Testing the Expansion of a Graph

Asaf Nachmias* Asaf Shapira †

Abstract

We study the problem of testing the expansion of graphs with bounded degree d in sublinear time. A graph is said to be an α-expander if every vertex set $U \subset V$ of size at most $\frac{1}{2}|V|$ has a neighborhood of size at least $\alpha|U|$.

We show that the algorithm proposed by Goldreich and Ron [9] (ECCC-2000) for testing the expansion of a graph distinguishes with high probability between α-expanders of degree bound d and graphs which are ϵ-far from having expansion at least $\Omega(\alpha^2)$. This improves a recent result of Czumaj and Sohler [3] (FOCS-07) who showed that this algorithm can distinguish between α-expanders of degree bound d and graphs which are ϵ-far from having expansion at least $\Omega(\alpha^2/\log n)$. It also improves a recent result of Kale and Seshadhri [11] (ECCC-2007) who showed that this algorithm can distinguish between α-expanders and graphs which are ϵ-far from having expansion at least $\Omega(\alpha^2)$ with twice the maximum degree. Finally, our result shows that the conjecture of Goldreich and Ron [9], on testing the second eigenvalue of the graph, holds when the second eigenvalue lies in a certain interval of constant size. Our methods combine the techniques of [3], [9] and [11].

*Department of Mathematics, U.C. Berkeley. Research supported in part by NSF grant #DMS-0605166.

†Microsoft Research

1 Introduction

1.1 Background on property testing

We consider testing properties of graphs in the *bounded degree* model, which was introduced by Goldreich and Ron [9]. In this model we fix a degree bound d and represent graphs using adjacency lists. More precisely, we assume that a graph G is represented as a function $f_G : [n]\times[d] \mapsto \{[n]\cup\{*\}\}$, where given a vertex $v \in V(G)$ and $1 \leq i \leq d$ the function $f(v,i)$ returns the i^{th} neighbor of v, in case v has at least i vertices. If v has less than i vertices then $f(v,i) = *$. A graph of bounded degree d is said to be ϵ-far from satisfying $\mathcal{P}$ if one needs to add and/or delete at least ϵdn edges to G in order to turn it into a graph satisfying $\mathcal{P}$. Observe that if we think of d as a fixed constant, which is independent of n, then being ϵ-far actually means that an ϵ-fraction of the edges should be modified in order to get a graph satisfying the property (assuming the graphs has $\Omega(n)$ edges).

A testing algorithm (or tester) for graph property $\mathcal{P}$ is a (possibly randomized) algorithm that distinguishes with probability at least 2/3 between graphs satisfying $\mathcal{P}$ from graphs that are ϵ-far from satisfying it. More precisely, if the input graph satisfies $\mathcal{P}$ the algorithm accepts it with probability at least 2/3, where the probability is taken over the coin tosses of the tester. Similarly if it is ϵ-far from satisfying $\mathcal{P}$ the algorithm should reject it with probability at least 2/3. The tester only has access to the function f_G, and its *query complexity* when executed on G is the number of f_G-calls that it performs. We say that a tester for property $\mathcal{P}$ has query complexity $q_{\mathcal{P}}(n,\epsilon)$ if for any $\epsilon > 0$ and any input graph of G on n vertices, the tester makes at most $q_{\mathcal{P}}(n,\epsilon)$ queries to f_G.

For more details on property testing and on graph property testing, see the surveys [4, 6, 12, 13]

1.2 Previous results on testing expansion

Our main result in this paper is related to testing the expansion of a graph. We start by introducing the standard notation and definitions related to expanders. See [10] for more details. For a set of vertices $U \subseteq V(G)$ in a graph $G = (V,E)$ we denote by $N(U)$ the set of vertices in $V(G)\setminus U$ that are connected with at least one vertex of U. For two disjoint vertex set A and B we denote by $E(A,B)$ the number of edges connecting a vertex of A with a vertex of B. We say that a graph is an *α-vertex-expander* or just *α-expander* if for every $U \subseteq V(G)$ satisfying $|U| \leq \frac{1}{2}|V(G)|$ we have $N(U) \geq \alpha|U|$.

