An expansion tester for bounded degree graphs

Satyen Kale*[1] and C. Seshadhri[2]

[1] Microsoft Research
One Microsoft Way
Redmond, WA 98052
satyen.kale@microsoft.com
[2] Dept. of Computer Science,
Princeton University
35 Olden St, Princeton, NJ 08540
csesha@cs.princeton.edu

Abstract. We consider the problem of testing graph expansion (either vertex or edge) in the bounded degree model [10]. We give a property tester that given a graph with degree bound d, an expansion bound α, and a parameter $\varepsilon > 0$, 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 $\tilde{O}(\frac{n^{(1+\mu)/2}d^2}{\varepsilon\alpha^2})$ for any given constant $\mu > 0$.

1 Introduction

With the presence of large data sets, reading the whole input may be a luxury. It becomes important to design algorithms which 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* [7, 14], we wish to accept inputs that satisfy some given 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 [5, 6, 13]). For property testing in graphs [7], 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 ([1, 2]).

The problem of property testing for bounded degree graphs was first dealt with by Goldreich and Ron [8]. The input graph G is assumed that 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

* Part of this work was done when the author was at Princeton University.

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 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, there aren't any general results about testable properties as in the case of dense graphs. Czumaj and Sohler [3] made the first attempt in this direction, and showed testability results for classes of graphs that do not contain expanders. Using random walks, Goldreich and Ron [9] proved that bipartiteness is testable with $\widetilde{O}(\sqrt{n})$ queries to the graph. In later work, Goldreich and Ron [10] posed the question of testing *expansion*. Given positive parameters $\lambda, \varepsilon < 1$, they provided a $\widetilde{O}(\sqrt{n})$-time algorithm that was *conjectured* to accept every graph G whose second largest eigenvalue $\lambda(G)$ is less than λ, and 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 [9].

One of the major parts of the analysis of the algorithm of [9] for bipartiteness deals with the expansion properties of the graph. Their main technique involves performing random walks on the graph. In the adjacency list model, the basic operation that we possess is that of walking in the graph, and random walks seem like a very natural operation to perform. This immediately raises the question of whether random walks can be used to test expansion. Furthermore, the results of [3] show that classes of graphs which do *not* contain expanders can be tested. All of this indicates that, regarding property testing in bounded degree graphs, testing expansion is a very natural and central issue. The problem of designing a property tester for expansion remained open for more than 6 years, until recently, when Czumaj and Sohler [4] 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 nodes in $\bar{S}$ that are adjacent to nodes 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 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 [4]. Their algorithm was based on that of Goldreich and Ron [10], 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 [11] which 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 which 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 tech report on ECCC [11], 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 [12] 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 node. Thus, for a vertex of degree $d' \leq d$, the probability of 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[3]) accept if $\Phi_G > \Phi$ and reject if G is ε-far from having $\Phi_G > c\Phi^2$ (for some absolute constant c). Our tester is almost identical to the one described in [10]. Now we present our main result:

Theorem 1. *Given any conductance parameter Φ, and any constant $\mu > 0$, there is an algorithm which 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, where c is a constant[4] which depends on μ.*

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

REMARK: In Theorem 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 [4], but the conductance bound becomes $\Omega(\Phi^2/\log(n))$.

In our bounded degree graph model, the following easy relations hold:

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

$$(\text{vertex expansion})/2 \;\geq\; \text{conductance} \;\geq\; (\text{vertex expansion})/2d.$$

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:

Corollary 1. *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, $\alpha' = \Omega(\frac{\alpha^2}{d^2})$.*

Goldreich and Ron's formulation of the problem [10] 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 lazy random walk) 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 [15]) 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 $\lambda \leq 1 - \alpha$ for some constant $\alpha > 0$, 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. 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 which is ε-far from having $\lambda(G) \leq \lambda'$. Thus, we have the following corollary to Theorem 1:

Corollary 2. *Given any parameter $\lambda < 1 - \Omega(1)$, and any constant $\mu > 0$, there is an algorithm which 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, for some $\lambda' \leq \lambda^{\Omega(1)}$.*

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 s.
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 $A/\binom{m}{2} \geq (1 + 2n^{-\mu})/n$, 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

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 [10]). 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 ${\boldsymbol{p}_v}^l$. 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 $\gamma_l(v) = \|\boldsymbol{p}_v\| = \sum_v (p^l_{v,u})^2$ (henceforth, we use $\|\cdot\|$ to denote the $\mathcal{L}_2$ norm). Let $\mathbf{1}$ denote the all 1's vector. The norm of the discrepancy from the stationary distribution will be denoted by $\Delta_l(v)$:

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

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 s 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 basically proven in [10]:

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

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

Corollary 3. *The following holds with probability at least* $5/6$*. For all vertices* v *in the random sample* S *chosen by the* Conductance Tester*, 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**.

