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

Testing Expansion in Bounded Degree Graphs

Satyen Kale C. Seshadhri
Dept. of Computer Science, Princeton University
35 Olden St, Princeton, NJ 08540
{satyen, csesha}@cs.princeton.edu

Abstract

We consider the problem of testing graph expansion in the bounded degree model. 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 degree bound $2d$) with expansion $\Omega(\alpha^2)$. The algorithm runs in time $\tilde{O}(\frac{n^{0.5+\mu}d^2}{\varepsilon\alpha^2})$ 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 expansion of the cut is $\frac{E(S,\bar{S})}{\min\{|S|,|\bar{S}|\}}$. The expansion of the graph, α_G, is the expansion of the minimum expansion cut in the graph.

We can estimate graph expansion using random walks. 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 conductance of the minimum conductance cut in the graph. Thus, $\Phi_G = \alpha_G/2d$.

The graph is represented by an adjacency list, so we have constant time access to the neighbors of any vertex. The tester is given two parameters Φ and ε. Each of these will be assumed to be at most some sufficiently small constant. 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), even when the degree is allowed to be $2d$. This means that G has to be changed at εdn edges (either removing or adding) to make the conductance $> c\Phi^2$ - keeping the degrees to be at most $2d$. Note that this formulation in terms of the conductance is equivalent to the property testing problem given in the abstract, since the conductance and expansion in the graph are related by scaling. We can always assume that $\varepsilon \leq 1/4d$. If not, then we can simply set ε to be $1/4d$ in our tester and this will not affect the results.

This formulation was first considered by Goldreich and Ron [2], who described an approach towards designing the required property tester. They proposed an algorithm, but analysis relied on an unproven combinatorial conjecture. Note that we solve a weaker version of their problem, because we only reject graphs that are far from being expanders with degree bound $2d$ (instead of d). Our algorithm uses the same ideas as their paper, but we use different algebraic techniques to

ISSN 1433-8092

prove different combinatorial results which suffice to complete the analysis. Now we present our main result:

Theorem 1.1 *For any constant $\mu > 0$, there is an algorithm which runs in time $O(\frac{n^{1/2+\mu}\log(n)\log(1/\varepsilon)}{\varepsilon\Phi^2})$ 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 $2d$.*

2 Description of the Property Tester

We now present the 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: $\ell = 3\ln n/\Phi^2$ and $m = \Omega(n^{(1+\mu)/2})$.

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.

Expansion 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 Expansion Tester aborts and outputs **Reject**.
3. Output **Accept**.

3 Proof of Theorem 1.1

Let us fix some notation. The probability of reaching v by performing a random walk of length l from s is $p^l_{s,v}$. The collision probability from s is denoted by $\gamma_l(s) = \sum_v (p^l_{s,v})^2$. The norm of the discrepancy from the stationary distribution will be denoted by $\Delta_t(s)$:

$$\Delta_l(s)^2 = \sum_{v\in V}(p^l_{s,v} - 1/n)^2 = \sum_{v\in V}(p^l_{s,v})^2 - 1/n = \gamma_l(s) - 1/n$$

Since l will usually be equal to ℓ, in that case we drop the subscripts. The relationship between $\Delta(s)$ and $\gamma(s)$ is central to the functioning of the tester. The paramater $\Delta(s)$ is a measure of how well a random walk from s mixes. The parameter $\gamma(s)$ 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 $\gamma(s)$, viz. $A/\binom{m}{2}$, provided by the* VERTEX TESTER *lies outside the range $[(1-2n^{-\mu})\gamma(s), (1+2n^{-\mu})\gamma(s)]$ 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 $\mathbf{E}[X_{ij}] = \gamma$ and $\mathbf{E}[A] = \gamma(s)\binom{m}{2}$. 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 $\overline{X}_{ij} = X_{ij} - \gamma(s)$.

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

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 $\sum_v p_{s,v}^3$. Using Cauchy-Schwartz, we can show that $\sum_v p_{s,v}^3 \leq \gamma(s)^{3/2}$. Thus,

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

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

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

This proves that the estimate provided by the VERTEX TESTER, viz. A/M, lies outside the range $[(1-2n^{-\mu})\gamma(s), (1+2n^{-\mu})\gamma(s)]$ with probability $< 1/3$. □

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

Corollary 3.2 *The following holds with probability of error $< \varepsilon/3$. Let $s \in S$, the random sample chosen by the* EXPANSION TESTER. *If $\gamma(s) < (1+\sigma)/n$, then the majority of the N trials of* VERTEX TESTER *run on v return* **Accept**. *If $\gamma(s) > (1+6\sigma)/n$, 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.