Let us also introduce two other notions of expansion that are frequently used in the literature. A graph as above is said to be an α*-edge-expander* if for every $U \subseteq V(G)$ satisfying $|U| \leq \frac{1}{2}|V(G)|$ we have $E(U, V \setminus U) \geq \alpha|U|$. Let us associate with a graph $G = (V, E)$ the standard adjacency matrix $A = A(G)$, and denote by $\lambda(G)$ the second largest eigenvalue of $\frac{1}{d}A$ (we normalize by $1/d$ so that all eigenvalues are in $[-1, 1]$). We say that G is an (n, d, λ)*-expander* if G has bounded degree d and $\lambda(G) \leq \lambda$.

Goldreich and Ron [9, 7] were the first to consider the problem of testing the expansion of a graph. More precisely, they considered the problem of distinguishing between an input that is an (n, d, λ)-expander and an input that is ϵ-far from being an (n, d, λ')-expander for some $\lambda' > \lambda$ [1]. It was already observed in [7] that this problem cannot be tested with $o(n^{0.5})$ queries. In [9], Goldreich and Ron suggested an algorithm, which is described in detail in the next section, that performs a sequence of random walks on the input and counts the number of pairwise collisions of these walks. This algorithm is parameterized by a real $\eta > 0$, and the conjecture of Goldreich and Ron (GR-conjecture) was that in time[2] $\tilde{O}(n^{0.5+\eta}\mathrm{poly}(1/\epsilon))$ the algorithm can distinguish between (n, d, λ)-expanders and graphs that are ϵ-far from being an $(n, d, \lambda^{\Theta(\eta)})$-expander. Note that as the running time of the conjectured algorithm is $\tilde{O}(n^{0.5+\eta}\mathrm{poly}(1/\epsilon))$ then so is its query complexity.

The GR-conjecture was recently addressed in [3, 11]. Czumaj and Solher [3] showed that the algorithm purposed in [9] successfully distinguishes between α-expanders and graphs that are ϵ-far from being an $O(\alpha^2/\log n)$-expanders. Note that this result is weaker than the GR-conjecture, where the algorithm is supposed to be able to distinguish between two *constant* expansions, while the analysis of [3] shows that the algorithm only rejects graphs that are far from having a *sub-constant* expansion.

In another recent paper, Kale and Seshadhri [11] have shown that the algorithm proposed in [9] successfully distinguishes between α-edge-expanders with bounded degree d and graphs that are ϵ-far even from being a $\Omega(\alpha^2)$-edge-expanders with bounded degree $2d$. While the algorithm of [11] considers constant expansion, it considers graphs that are far from being expanders even when the degree can be twice as large.

Our main result in this paper deals with vertex and edge expansion as in [3, 11]. We simultaneously improve the results of [3] and [11] in that we

---

[1] The reader may have noticed that this is a relaxed version of the usual notion of testing a property as we have defined in the previous subsection, because we are only asked to reject a graph that is far from satisfying a property that is weaker than the property which should make the algorithm accept.

[2] The $\tilde{O}$ notation hides $\log^{O(1)} n$ factors.

consider constant expansion (unlike the sub-constat expansion considered in [3]) and we consider graphs that are far from being expanders of the same degree (unlike [11] that consider graphs with a larger degree). Although our analysis does not fully resolve the GR-conjecture, as we cannot address the entire range of $\lambda(G)$, we can verify their conjecture when $\lambda(G)$ lies in a certain interval. See Section 2.

The main idea of our proof is to combine the central combinatorial argument of Czumaj and Sohler [3], with a spectral lemma from the analysis of Kale and Seshadhri [11] as well as a second moment estimation from Goldreich and Ron [9].

