E}[X] \leq 5dm^2/(4n)$. By a Chernoff bound (Theorem 4.1 of [MR00]), for any positive δ,

   $$\Pr[X > (1+\delta)5dm^2/(4n)] < \left[\frac{e^\delta}{(1+\delta)^{1+\delta}}\right]^{5dm^2/(4n)}$$

   Since $|S_{\bar{A}}| \leq |S_A| \leq |A| \leq 2\varepsilon n/5$, we can bound $m \leq 4\varepsilon n/5$. We set $(1+\delta) = n/(10m)$. Since ε is a sufficiently small constant, we have $\delta > 0$. We get

   $$\Pr[|E(S, \bar{S})| \leq md/8] \leq \Pr[X > md/8] < \left(\frac{n}{10em}\right)^{-md/8}$$

We now take a union bound over all sets S. For case 1, the set S has conductance at least $\Omega(\Phi^2)$. For case 2, we have an upper bound on the probability that S has conductance $< 1/16$. There are

$\binom{n}{m}$ sets of size m. We assume that d is a sufficiently large constant. The probability that the final graph has conductance less than $c\Phi^2/2$ is at most

$$\sum_{m \leq 4\varepsilon n/5} \binom{n}{m} \left(\frac{10em}{n}\right)^{md/8} \leq \sum_{m \leq 4\varepsilon n/5} \left(\frac{en}{m}\right)^{m} \left(\frac{10em}{n}\right)^{md/8} \leq \sum_{m \leq 4\varepsilon n/5} \frac{(10e)^{md/4}}{(n/m)^{md/10}} < 1$$

□

References

[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. Conference version in STOC 2006.

[AS08] N. Alon and A. Shapira. A charaterization of the (natural) graph properties testable with one-sided error. *SIAM Journal on Computing*, 37(6):1703–1727, 2008. Conference version in FOCS 2005.

[BSS10] I. Benjamini, O. Schramm, and A. Shapira. Every minor-closed property of sparse graphs is testable. *Advances in Mathematics*, 223(6):2200–2218, 2010. Conference version in STOC 2009.

[Chu92] F. Chung. *Spectral Graph Theory.* AMS Publications, 1992. Available at http://www.math.ucsd.edu/∼fan/research/revised.html.

[CS07] A. Czumaj and C. Sohler. Testing expansion in bounded degree graphs. *Proceedings of the 48th Annual Annual Foundations of Computer Science (FOCS)*, pages 570–578, 2007.

[CSS09] A. Czumaj, A. Shapira, and C. Sohler. Testing hereditary properties of nonexpanding bounded-degree graphs. *SIAM Journal on Computing*, 38(6):2499–2510, 2009. Conference version titled "On Testable Properties in Bounded Degree Graphs" in SODA 2007.

[Fis01] E. Fischer. The art of uninformed decisions: A primer to property testing. *Bulletin of EATCS*, 75:97–126, 2001.

[GGR98] O. Goldreich, S. Goldwasser, and D. Ron. Property testing and its connection to learning and approximation. *Journal of the ACM*, 45(4):653–750, 1998. Conference version in FOCS 1996.

[Gol98] O. Goldreich. Combinatorial property testing - a survey. *Randomization Methods in Algorithm Design*, 75:45–60, 1998.

[Gol10] O. Goldreich. Introduction to testing graph properties. Technical Report TR10-082, ECCC, 2010.

[GR99] O. Goldreich and D. Ron. A sublinear bipartite tester for bounded degree graphs. *Combinatorica*, 19(3):335–373, 1999. Conference version in STOC 1998.

[GR00] O. Goldreich and D. Ron. On testing expansion in bounded-degree graphs. *ECCC*, TR00-020, 2000.

[GR02] O. Goldreich and D. Ron. Property testing in bounded degree graphs. *Algorithmica*, 32(2):302–343, 2002. Conference version in STOC 1997.

[KS07] S. Kale and C. Seshadhri. Testing expansion in bounded degree graphs. Technical Report TR07-076, ECCC, 2007.

[MR00] R. Motwani and P. Raghavan. *Randomized Algorithms*. Cambridge University Press, 2000.

[NS07] A. Nachmias and A. Shapira. Testing the expansion of a graph. *ECCC*, TR07-118, 2007.

[Ron01] D. Ron. Property testing. *Handbook on Randomization*, II:597–649, 2001.

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

[RS96] R. Rubinfeld and M. Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM Journal on Computing*, 25(2):252–271, 1996.

[Sin93] A. Sinclair. *Algorithms for Random Generation and Counting: a Markov Chain Approach*. Birkhaser Progress In Theoretical Computer Science Series, 1993.