Claim 3.3 *If $\Phi_G > \Phi$, then the* EXPANSION TESTER *accepts with probability at least $2/3$.*

Proof: Let λ_G be the second largest eigenvalue of the transition matrix of the random walk on G. It is well known (see, e.g., [3]) that $\lambda_G \le 1 - \Phi_G^2/2 < 1 - \Phi^2/2$. By the standard rapid mixing analysis, for all $s, v \in V$:

$$|p_{s,v} - 1/n| \le \lambda_G^\ell < (1 - \Phi^2/2)^{3\Phi^{-2}\ln n} \le 1/n^{1.5}$$

As a result, $\Delta(s)^2 < 1/n^2$, and $\gamma(s) < (1+\sigma)/n$ for all $s \in V$. By Corollary 3.2, the tester accepts with probability $> 2/3$. $\square$

We now show that if G is ε-far from having conductance $\Omega(\Phi^2)$, then the tester rejects with high probability. Call a vertex s *weak* if $\gamma(s) > (1+6\sigma)/n$, all others will be called *strong*. Suppose there are more than εn weak vertices. Then with high probability, the random sample S chosen by the Expansion Tester has a weak vertex, since the sample has $\Omega(\varepsilon^{-1})$ random vertices. Thus, the Expansion Tester will reject with high probability.

Let us therefore assume that there are at most εn weak vertices. Now, we will show that εdn edges can be added to make the conductance $\Omega(\Phi^2)$.

We first start with a useful lemma. Let M denote the transition matrix of the random walk. The matrix $L = I - M$ is the Laplacian (I denotes the identity matrix). The eigenvalues of this matrix are of the form $(1-\lambda)$, where λ is an eigenvalue of M.

Lemma 3.4 *Consider a set $S \subset V$ of size $s < n/2$ such that the cut $(S, \overline{S})$ has conductance less than δ. Then, for any integer $l > 0$, there exists a $v \in S$ such that $\Delta_l(v) > (2\sqrt{s})^{-1}(1-\delta)^l$.*

Proof: Denote the size of S by s ($s < n/2$). 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 $\vec{u}M^t = \vec{p}M^t - \vec{1}/n$. For clarity -

$$u_v = \begin{cases} 1/s - 1/n & v \in S \\ -1/n & v \notin S \end{cases}$$

Let $1 = \lambda_1 \ge \lambda_2 \cdots \ge \lambda_n > 0$ be the eigenvalues of M and $\vec{e}_1, \vec{e}_2, \cdots, \vec{e}_n$ be the corresponding orthogonal normalized eigenvectors. Note that $\vec{e}_1$ is the $\vec{1}/\sqrt{n}$. We can represent $\vec{u} = \sum_i \alpha_i \vec{e}_i$. 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 Laplacian L:

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

On the other hand:

$$\vec{u}^\top L\vec{u} = \sum_{i<j} M_{ij}(u_i - u_j)^2 < \frac{\delta ds}{2} \times \frac{1}{2d} \times \frac{1}{s^2} = \frac{\delta}{4s}$$

Putting the above together:

$$\begin{aligned}
\sum_i \alpha_i^2 \lambda_i &> \left(\frac{1}{s} - \frac{1}{n}\right) - \frac{\delta}{4s} \\
&= \frac{1}{s}\left(1 - \frac{\delta}{4}\right) - \frac{1}{n}
\end{aligned}$$

If $\lambda_i > (1-\delta)$, 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.

$$\sum_{i\in H} \alpha_i^2 \lambda_i + \sum_{i \in L} \alpha_i^2 \lambda_i > \frac{1}{s}\left(1 - \frac{\delta}{4}\right) - \frac{1}{n}$$

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

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

We therefore get:

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

We note that $\vec{u}M^l = \sum_i \alpha_i \lambda^l \vec{e}_i$.