1.3 Organization

The rest of the paper is organized as follows. In Section 2 we describe the algorithm suggested in [9] for testing the expansion of a graph. In this section we also state our main result that improves those presented in [3] and [11] and partially resolves the GR-conjecture. The proof of the main result appears in Section 3 and in section 4 we discuss some concluding remarks and open problems.

2 The Goldreich-Ron Algorithm and Statement of Main Result

We shall use the algorithm `ExpansionTester` suggested in [9]. In this algorithm we use a modified version of the lazy simple random walk on G. In this walk the probability of taking any outgoing edge is $\frac{1}{2d}$ and with the remaining probability the random walk stays at the same vertex. The algorithm receives 4 parameters (t, m, M, N) and operates as follows. It repeats N times the following procedure: pick a random uniform vertex $v \in G$ and perform m random walks of length t from v. Let X count the number of pairwise collisions of the endpoints of the m walks. The algorithm accepts if in all N trials $X \leq M$ and rejects if in one of the trials we have $X > M$.

Theorem 2.1. **[Main Result]** *For any $\alpha \in (0,1)$, integer $d \geq 3$ and $\mu \in (0,1/4)$ we set*

$$t = \frac{16d^2}{\alpha^2}\log n\,, \quad m = n^{1/2+\mu}\,, \quad M = \frac{n^{2\mu}}{2} + \frac{n^{7\mu/4}}{128}\,, \quad N = \frac{300}{\epsilon}\,.$$

Then there is a constant $c = c(d) > 0$ such that for any $\epsilon \in (0,1)$ we have

1. *If G is an α-vertex-expander, then with probability at least $2/3$ the algorithm* `ExpansionTester` *accepts.*

2. *If G is ϵ-far from a $c\mu\alpha^2$-vertex-expander of degree bound d then with probability at least $2/3$ the algorithm* `ExpansionTester` *rejects.*

As we have mentioned before, the Goldreich-Ron algorithm when applied to the vertex expansion problem was analyzed in [3], where it was shown that it can distinguish between α-expanders and graphs that are far from $\alpha^2/\log n$-expanders. Theorem 2.1 thus improves the result of [3] by completely removing the dependence on n in the definition of the minimal expansion that should be rejected.

As every α-expander is also an α-edge-expander and every α-edge-expander of bounded degree d is also an α/d-expander we immediately get the following

Corollary 2.2. *Let $\beta \in (0,1)$ and consider the algorithm of Theorem 2.1 with $\alpha = \beta/d$. Then this algorithm distinguishes with high probability between β-edge-expanders and graphs that are ϵ-far from being a $c\mu\beta^2/d^2$-edge-expander.*

As we have mentioned before, the Goldreich-Ron algorithm when applied to the edge-expansion problem was analyzed in [11], where it was shown that it can distinguish between β-expanders and graphs that are far from being $\Omega(\beta^2)$-expanders with twice the maximum degree. Theorem 2.1 thus improves the result of [11] by considering the case when the graph is far from being an expander with the same maximum degree.

Recall that the *spectral gap* g of a Markov chain p is defined as $1 - \lambda_2$ where λ_2 is the second largest eigenvalue of p. By the classical relations between the expansion of a graph and the spectral gap of the random walk on it (see Theorem 3.5) we can derive the following.

Corollary 2.3. *For any $g \in (0,1)$, the algorithm* `ExpansionTester` *with the parameters defined in Theorem 2.1 distinguishes with high probability between graphs of bounded degree d with spectral gap at least g and graphs which are ϵ-far from having spectral gap at least $\Omega(\mu g^4)$ (note that we are referring to the spectral gap of the modified random walk).*

Remark. By following the proof of Theorem 2.1 one can improve the $\Omega(\mu g^4)$ appearing in Corollary 2.3 to $\Omega(\mu g^2)$. We omit the details.

