On Testing Expansion in Bounded-Degree Graphs

Oded Goldreich*
Dept. of Computer Science
Weizmann Institute of Science
Rehovot, ISRAEL
oded@wisdom.weizmann.ac.il

Dana Ron
Dept. of EE – Systems
Tel Aviv University
Ramat Aviv, ISRAEL
danar@eng.tau.ac.il

June 15, 2000

Abstract

We consider testing graph expansion in the bounded-degree graph model (as formulated in [1]). Specifically, we refer to algorithms for testing whether the graph has a second eigenvalue bounded above by a given threshold or is far from any graph with such (or related) property.

We present a natural algorithm aimed towards achieving the above task. The algorithm is given a (normalized) second eigenvalue bound $\lambda < 1$, oracle access to a bounded-degree N-vertex graph, and two additional parameters $\epsilon, \alpha > 0$. The algorithm runs in time $N^{0.5+\alpha}/\text{poly}(\epsilon)$, and accepts any graph having (normalized) second eigenvalue at most λ. We believe that the algorithm rejects any graph that is ϵ-far from having second eigenvalue at most $\lambda^{\alpha/O(1)}$, and prove the validity of this belief under an appealing combinatorial conjecture.

*Supported by MINERVA Foundation, Germany.

1 Brief Introduction

This memo reports partial results regarding the task of testing whether a given bounded-degree graph is an expander. The model is of testing graph properties as formulated in [1]: The (randomized) algorithm is given integers d and N, a distance parameter ϵ (as well as some problem-specific parameters), and oracle access to a N-vertex graph G with degree bound d; that is, query $(v,i) \in [N] \times [d]$ is answered by the $i^{\rm th}$ neighbor of v in G (or by a special symbol in case v has less than i neighbors). For a predetermined property $\mathcal{P}$, the algorithm is required to accept (with probability at least 2/3) any graph having property $\mathcal{P}$, and reject (with probability at least 2/3) any graph that is ϵ-far from having property $\mathcal{P}$, where distance between graphs is defined as the fraction of edges (over dN) on which the graphs differ.

Loosely speaking, the specific property considered here is being an expander. More precisely, for a given bound $\lambda < 1$, we consider the property, denoted $\mathcal{E}_\lambda$, of having a normalized by d adjacency matrix with second eigenvalue at most λ. Actually, we further relax the property testing formulation (as in [4]): Using an additional parameter $\lambda' \geq \lambda$, we only require that

- the algorithm must accept (with probability at least 2/3) any graph having property $\mathcal{E}_\lambda$ (i.e., having second eigenvalue at most λ); and
- the algorithm must reject (with probability at least 2/3) any graph that is ϵ-far from having property $\mathcal{E}_{\lambda'}$ (i.e., from any graph that has second eigenvalue at most λ').

Setting $\lambda' = \lambda$ we regain the more strict formulation of testing whether a graph has second eigenvalue at most λ.

We mention that the $\Omega(\sqrt{N})$ lower bound on "testing expansion" (presented in [1]) continues to hold for the relaxed formulation above, provided that $\lambda' < 1$. This is the case since the lower bound is established by showing that any $o(\sqrt{N})$-query algorithm fails to distinguish between a very good expander and an unconnected graph with several huge connected components.[1]

In view of the above, we shall be content with any sub-linear time algorithm for testing expansion. Below, we present a parameterized family of algorithms. For any $\alpha > 0$, the algorithm has running-time $n^{0.5+\alpha}/\text{poly}(\epsilon)$ and is supposed to satisfy the above requirement with $\lambda' = \lambda^{\alpha/7}$. Unfortunately, we can only prove that this is indeed the case provided that a certain combinatorial conjecture (presented in Section 4.2) holds.

2 Conventions and Notations

We consider N-vertex graphs of degree bound d, which should be thought of as fixed. We consider the stochastic matrix representing a canonical random walk on this graph, where canonical is anything reasonable (e.g., go to each neighbor with probability $1/2d$). The eigenvalues below refer to this matrix.

By λ we denote the claimed second eigenvalue (i.e., we need to accept graphs having second eigenvalue at most λ). By ϵ we denote the distance parameter: we need to reject graphs that are ϵ-far from having second eigenvalue at most λ', where $\lambda' > \lambda$ is related to λ.