This is an easy consequence of the fact that we run $N = \Omega(\log(\varepsilon^{-1}))$ trials, by an direct application of Chernoff's bound and using Lemma 1. 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 $\mathbf{1}$. 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 2. *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., [15]) that $\lambda_G \leq 1 - \Phi_G^2/2 \leq 1 - \Phi^2/2$. Thus, we have for any $v \in V$, if $\boldsymbol{e}_v$ denotes the row vector which is 1 on coordinate v and zero elsewhere,

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

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

We now show that if G is ε-far from having conductance $\Omega(\Phi^2)$, then the tester rejects with high probability. Actually, we will prove the contrapositive : if the tester does *not* reject with high probability, then G is ε-close to having conductance $\Omega(\Phi^2)$. Call a vertex s *weak* if $\gamma(v) > (1+6\sigma)/n$, all others will be called *strong*. 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. Thus, the CONDUCTANCE TESTER will reject with high probability.

Let us therefore assume that there are at most $\frac{1}{25}\varepsilon n$ weak vertices. Now, we will show that εnd edges can be added to 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.

Lemma 3. *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 node $v \in S$ 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 $\boldsymbol{p}$ where:

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

Let $\boldsymbol{u} = \boldsymbol{p} - \mathbf{1}/n$. Note that $\boldsymbol{u}M^l = \boldsymbol{p}M^l - \mathbf{1}/n$. Let $1 = \lambda_1 \geq \lambda_2 \cdots \geq \lambda_n > 0$ be the eigenvalues of M and $\boldsymbol{f}_1, \boldsymbol{f}_2, \cdots, \boldsymbol{f}_n$ be the corresponding orthogonal unit eigenvectors. Note that $\boldsymbol{f}_1 = \mathbf{1}/\sqrt{n}$. We represent $\boldsymbol{u}$ in the orthonormal basis formed by the eigenvectors of M as $\boldsymbol{u} = \sum_i \alpha_i \boldsymbol{f}_i$. Here, $\alpha_1 = 0$, since $\boldsymbol{u} \cdot \mathbf{1} = 0$.

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

Taking the Rayleigh quotient with the Laplacian L:

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

On the other hand, using the fact that the conductance of the cut $(S, \bar{S})$ is less than δ, we have

$$\boldsymbol{u}^\top L\boldsymbol{u} \;=\; \sum_{i<j} M_{ij}(u_i - u_j)^2 \;<\; 2\delta ds \times \frac{1}{2d} \times \frac{1}{s^2} \;=\; \frac{\delta}{s}.$$

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 $\boldsymbol{p}$ 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:

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

Note that $\boldsymbol{u}M^l = \sum_i \alpha_i \lambda^l \boldsymbol{f}_i$. Thus,

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

So, $\|\boldsymbol{u}M^l\|_2 > \frac{1}{2\sqrt{s}}(1-4\delta)^l$. Note that $\boldsymbol{u} = \frac{1}{s}\sum_{v\in S}(\boldsymbol{e}_v - \frac{\boldsymbol{1}}{n})$, and hence $\boldsymbol{u}M^l = \frac{1}{s}\sum_{v\in S}(\boldsymbol{e}_v M^l - \frac{\boldsymbol{1}}{n})$. Now, $\boldsymbol{e}_v M^l - \frac{\boldsymbol{1}}{n}$ 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

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

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

Lemma 4. *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 node $v \in T$ such that $\Delta_l(v) > \frac{(1-2\sqrt{2\theta})}{2\sqrt{s}}(1-4\delta)^l$.*

Proof. Let $\boldsymbol{u}_S$ (resp., $\boldsymbol{u}_T$) be the uniform distribution over S (resp., T) minus $\frac{\mathbf{1}}{n}$. Let s and t be the sizes of S and T resp. Let $\boldsymbol{u}_S = \sum_i \alpha_i \boldsymbol{f}_i$ and $\boldsymbol{u}_T = \sum_i \beta_i \boldsymbol{f}_i$ be representation of $\boldsymbol{u}_S$ and $\boldsymbol{u}_T$ in the basis $\{\boldsymbol{f}_1, \ldots, \boldsymbol{f}_n\}$, the unit eigenvectors of M. Note that $\alpha_1 = \beta_1 = 0$ since $\boldsymbol{u}_S$ and $\boldsymbol{u}_T$ are orthogonal to $\mathbf{1}$.

Since the conductance of S is less than δ, by applying inequality (1) from Lemma 3, we have that

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

We have

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

Furthermore,

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

Using the triangle inequality $\|\boldsymbol{a}-\boldsymbol{b}\| \geq \|\boldsymbol{a}\| - \|\boldsymbol{b}\|$, 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, we get that $\|\boldsymbol{u}_T M^l\| > \frac{(1-2\sqrt{2\theta})}{2\sqrt{s}}(1-\delta)^l$, 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 4. *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 $\theta|S|$ nodes $v \in S$ 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 5. *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. It has conductance at most $c\Phi^2$. Now, Corollary 4 implies (with $\theta = 1/10$) that there are at least $\frac{1}{10}|A| > \frac{1}{25}\varepsilon n$ nodes in A such that for all such nodes 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 suitable choice of c in terms of μ (say, $c = \mu/200$ suffices).