The GR-conjecture states that the algorithm described above distinguishes in time $O(n^{0.5+\eta})$ between graphs with second eigenvalue at most λ and graphs which are ϵ-far from having second eigenvalue at most $\lambda^{a\mu}$ where $a > 0$ is a small universal constant (note that here, as in [9], we refer to the second largest eigenvalue of the transition matrix of the modified random walk). Our analysis of testing expansion shows that this conjecture holds when λ is contained within a certain (non-empty) interval of $[0, 1]$.

Indeed, we can show that for every $\zeta > 0$ there are constants $a > 0$ and $\zeta' > \zeta$ such that the GR-conjecture holds when $\lambda \in [1-\zeta', 1-\zeta]$. To see this, recall that Corollary 2.3 shows that the algorithm of Theorem 2.1 can distinguish in time $O(n^{0.5+\eta})$ between graphs with spectral gap ζ from graphs that are ϵ-far from having spectral gap $c\mu\zeta^4$, where c depends only on d. In other words, it can distinguish between graphs with $\lambda(G) \leq 1-\zeta$ from graphs that are ϵ-far from having $\lambda(G) \leq 1 - c\mu\zeta^4$. Now, clearly, if $a > 0$ is small enough (in terms of ζ and c) then $(1 - c\mu\zeta^4)^{\frac{1}{a\mu}} < (1 - \zeta)$. Therefore, for some $\zeta' > \zeta$, we have $(1 - c\mu x^4)^{\frac{1}{a\mu}} \leq (1 - x)$ for every $x \in [1 - \zeta', 1 - \zeta]$. Therefore, when $\lambda \in [1-\zeta', 1-\zeta]$ the algorithm distinguishes between graphs with second eigenvalue at most λ and graph that are far from having second eigenvalue at most $\lambda^{a\mu}$.

3 Proof of Theorem 2.1

Let us recall some definitions. Let p be a reversible aperiodic Markov chain on a finite state space V with stationary distribution π. The *Cheeger* constant of the chain Φ_* is defined as

$$\Phi_* = \min_{S \subset V \,:\, \pi(S) \leq 1/2} \frac{\sum_{x \in S, y \in S^c} \pi(x) p(x,y)}{\pi(S)}$$

where $\pi(S) = \sum_{v \in S} \pi(v)$. Our modified random walk has $p(x, y) = \frac{1}{2d}$ if (x, y) is an edge of G and $p(x, x) = 1 - \frac{\deg(x)}{2d} \geq 1/2$. The stationary distribution of this chain is the uniform distribution. It is immediate from the definitions that if G is an α-expander, then the Cheeger constant of the modified random walk satisfies $\Phi_* \geq \frac{\alpha}{2d}$.

For the proof of Theorem 2.1 we will need to first recall some lemmas from [3, 9, 11]. First, we will need the following combinatorial result of [3].

Lemma 3.1. [Corollary 4.6 of [3]] *Let $G = (V, E)$ be a graph with $|V| = n$ and bounded degree d. There exists a constant $C = C(d) > 0$ such that the following holds. If G is ϵ-far from a β-expander with $\beta \leq \frac{1}{3}$, then there* *is a subset of vertices $A \subset V$ with $\epsilon n/4 \le |A| \le (1+\epsilon)n/2$ such that $|N(A)| \le C\beta|A|$.*

Next, let $p(x,y)$ be the transition matrix of our modified random walk on G. We denote the distance in ℓ_2 of $p^t(x,\cdot)$ from the stationary distribution by $\Delta_t(x)$, i.e.,

$$\Delta_t^2(x) = \sum_{y \in V} \left[p^t(x,y) - \frac{1}{n}\right]^2 .$$

The following lemma is obtained immediately by putting $\Theta = \frac{1}{32}$ in Lemma 3.5 of [11].

Lemma 3.2. [Lemma 3.5 of [11]] *Consider a set $A \subset V$ of size $|A| < n/2$. For any $\delta \in (0,1)$ we have that if $E(A, V \setminus A) < 2\delta d|A|$ then at least $\frac{1}{32}|A|$ vertices x of A satisfy*

$$\Delta_t^2(x) \ge \frac{(1-\delta)^{2t}}{64|A|} ,$$

for any integer $t > 0$.