The algorithm presented below is parameterized by a small constant $\alpha > 0$ that determines both its complexity (i.e., $O(N^{0.5+\alpha}/\text{poly}(\epsilon))$) and its performance (i.e., $\lambda' = \lambda^{\alpha/O(1)}$). To be of interest, the algorithm must use $\alpha < 0.5$.

[1] In the latter case, the graph has (normalized) second eigenvalue equal 1.

3 The algorithm

We set $L = \frac{1.5 \ln N}{\ln(1/\lambda)}$. This guarantees that a graph with second eigenvalue at most λ mixes well in L steps (i.e., the deviation in max-norm of the end probability from the uniform distribution is at most $N^{-1.5}$). The following algorithm evaluates the distance of the end probability (of an L-step random walk starting at a fixed vertex) from the uniform probability distribution. It is based on the fact that the uniform distribution over a set has the smallest possible collision probability, among all distributions over this set.

> *Repeat* $t \stackrel{\mathrm{def}}{=} \Theta(1/\epsilon)$ *times*
>
> *(1) Select uniformly a start vertex, denoted* s*.*
>
> *(2) Perform* $m \stackrel{\mathrm{def}}{=} \Theta(N^{0.5+\alpha}/\epsilon)$ *random walks of length* L*, starting from vertex* s*.*
>
> *(3) Count the number of pairwise collisions between the endpoints of these* m *walks.*
>
> *(4) If the count is greater than* $\frac{1+0.5 \cdot N^{-\alpha/2}}{N} \cdot \binom{m}{2}$ *then* `reject`
>
> *If all repetitions were completed without rejection then* `accept`.

Comment: Random walks were used before in the context of testing graph properties (in the bounded-degree model): Specifically, $\widetilde{O}(\sqrt{N}/\mathrm{poly}(\epsilon))$ such walks were used by the bipartitness tester of [2]. Random walks seem much more natural here.

4 Analysis

Fixing any start vertex s, we denote by $p_{s,v}$ the probability that a random walk of length L starting at s ends in v. The collision probability of L-walks starting at s is given by

$$\sum_v p_{s,v}^2 \ \geq \ \frac{1}{N} \tag{1}$$

By our choice of L, if the graph has eigenvalue at most λ then (for any starting vertex s) the collision probability of L-walks starting at s is very close to $1/N$ (i.e., is smaller than $(1/N) + (1/N^2)$).

4.1 Approximation of the collision probabilities

The first issue to address is the approximation to Eq. (1) provided by Steps (2)–(3) of the algorithm.

Lemma 1 *With probability at least* $1 - (1/3t)$*, the approximation provided by Steps (2)–(3) is within a factor of* $1 \pm \frac{1}{4} \cdot N^{-\alpha/2}$ *of* Eq. (1).

Thus, with probability at least 2/3, all approximations provided by the algorithms are within a factor of $1 \pm \frac{1}{4} \cdot N^{-\alpha/2}$ of the correct value.

Proof: For every $i < j$, define a 0-1 random variable $\zeta_{i,j}$ so that $\zeta_{i,j} = 1$ if the endpoint of the i^{th} path is equal to the endpoint of the j^{th} path. Clearly, $\mu \stackrel{\mathrm{def}}{=} \mathrm{E}[\zeta_{i,j}] = \sum_v p_{s,v}^2$, for every $i < j$. Using Chebyshev's inequality we bound the probability that the count provided by Steps (2)–(3) deviates from its (correct) expected value. Let $P \stackrel{\mathrm{def}}{=} \{(i,j) : 1 \leq i < j \leq m\}$ and $\delta = \frac{1}{4} \cdot N^{-\alpha/2}$.

$$\Pr\left[\left|\sum_{(i,j)\in P} \zeta_{i,j} - |P| \cdot \mu\right| > \delta \cdot |P| \cdot \mu\right] \ \leq \ \frac{\mathrm{Var}[\sum_{(i,j)\in P} \zeta_{i,j}]}{(\delta \cdot |P| \cdot \mu)^2} \tag{2}$$