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

For every node $v \in V$, let $\vec{f_v}$ be the distribution that is completely concentrated on v. Then $\vec{u} = \frac{1}{|S|}\sum_{v\in S}(\vec{f_v} - \frac{\vec{1}}{n})$, and hence $\vec{u}M^l = \frac{1}{|S|}\sum_{v\in S}(\vec{f_v}M^l - \frac{\vec{1}}{n})$. Now, $\vec{f_v}M^l - \frac{\vec{1}}{n}$ 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 $\frac{1}{|S|}\sum_{v\in S}\Delta_l(v) \geq \|\vec{u}M^l\| > \frac{1}{2\sqrt{s}}(1-\delta)^l$. Hence, there is some $v \in S$ for which $\Delta_l(v) > (2\sqrt{s})^{-1}(1-\delta)^l$. □

Lemma 3.5 *Consider sets $T \subseteq S \subseteq V$ such that the cut $(S, \bar{S})$ has conductance $< \delta$. Let $|T| = (1-\theta)|S|$. Assume $0 < \theta \leq \frac{1}{8}$. Then for some $v \in T$, $\Delta_t(v) > \frac{(1-2\sqrt{2\theta})^2}{2\sqrt{s}}(1-\delta)^t$.*

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 $\vec{u}_S = \sum_i \alpha_i \vec{e}_i$ and $\vec{u}_T = \sum_i \beta_i \vec{e}_i$ be representation of $\vec{u}_S$ and $\vec{u}_T$ in the basis $\{\vec{e}_1, \ldots, \vec{e}_n\}$, 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 $< \delta$, by (1), 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 before, we get that $\|\vec{u}_T M^t\| > \frac{(1-2\sqrt{2\theta})^2}{2\sqrt{s}}(1-\delta)^t$, and thus, by Jensen's inequality, there is a $v \in T$ such that $\Delta_t(v) > \frac{(1-2\sqrt{2\theta})^2}{2\sqrt{s}}(1-\delta)^t$. □

Lemma 3.6 *If there are less than εn weak vertices, then εdn edges can be added to make the conductance $\Omega(\Phi^2)$, while ensuring that all degrees are at most $2d$.*

Proof: We iterate over the weak vertices, and start adding edges randomly as follows. At any stage, there is a subset $F \subseteq V$ of at most $n/4$ "forbidden" nodes. Initially, F is empty. For each weak vertex v, we repeat the following process d times. We choose a random node u in $V \setminus F$ and add the edge $\{u, v\}$. We then add u to F. Once we have d edges to v, we move on to the next weak vertex. Since we add at most εdn edges this way (note that we can assume this is less than $n/4$), the size of F can never be more than $n/4$. This process gives us a new graph G'. Note that no node in G' has degree more than $2d$. We now show that G' (with non-zero probability) has conductance at least $\Omega(\Phi^2)$.

Consider any subset $S \subset V$ of size m ($m < n/2$). Suppose that at least a 1/10-fraction of the vertices in S are weak. At least $dm/10$ edges with an endpoint in S were randomly chosen. The probability that the random edge has both endpoints in S is at most $\frac{m}{3n/4} \leq 2/3$, since $|V \setminus F| \geq 3n/4$ at any stage, and $m \leq n/2$.

The expected number of these edges that lie completely in S is $\leq dm/15$. By the Chernoff-Hoeffding bounds, the probability that more than $md/12$ randomly chosen edges lie completely in S is less than $n^{-\Omega(md)} \leq 1/3n^{m+1}$, if we assume d is 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 $md/60$ edges cross the cut $(S, \overline{S})$. Therefore, the conductance of this cut is at least $1/120 > \Omega(\Phi^2)$, since $\Phi \leq 1$.

Now, suppose that at most a 1/10-fraction of the vertices in S are weak. For any strong vertex s, $\gamma(s) \leq (1+6\sigma)/n$ and thus $\Delta(s) \leq \sqrt{6\sigma/n}$. Let T be the set of all strong vertices in S. Note that $|T| \geq (9/10)|S|$. Consider the cut $(S,\overline{S})$. By Lemma 3.5, if the cut $(S,\overline{S})$ has conductance δ (before adding the new edges), then for some constant $b = \frac{(1-2\sqrt{1/5})^2}{\sqrt{2}}$, we have

$$\frac{b}{\sqrt{n}}(1-\delta)^{\ell} < \sqrt{\frac{6}{n^{1+\mu}}}$$

This implies $\delta = \Omega(\Phi^2)$. Since adding the new edges only increases the conductance, the conductance of the cut $(S,\bar{S})$ is also $\Omega(\Phi^2)$.

Therefore, $\Phi_{G'} = \Omega(\Phi^2)$. □

References

[1] Goldreich, O., Ron, D. *A Sublinear Bipartite Tester for Bounded Degree Graphs*, Combinatorica, Vol. 19(3), 335–373, 1999.

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

[3] 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/