Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 76 (2007)

Testing Expansion in Bounded Degree Graphs

Satyen Kale*
Microsoft Research
One Microsoft Way
Redmond, WA 98052
satyen.kale@microsoft.com

C. Seshadhri
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 [2]. 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 a 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 Property Testing of Expansion

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 εnd edges (either removing or adding) to make the conductance at least α'.

This problem was first defined by Goldreich and Ron [2], where an approach was described towards designing the required property tester. They proposed an algorithm, but their analysis relied on an unproven combinatorial conjecture. Our algorithm uses the same ideas as their paper, but we use algebraic techniques to prove different combinatorial results which suffice to complete the analysis. Progress towards this was made by Czumaj and Sohler [1], where the obtained a property tester for vertex expansion with $\alpha' = \Theta(\frac{\alpha^2}{d^2 \log n})$, using combinatorial techniques. Independently, we obtained the improvement $\alpha' = \Theta(\frac{\alpha^2}{d^2})$ for vertex expansion and $\alpha' = \Theta(\frac{\alpha^2}{d})$ for edge expansion using algebraic techniques.

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

ISSN 1433-8092

However, the first version of this paper could only 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. This revised version of the paper gives a small modification to the earlier paper that improves the degree bound to d. We recently found out that independently, the degree bound improvement was also obtained by Nachmias and Schapira [3] using a combination of our techniques in the earlier paper and those of Czumaj and Sohler.

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. Thus, the conductance of a cut $(S, \bar{S})$ in the graph is exactly its expansion divided by $2d$. The conductance of the graph, Φ_G, is the minimum conductance of any cut in the graph. We design a property tester for expansion.

The tester is given two parameters Φ and ε. The tester must (with high probability) 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 [2]. Now we present our main result:

Theorem 1.1 *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 a graph of conductance at least $\Omega(\Phi^2)$ with degree bound d.*

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 1.1:

Corollary 1.1 *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 a graph of expansion at least α' with degree bound d. For edge expansion, $\alpha' = \Omega(\frac{\alpha^2}{d})$, and for vertex expansion, $\alpha' = \Omega(\frac{\alpha^2}{d^2})$.*

2 Description of the Property Tester

We first define a procedure called Vertex Tester which will be used by the actual tester.

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

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 actual property 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

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 a random walk of length l starting from v is denoted by $\gamma_l(v) = \sum_v (p^l_{v,u})^2$. 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 paramater $\Delta(v)$ is a measure of how well a random walk from s mixes. The parameter $\gamma(v)$ is something that can be estimated in sublinear time, and by the relationship, allows us to test mixing of random walks in sublinear time. The following is basically proven in [2]:

Lemma 3.1 *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: For every $i < j \leq m$, define a 0/1 random variable X_{ij} which is 1 iff the ith and jth walks share the same endpoint. Let $A = \sum_{i,j} X_{ij}$, the total number of pairwise collisions. Note that $\mathbf{E}[X_{ij}] = \gamma(v)$ and $\mathbf{E}[A] = \gamma(v)\binom{m}{2}$. We now bound the variance $var(A)$.

Note that X_{ij} and X_{kl} are independent when $\{i, j\}$ and $\{k, l\}$ are disjoint. For clarity, we will denote $\binom{m}{2}$ by M. Set $\overline{X}_{ij} = X_{ij} - \gamma(v)$.

$$\begin{aligned}
var(A) &= \mathbf{E}[(A - \gamma(v)M)^2] \\
&= \mathbf{E}[(\sum_{i,j} \overline{X}_{ij})^2] \\
&\leq \sum_{i,j} \mathbf{E}[\overline{X}_{ij}^2] + \sum_{\substack{(i,j),(i',j') \\ i \neq i', j \neq j'}} \mathbf{E}[\overline{X}_{ij}\overline{X}_{i'j'}] + 6 \sum_{i<j<k} \mathbf{E}[\overline{X}_{ij}\overline{X}_{ik}] \\
&\leq \gamma(v)M + 0 + 6\binom{m}{3} \sum_u p^3_{v,u}
\end{aligned}$$