Following the notation of [11], for a vertex $x \in V$ write

$$\gamma_t(x) = \sum_{y \in V} [p^t(x,y)]^2 ,$$

and observe that we have the following relation between $\Delta_t^2(x)$ and $\gamma_t(x)$

$$\Delta_t^2(x) = \gamma_t(x) - \frac{1}{n} . \tag{3.1}$$

The previous two lemmas yield the following statement.

Lemma 3.3. *Let $\alpha \le \frac{1}{3}$ and put $t = \frac{16d^2}{\alpha^2}\log n$. Then there exists some $\beta_0 = \beta_0(d) > 0$ such that the for any $\beta \le \beta_0$ the following statement holds: if G is ϵ-far from being a β-expander, then there are at least $\epsilon n/128$ vertices x of G for which*

$$\Delta_t^2(x) \ge \frac{n^{-\frac{32Cd^2\beta}{\alpha^2}}}{32n} .$$

Proof. Lemma 3.1 shows that there exists a subset $A \subset V$ with $\epsilon n/4 \le |A| \le n/2$ satisfying $E[A, A^c] \le Cd\beta|A|$ (according to the lemma it is possible that $n/2 < |A| \le (1+\epsilon)n/2$ and in that case we just take A^c). Lemma

3.2 then implies that as long as $C\beta < 2$ we have for at least $\epsilon n/128$ vertices x of G that

$$\Delta_t^2(x) \ge \frac{(1 - C\beta/2)^{2t}}{32n} \ge \frac{n^{-\frac{32Cd^2\beta}{\alpha^2}}}{32n} .$$

$\square$

For the proof of Theorem 2.1 we will also need the following second-moment statement

Lemma 3.4. [Lemma 1 of [9] and Lemma 3.1 of [11]] *For an integer $m > 0$ consider m independent modified random walks of length t starting from a vertex x. Let X count the number of pairwise collisions of these walks. We have that*

$$\mathbf{E}X = \binom{m}{2}\gamma_t(x)$$

and for any $a > 0$

$$\mathbf{P}\Big(|X - \mathbf{E}X| \ge a\Big) \le \frac{5\gamma_t(x)^{3/2}m^3}{a^2} .$$

The last ingredient we will need for the proof of Theorem 2.1 is the following classical result on the relation between the Cheeger constant of a Markov chain and its second eigenvalue.

Theorem 3.5. [[1], [2] ,[5]] *Let p be a reversible lazy chain (i.e., $p(x,x) \ge 1/2$ for all x) with Cheeger constant Φ_*. Write λ_2 for the second largest eigenvalue of the transition matrix. Then,*

$$\frac{\Phi_*^2}{2} \le 1 - \lambda_2 \le 2\Phi_* .$$

Proof of Theorem 2.1. Let us start by showing that the algorithm accepts with high probability the input graph is an α-expander. Denote by λ_2 the second largest eigenvalue of p, and recall the classical fact that

$$p^t(x,y) - \frac{1}{n} \le \lambda_2^t \le e^{-t(1-\lambda_2)} .$$

Theorem 3.5 implies that

$$p^t(x,y) - \frac{1}{n} \le e^{-t\Phi_*^2/2} \le e^{-\frac{t\alpha^2}{8d^2}}$$

since $\Phi_* \geq \alpha/2d$. Thus, for $t = \frac{16d^2}{\alpha^2}\log n$ we have that $\gamma_t(x) \leq \frac{1+O(n^{-1})}{n}$ for any x. We put $a = n^{7\mu/4}/200$ in Lemma 3.4 which in turn implies that the probability of having more than $M = n^{2\mu}/2 + n^{7\mu/4}/128$ pairwise collisions in each trial of the algorithm has probability at most $O(n^{-\mu/2})$. Since $N = 300/\epsilon$ we have that the algorithm accepts with high probability when G is an α-expander.

