Testing Expansion in Bounded-Degree Graphs

Artur Czumaj *
Department of Computer Science
University of Warwick
Coventry CV4 7AL, UK
czumaj@dcs.warwick.ac.uk

Christian Sohler †
Department of Computer Science
University of Paderborn
33102 Paderborn, Germany
csohler@uni-paderborn.de

Abstract

We consider the problem of *testing expansion in bounded degree graphs*. We focus on the notion of *vertex-expansion*: an α-expander is a graph $G = (V, E)$ in which every subset $U \subseteq V$ of at most $|V|/2$ vertices has a neighborhood of size at least $\alpha \cdot |U|$. Our main result is that one can distinguish good expanders from graphs that are far from being weak expanders in time $\widetilde{\mathcal{O}}(\sqrt{n})$. We prove that the property testing algorithm proposed by Goldreich and Ron (2000) with appropriately set parameters accepts every α-expander with probability at least $\frac{2}{3}$ and rejects every graph that is ϵ-far from an α^*-expander with probability at least $\frac{2}{3}$, where $\alpha^* = \Theta\left(\frac{\alpha^2}{d^2 \log(n/\epsilon)}\right)$ and d is the maximum degree of the graphs. The algorithm assumes the bounded-degree graphs model with adjacency list graph representation and its running time is $\mathcal{O}\left(\frac{d^2 \sqrt{n} \log(n/\epsilon)}{\alpha^2 \epsilon^3}\right)$.

1 Introduction

Property testing is a relaxation of a standard decision problem: the task is to distinguish between an input object (for example, a graph, a function, or a point set) satisfying a certain predetermined property (for example, being bipartite, monotone, or in convex position) and an object that it is *"far"* from satisfying the property. The notion of being far from a property is parameterized by a distance parameter ϵ. An object is ϵ-far from having a property Π if it differs in more than an ϵ-fraction of its description from any object having the property Π. For example, when the object is a (dense) graph represented by an adjacency matrix and the considered property is bipartiteness, then a graph is ϵ-far from bipartite if one has to change more than $\epsilon\, n^2$ entries in the adjacency matrix to make it bipartite.

*Research supported in part by the *Centre for Discrete Mathematics and its Applications (DIMAP)*, University of Warwick.

†Research supported by DFG grant Me 872/8-3.

The main goal of property testing is to design randomized algorithms that can test a property without looking at the entire input. The algorithms sought should accept with probability at least $\frac{2}{3}$ every object that has the predetermined property and should reject with probability at least $\frac{2}{3}$ every object that is ϵ-far from having the property. Typically, one assumes that the input object is given as an oracle. Given access to this oracle, many properties can be tested in time *sublinear* in the input description size, and in some cases we can even achieve *running time independent of the input size*. In the last decade, property testing algorithms have been successfully developed for many different properties of combinatorial objects such as graphs and hypergraphs, functions, point sets, formal languages, and many other structures (for the references, see the surveys [10, 13, 14, 15, 20, 23]). In the area of graph properties, recently we have seen a vast amount of works dealing with properties of *dense graphs* represented by adjacency matrices. After a series of results for specific problems, recently very powerful general results characterizing testability of entire classes of graph properties have been established. For example, it has been proved that all hereditary properties are testable (even with one-sided error) in time independent of the input size (see [4] and [8]; cf. also [2]). These results imply that in the adjacency matrix model essentially any "natural" graph property can be tested with a constant number of queries.

While property testing in dense graphs is relatively well-understood, not much is known about property testing in *sparse graphs*. To study properties of sparse graphs one typically assumes that the input graph is given in the *bounded degree graph model* introduced by Goldreich and Ron [18]. In this model, the input graph G is represented by n *adjacency lists* of length at most d each, where n denotes the number of vertices of G and d, the maximum degree, is a constant independent of n. A property testing algorithm has oracle access to any entry in the adjacency list by making a query to the ith neighbor of a given vertex v. The number of accesses to the adjacency list is the query complexity of the tester. The task of property testing algorithm is to distinguish between graphs that satisfy a predetermined graph property Π and graphs that are ϵ-*far* from property Π; here we say that a graph G is ϵ-far from Π if one has to modify more than $\epsilon\, d\, n$ edges in G to obtain a graph having property Π.