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

Since $|A| \leq \frac{2}{5}\varepsilon n < n/2$, 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 5, 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 5.
2. Remove all edges incident on nodes in A.
3. For each node $u \in A$, 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 nodes in $\bar{A}$ with degree less than d or having an edge to another node in $\bar{A}$ is non-empty. In fact, we can show a stronger fact:

Lemma 6. *At any stage in the patch-up algorithm, there are at least* $\frac{1}{4}|\bar{A}| \geq \frac{1}{4}(1-2\varepsilon/5)n$ *nodes in* $\bar{A}$ *with degree less than* d *or having an edge to another node in* $\bar{A}$.

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

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

Now, we prove that the patch-up algorithm works:

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

Proof. We run the patch-up algorithm on the given graph. It is easy to see that at the end of the algorithm, every node 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.

Now, let $(S, \bar{S})$ be a cut in the graph with $|S| \leq n/2$. Let $S_A = S \cap A$, and $S_{\bar{A}} = S \cap \bar{A}$. Let $m := |S|$. We have two cases now:

1. $|S_{\bar{A}}| \geq m/2$: In this case, note that in the subgraph of original graph induced on $\bar{A}$, the set $S_{\bar{A}}$ had conductance at least $c\Phi^2$, and 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.
   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 $u \in A$. Now it is easy to check that 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 $\frac{c}{2}\Phi^2$.
2. $|S_{\bar{A}}| \leq m/2$: In this case, for each node $u \in S_A$, we chose at least $d/2$ random edges connecting u to nodes in $\bar{A}$ (for now, disregarding one edge in every set of paired edge from step 3.). By Lemma 6, and since $|S_{\bar{A}}| \leq |S_A| \leq |A| \leq 2\varepsilon n/5$, the probability that for any such edge, the endpoint in $\bar{A}$ was actually in $S_{\bar{A}}$ is at most

$$\frac{|S_{\bar{A}}|}{\frac{1}{4}|\bar{A}|} \leq \frac{2\varepsilon/5}{\frac{1}{4}(1-2\varepsilon/5)} \leq 1/4$$

assuming $\varepsilon \leq 1/8$.

Since $|S_A| \geq m/2$, the total number of edges added to nodes in S_A is at least $md/4$ (again, disregarding one edge out of every set of paired edges). The expected number of these edges going into $S_{\bar{A}}$ is at most $md/16$. By the Chernoff-Hoeffding bounds, the probability that more than $md/8$ randomly chosen edges lie completely in S is less than $n^{-\Omega(md)} \leq 1/3n^{m+1}$, if we assume d is at least a large enough constant.

Taking a union bound over all sets of size m (the number of which is at most n^m), and then summing over all m, we get the with probability at least 2/3, none of these events happen, and thus at least at least $md/8$ edges cross the cut $(S, \overline{S})$. Therefore, the conductance of this cut is at least $1/16 > \Omega(\Phi^2)$, since $\Phi \leq 1$.

□

References

1. Alon, N., Fischer, E., Newman, I., Shapira, A., *A Combinatorial Characterization of the Testable Graph Properties : it's all about Regularity* Proc. 38th STOC, 2006, 251–260.
2. Alon, N., Shapira, A., *A Charaterization of the (Natural) Graph Properties Testable with One-sided Error* Proc. 46th FOCS, 2005, 429–438.
3. Czumaj, A., Sohler, C. *On Testable Properties in Bounded Degree Graphs*, Proc. 18th SODA, 2007, 494–501.
4. Czumaj, A., Sohler, C. *Testing Expansion in Bounded Degree Graphs*, Proc. 48th FOCS, 2007, 570–578.
5. Fischer, E. *The Art of Uninformed Decisions: A Primer to Property Testing*, Bulletin of EATCS, 75: 97-126, 2001.
6. Goldreich, O. *Combinatorial property testing - A survey*, in "Randomization Methods in Algorithm Design," 45-60, 1998.
7. Goldreich, O., Goldwasser, S., Ron, D. *Property Testing and its Connection to Learning and Approximation*, J. ACM 45 (1998), 653–750.
8. Goldreich, O., Ron, D. *Property Testing in Bounded Degree Graphs*, Algorithmica, 32(2), 2002, 302–343.
9. Goldreich, O., Ron, D. *A Sublinear Bipartiteness Property Tester for Bounded Degree Graphs*, Combinatorica, 19(3), 1999, 335–373.
10. Goldreich, O., Ron, D. *On Testing Expansion in Bounded-Degree Graphs*, ECCC, TR00-020, 2000.
11. Kale, S., Seshadhri, C. *Testing Expansion in Bounded Degree Graphs*, ECCC, TR07-076, 2007.
12. Nachmias, A., Shapira, A. Testing the Expansion of a Graph ECCC, TR07-118, 2007.
13. Ron, D. *Property testing*, in "Handbook on Randomization," Volume II, 597-649, 2001.
14. Rubinfeld, R., Sudan, M. *Robust characterization of polynomials with applications to program testing*, SIAM J. Comput. 25 (1996), 647–668.
15. Sinclair, A. *Algorithms for Random Generation and Counting: a Markov Chain Approach*, Birkhaser Progress In Theoretical Computer Science Series, 1993.