Assume now that G is ϵ-far from a β-expander of degree bound d. Lemma 3.3 shows that for at least $\epsilon n/128$ vertices x we have

$$\Delta_t^2(x) \geq \frac{n^{-\frac{32Cd^2\beta}{\alpha^2}}}{32n} .$$

Thus, as long as β satisfies

$$\beta \leq \frac{\mu\alpha^2}{128Cd^2} ,$$

we have by (3.1) that $\gamma_t(x) \geq \frac{1}{n} + \frac{n^{-1-\mu/4}}{32}$ for at least $\epsilon/128$ fraction of the vertices. In the notation of Lemma 3.4 this implies that for these vertices we have $\mathbf{E}X \geq n^{2\mu}/2 + n^{7\mu/4}/64$. Lemma 3.4 with $a = n^{7\mu/4}/128$ then implies that the probability of having less than $M = n^{2\mu}/2 + n^{7\mu/4}/128$ pairwise collisions in each trial of the algorithm has probability at most $O(n^{-\mu/2})$. Since $N = 300/\epsilon$ the algorithm will reject G with high probability. $\square$

4 Concluding Remarks

- Our main result in this paper is a tighter analysis of the Goldreich-Ron algorithm when applied to the problem of testing the vertex-expansion and edge-expansion of a graph. It seems interesting to check if the analysis can be further improved to show that the algorithm can distinguish between α-expanders and graphs that are ϵ-far from α-expanders.

- As we have explained at the end of Section 2, our main result can be used to partially resolve the conjecture of [9] on testing expansion that is defined by the second eigenvalue $\lambda(G)$ of the graph. The values of $\lambda(G)$ that our solutions covers is when $\lambda(G)$ lies in a certain interval. The conjecture of [9] for other values of $\lambda(G)$ remains open.

Acknowledgements

The first author would like to thank Microsoft Research, where this research was conducted, for their kind hospitality.

References

[1] N. Alon, Eigenvalues and expanders, Combinatorica 6(1986), 83-96.

[2] N. Alon and V. D. Milman, Eigenvalues, expanders and superconcentrators, Proc. 25^{th} Annual Symp. on Foundations of Computer Science (FOCS), Singer Island, Florida, IEEE (1984), 320-322. Also: λ_1, isoperimetric inequalities for graphs and superconcentrators, J. Combinatorial Theory, Ser. B 38(1985), 73-88.

[3] A. Czumaj and C. Sohler, Testing expansion in bounded-degree graphs, Proc. of FOCS 2007.

[4] A. Czumaj and C. Sohler, Sublinear-time algorithms, Bulletin of the EATCS, 89 (2006) 23-47.

[5] J. Dodziuk, Difference equations, isoperimetric inequality and transience of certain random walks, Trans. Amer. Math. Soc. 284 (1984), 787–794.

[6] E. Fischer, The art of uninformed decisions: A primer to property testing, The Computational Complexity Column of The Bulletin of the European Association for Theoretical Computer Science 75 (2001), 97-126.

[7] O. Goldreich and D. Ron, Property Testing in Bounded-Degree Graphs, Algorithmica 32 (2002), 302-343. Also, Proc. of STOC 1997, 406-415.

[8] O. Goldreich and D. Ron, A sublinear bipartiteness tester for bounded degree graphs, Combinatorica, 19 (1999), 335–373.

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

[10] S. Hoory, N. Linial and A. Wigderson, Expander graphs and their applications, Bulletin of the AMS, Vol 43 (4), 2006, 439-561.

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

[12] R. Rubinfeld, Sublinear time algorithms, Proc. of ICM 2006.

[13] D. Ron, Property testing, in: P. M. Pardalos, S. Rajasekaran, J. Reif and J. D. P. Rolim, editors, *Handbook of Randomized Computing*, Vol. II, Kluwer Academic Publishers, 2001, 597–649.