Denote $\overline{\zeta}_{i,j} \stackrel{\rm def}{=} \zeta_{i,j} - \mu$. The rest of the proof needs to deal with the fact that the random variables associated with P are NOT pairwise independent. Specifically, for four *distinct* i, j, i', j', indeed $\zeta_{i,j}$ and $\zeta_{i',j'}$ are independent, and so $\mathrm{E}[\overline{\zeta}_{i,j}\overline{\zeta}_{i',j'}] = \mathrm{E}[\overline{\zeta}_{i,j}] \cdot \mathrm{E}[\overline{\zeta}_{i',j'}] = 0$; but for $i < j \neq k$ the random variables $\zeta_{i,j}$ and $\zeta_{i,k}$ are *not* independent (since they both depend on the same $i^{\rm th}$ walk). Still

$$\begin{aligned}
\mathrm{Var}\left[\sum_{(i,j)\in P} \zeta_{i,j}\right] &= \mathrm{E}\left[\left(\sum_{(i,j)\in P} \overline{\zeta}_{i,j}\right)^2\right] \\
&= \sum_{(i,j)\in P} \mathrm{E}\left[\overline{\zeta}_{i,j}^2\right] + 2 \cdot \sum_{(i,j),(i,k)\in P \ \&\ j\neq k} \mathrm{E}\left[\overline{\zeta}_{i,j}\overline{\zeta}_{i,k}\right] \\
&\leq \sum_{(i,j)\in P} \mathrm{E}\left[\zeta_{i,j}^2\right] + 2 \cdot \sum_{i\in[m]} \sum_{j\neq k\in\{i+1,\ldots,m\}} \mathrm{E}\left[\zeta_{i,j}\zeta_{i,k}\right] \\
&\leq |P| \cdot \mu + 2 \cdot \frac{m^3}{6} \cdot \sum_v p_{s,v}^3
\end{aligned}$$

since $\zeta_{i,j}\zeta_{i,k} = 1$ if and only if all three random walks end at the same vertex. Using $(\sum_v p_{s,v}^3)^{1/3} \leq (\sum_v p_{s,v}^2)^{1/2}$, and $m^2 < 3 \cdot |P|$, we obtain

$$\mathrm{Var}\left[\sum_{(i,j)\in P} \zeta_{i,j}\right] \leq |P| \cdot \mu + |P|^{3/2} \cdot \mu^{3/2} < 2 \cdot (|P| \cdot \mu)^{3/2} \tag{3}$$

Combining Eq. (2) and (3), we obtain

$$\Pr\left[\left|\sum_{(i,j)\in P} \zeta_{i,j} - |P| \cdot \mu\right| > \delta \cdot |P| \cdot \mu\right] < \frac{2}{\delta^2 \cdot (|P| \cdot \mu)^{1/2}}$$

Using $\mu \geq 1/N$ and $|P| > \frac{m^2}{4} = \Theta(\frac{N^{1+2\alpha}}{\epsilon^2})$, the denominator is at least $\delta^2 \cdot \Theta(\frac{N^\alpha}{\epsilon})$. Recalling that $\delta = \frac{1}{4} \cdot N^{-\alpha/2}$ and $t = O(1/\epsilon)$, the lemma follows. ■

As an immediate corollary we get

Corollary 2 *If the graph has second eigenvalue at most λ then the above algorithm accepts it with probability at least* 2/3.

Another immediate corollary is the following

Corollary 3 *Suppose that for at least a $\epsilon/O(1)$ fraction of the vertices s in G the collision probability of L-walks starting at s is greater than $\frac{1+0.8N^{-\alpha/2}}{N}$. Then the algorithm rejects with probability at least* 2/3.

Thus, if a graph passes the test (with probability greater than 1/3) then it must have less than $(\epsilon/O(1)) \cdot N$ *exceptional vertices*; that is, vertices s for which the collision probability of L-walks starting at s is greater than $\frac{1+0.8N^{-\alpha/2}}{N}$.