Since X_{ij} and $X_{i'j'}$ are independent, $\mathbf{E}[\overline{X}_{ij}\overline{X}_{i'j'}] = \mathbf{E}[\overline{X}_{ij}]\mathbf{E}[\overline{X}_{i'j'}] = 0$. The product $\overline{X}_{ij}\overline{X}_{ik}$ is 1 iff the ith, jth, and kth walks end at the same vertex, and the probability of that is $\sum_u p_{v,u}^3$. Using the Cauchy-Schwartz inequality, we can show that $\sum_u p_{v,u}^3 \le \gamma(v)^{3/2}$. We know that $\gamma(v) \ge 1/n$ and M is chosen to be larger than n. Therefore $\gamma(v)M \ge 1$ and

$$var(A) \le \gamma(v)M + 4(\gamma(v)M)^{3/2} \le 5(\gamma(v)M)^{3/2}$$

By Chebyschev's inequality, for any $k > 0$,

$$Pr[|A - \gamma(v)M| > k(\gamma(v)M)^{3/4}] < 1/k^2$$

Since $\gamma(v) \ge 1/n$ and $M \ge \Omega(n^{1+\mu})$,

$$(\gamma(v)M)^{3/4} \ \le \ n^{-\mu/4}\gamma(v)M$$

since $m = 2n^{(1+\mu)/2}$. Choosing $k = 2$, the bound above shows that the estimate provided by the Vertex Tester, viz. A/M, lies outside the range $[(1-2n^{-\mu/4})\gamma(v), (1+2n^{-\mu/4})\gamma(v)]$ with probability $< 1/4$. □

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

Corollary 3.2 *The following holds with probability at least* 1/3. *Let $v \in S$, the random sample 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 3.1. 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 $\vec{1}$. We will also need the matrix $L = I - M$, which is the 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 $\Phi_G > \Phi$, 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., [4]) that $\lambda_G \le 1 - \Phi_G^2/2 < 1 - \Phi^2/2$. 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,

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

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

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 high probability, 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 $\Omega(\Phi^2)$. We need a few useful lemmas first.

Lemma 3.4 *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 $\vec{p}$ 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 $\vec{f_1}$ is the $\vec{1}/\sqrt{n}$. We can represent $\vec{u} = \sum_i \alpha_i \vec{f_i}$. Here, $\alpha_1 = 0$, 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:

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

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

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

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 L the index set of the light ones. 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

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

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:

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

We note that $\vec{u}M^l = \sum_i \alpha_i \lambda^l \vec{f_i}$. Thus,

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

Now, 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 $\Delta_l(v) > (2\sqrt{s})^{-1}(1-4\delta)^l$. □

Lemma 3.5 *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 $\vec{u}_S$ (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. Note that $\alpha_1 = \beta_1 = 0$ since $\vec{u}_S$ and $\vec{u}_T$ are orthogonal to $\vec{1}$.

Since the conductance of S is less than δ, by applying inequality (1) from Lemma 3.4, 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 3.4, we get that $\|\vec{u}_T M^l\| > \frac{(1-2\sqrt{2\theta})}{2\sqrt{s}}(1-\delta)^l$, and thus, by Jensen's inequality, there is a $v \in T$ such that $\Delta_l(v) > \frac{(1-2\sqrt{2\theta})}{2\sqrt{s}}(1-4\delta)^l$. □

Lemma 3.6 *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: the second one is obvious from the construction, and as for the first one, 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, lemma 3.5 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. □

Armed with this partitioning algorithm, we are ready to present the patch-up algorithm, which changes the graph in εnd edges and raises its conductance to $\Omega(\Phi^2)$:

PATCH-UP ALGORITHM

1. Partition the graph into two pieces A and $\bar{A}$ with the properties given in Lemma 3.6.
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 $\{u, w\}$. Otherwise, re-sample the vertex v from $\bar{A}$, and repeat.

To implement the second step, 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 3.7 *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. Thus, 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 $\bar{A}$. Thus, 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 3.1 *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 $\bar{A}$ (we disregard the chosen paired edges for now). By Lemma 3.7, 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 the 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 $\Phi \leq 1$.

$\square$

References

[1] Czumaj, S., Sohler, C. *Testing Expansion in Bounded Degree Graphs*, Proc. 48th FOCS, 2007, 570–578.

[2] Goldreich, O., Ron, D. *On Testing Expansion in Bounded-Degree Graphs*, ECCC, TR00-020, 2000.

[3] Nachmias, A., Shapira, A. Testing the Expansion of a Graph ECCC, TR07-118, 2007.

[4] Sinclair, A. *Algorithms for random generation and counting: a Markov chain approach*, Birkhaser Progress In Theoretical Computer Science Series, 1993.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de/