Comment: Note that by changing parameters in the algorithm (i.e., $t = \Theta(N^\alpha/\epsilon)$ and $m = \Theta(N^{0.5+2\alpha}/\epsilon)$), we can make the fraction of exceptional vertices smaller than $\epsilon N^{-\alpha}$. This may help in closing the gap (below), and only increases the complexity from $N^{0.5+\alpha}/\mathrm{poly}(\epsilon)$ to $N^{0.5+3\alpha}/\mathrm{poly}(\epsilon)$.

4.2 The gap

We believe that the following conjecture (or something similar) is true.

Conjecture: *Let G be an N-vertex graph of degree-bound d. Suppose that* for all but at most $\epsilon/O(1)$ fraction *of the vertices s in G the collision probability of L-walks starting at s is at most $\frac{1+0.8N^{-\alpha/2}}{N}$. Then G is ϵ-close to a N-vertex graph* (of degree-bound d) *in which the collision probability of L-walks starting at* any vertex *is at most $\frac{1+N^{-\alpha/2}}{N}$.*

The conjecture is very appealing: Supposedly, you add ϵdN edges connecting at random the exceptional vertices to the rest of the graph. Ignoring for a moment the issue of preserving the degree bounds, this seems to work – but we cannot prove it. Indeed, one can show that the previously exceptional vertices enjoy rapid mixing, but it is not clear that the added edges will not cause harm to the mixing properties of non-exceptional vertices.

4.3 Finishing it off

Once the gap is closed, we have the following situation: If the algorithm rejects with probability smaller than $2/3$ then the input graph is ϵ-close to a graph in which the collision probability of L-walks starting at *any vertex* is at most $\frac{1+N^{-\alpha/2}}{N}$. But the excess of the collision probability beyond $1/N$ is nothing but the square of the distance, in norm 2, of the probability vector $(p_{s,v})_{v\in[N]}$ from the uniform probability vector (i.e., $(\sum_v p_{s,v}^2) - (1/N) = \sum_v (p_{s,v} - (1/N))^2$). Thus, for every s the distance, in norm 2, of the probability vector $(p_{s,v})_{v\in[N]}$ from the uniform probability vector is at most $\sqrt{\frac{N^{-\alpha/2}}{N}} = N^{-(0.5+\beta)}$, where $\beta = \alpha/4$.

The plan now is to "reverse" the standard eigenvalue to rapid-mixing connection. That is, infer from the rapid-mixing feature that the graph has a small second eigenvalue. Such a lemma has appeared in [3]:

Lemma 4 (Lemma 4.6 in [3]): *Consider a regular connected graph on N vertices, let A be its normalized adjacency matrix and λ_2 denote the absolute value of the second eigenvalue of A. Let ℓ be an integer and Δ_ℓ denote an upper bound on the maximum, taken over all possible start vertices s, of the difference in Norm2 between the distribution induced by an ℓ-step random walk starting at s and the uniform distribution. Then $\lambda_2 \leq (N \cdot \Delta_\ell)^{1/\ell}$.*

Note that by the above, we have $\Delta_L < N^{-(0.5+\beta)}$. This does not give anything useful when applying the lemma directly. Instead, we apply the lemma after bounding Δ_ℓ for $\ell = O(L)$. (The following may be an oversight, but that's how we argue it now.)

Claim 5 *Let Δ_ℓ be define as in Lemma 4. Then $\Delta_{k\ell} \leq (\sqrt{N} \cdot \Delta_\ell)^k$, for every integer k.*

Proof: Let $B = A^\ell$ be the stochastic matrix representing an ℓ-step random walk, and let $\vec{e}_1, ..., \vec{e}_N$ denote probability vectors in which all the mass is on one vertex. Let $\vec{\nu}$ denote the uniform probability vector. Then Δ_ℓ (resp., $\Delta_{k\ell}$) equals the maximum of $\|B\vec{e}_i - \vec{\nu}\|$ (resp., $\|B^k\vec{e}_i - \vec{\nu}\|$) taken over all the $\vec{e}_i$'s.

Considering the basis of $\vec{e}_i$'s, let $\vec{z}$ be a zero-sum vector (such as $\vec{e}_i - \vec{\nu}$). That is, $\vec{z}$ is written in the basis of $\vec{e}_i$'s as $\vec{z} = \sum_i z_i\vec{e}_i$, and $\sum_i z_i = 0$. We obtain

$$
\begin{aligned}
\|B\vec{z}\| &= \left\| B\left(\sum_i z_i \vec{e}_i\right) - \sum_i z_i B\vec{\nu} \right\| \\
&= \left\| \sum_i z_i B(\vec{e}_i - \vec{\nu}) \right\| \\
&\leq \sum_i \|z_i B(\vec{e}_i - \vec{\nu})\| \\
&= \sum_i |z_i| \cdot \|B(\vec{e}_i - \vec{\nu})\| \\
&\leq \left(\sum_i |z_i|\right) \cdot \Delta_\ell
\end{aligned}
$$

Since $\sum_i |z_i| \leq \sqrt{N} \cdot \sqrt{\sum_i z_i^2} = \sqrt{N} \cdot \|\vec{z}\|$, we get

$$
\|B\vec{z}\| \leq \sqrt{N} \cdot \Delta_\ell \cdot \|\vec{z}\|
$$

Using $B\vec{\nu} = \vec{\nu}$, we get for every i

$$
\begin{aligned}
\|B^k \vec{e}_i - \vec{\nu}\| &= \|B(B^{k-1}\vec{e}_i - \vec{\nu})\| \\
&\leq \Delta_\ell \cdot \sqrt{N} \cdot \|B^{k-1}\vec{e}_i - \vec{\nu}\| \\
&< \left(\Delta_\ell \cdot \sqrt{N}\right)^k
\end{aligned}
$$

and the claim follows. ■

Combining Lemma 4 and Claim 5, we obtain the following

Corollary 6 *Suppose that for every s the distance, in norm 2, of the probability vector $(p_{s,v})_{v\in[N]}$ from the uniform probability vector is at most $N^{-(0.5+\beta)}$. Then, for every constant $\gamma < 2\beta/3$, the second eigenvalue of the graph is at most λ^γ.*

So once the gap is filled, we are done (using $\beta = \alpha/4$ and $\gamma \approx 2\beta/3$).

Proof: Let λ' be the second eigenvalue of the graph. Then, for every k we have

$$
\begin{aligned}
\lambda' &\leq (N \cdot \Delta_{kL})^{1/kL} \\
&\leq \left(N \cdot \left(\sqrt{N} \cdot \Delta_L\right)^k\right)^{1/kL} \\
&\leq \left(N \cdot \left(N^{-\beta}\right)^k\right)^{1/kL} \\
&= \exp\left(\frac{(1-k\beta)\cdot \ln N}{kL}\right)
\end{aligned}
$$

Substituting for $L = \frac{1.5 \ln N}{\ln(1/\lambda)}$, we get

$$
\begin{aligned}
\frac{(1-k\beta)\cdot \ln N}{kL} &= \frac{(1-k\beta)\cdot \ln N}{k \cdot ((1.5 \ln N)/\ln(1/\lambda))} \\
&= \left(\frac{2\beta}{3} - \frac{2}{3k}\right) \cdot \ln \lambda \\
&> \gamma \cdot \ln \lambda
\end{aligned}
$$

for sufficiently large k (since $\gamma < 2\beta/3$). We get $\ln \lambda' > \gamma \cdot \ln \lambda$, and the corollary follows. ■

Comment: We have $\lambda' \leq \lambda^\gamma$ for any $\gamma < \alpha/6$ (e.g., $\gamma = \alpha/7$ will do). One may be able to increase the exponent (i.e., γ) somewhat, but a linear dependency (of the exponent γ) on α seems unavoidable (under the current approach).

References

[1] O. Goldreich and D. Ron. Property Testing in Bounded Degree Graphs. In *Proc. of the 29th ACM Symp. on Theory of Computing*, pages 406–415, 1997.

[2] O. Goldreich and D. Ron. A sublinear Bipartite Tester for Bounded Degree Graphs. *Combinatorica*, Vol. 19 (2), pages 1–39, 1999.

[3] O. Goldreich and M. Sudan. Computational Indistinguishability: A Sample Hierarchy. *JCSS*, Vol. 59, pages 253–269, 1999.

[4] M. Parnas and D. Ron. Testing the diameter of graphs. In *Proceedings of Random99*, 1999.