In the bounded degree graph model we do not have any general results about testability of graph properties, and only a few, very simple graph properties (like connectivity) are known to be testable with constant number of queries [12, 18]. Furthermore, for most of non-trivial graph properties super-constant lower bounds exist (even for hereditary properties, for which constant-time testing is known in the adjacency matrix model). For example, acyclicity in directed graphs has $\Omega(n^{1/3})$ query complexity [6], the property of being bipartite has query complexity $\Omega(\sqrt{n})$ [18], and the query complexity of testing 3-colorability is $\Omega(n)$ [7]. Since one cannot hope to get property testers with constant query complexity, the focus turned on property testers with *sublinear* query complexity. One example of such a tester is the algorithm of [18] for testing bipartiteness with $\mathcal{O}(\sqrt{n})$ queries for constant ϵ, which combines the idea of random sampling with the approach of random walks on the input graph. Another approach to property testing in bounded degree graphs has been made recently in [11], where the authors assume that the input graph comes from a hereditary class of graphs that does not contain expanders, e.g., the class of planar graphs. Under this assumption, it is possible to show that every hereditary graph property is testable with constant number of queries.

In many of these property testing algorithms the crucial role has been played by the expansion of the graphs at hand. For example, in [18], the expansion properties of the graph are critical to "control" the analysis of the random walks and in [12], the constant-time testing is shown for classes of graphs with excluded expanders. In view of this, it is natural to ask if one can efficiently (in sublinear-time) test if a given graph is a good expander.

It is known that testing an expander requires at least $\Omega(\sqrt{n})$ time, which is the time needed even to distinguish between a very good expander and a disconnected graph with several large connected components [18]. Goldreich and Ron [17] made a major *attempt* to design a sublinear-time algorithm for testing expansion, however, unlike in our paper, they considered the *algebraic characterization of expanders*. In the contrast to majority of the property testing papers, Goldreich and Ron's goal was to consider a relaxation of the notion of property testing: their goal was to distinguish between graphs that are *good expanders* and those that are far from (relatively) *weak expanders*. More formally, their goal was to design an algorithm that for any given λ, $0 < \lambda < 1$, has query complexity of $\widetilde{\mathcal{O}}(\sqrt{n})$ and (with probability at least $\frac{2}{3}$) accepts every graph whose (normalized) second largest eigenvalue satisfies $\lambda(G) \leq \lambda$, and rejects every graph that is ϵ-far from any graph G' with $\lambda(G') \leq \lambda^*$, where $\lambda^* \gg \lambda$ but possibly $\lambda^* \leq \lambda^{\mathcal{O}(1)}$. Goldreich and Ron [17] give a property testing algorithm and show that under a certain appealing conjecture their algorithm is a property tester for expansion. Despite research efforts, the conjecture posed in [17] has not been resolved yet, and no sublinear-time algorithm testing expansion has been known; this problem has been considered as one of the major open questions in property testing.

1.1 Our result

In this paper we study the problem of testing expanders, but unlike Goldreich and Ron, we consider the combinatorial notion of *vertex-expansion*. We reconsider the algorithm of Goldreich and Ron [17] and prove that when run with appropriate parameters, this algorithm distinguishes between good expanders and graphs far from weak expanders with $\widetilde{\mathcal{O}}(\sqrt{n})$ queries. In particular, for any $\alpha > 0$, the algorithm has query complexity of $\mathcal{O}\left(\frac{d^2 \cdot \ln(n/\epsilon) \cdot \sqrt{n}}{\alpha^2 \cdot \epsilon^3}\right)$ and with probability $\frac{2}{3}$, it distinguishes between an α-expander and a graph that is ϵ-far from any α^*-expander, where $\alpha^* = \mathcal{O}\left(\frac{\alpha^2}{d^2 \log(n/\epsilon)}\right)$.

Our property tester uses the algorithm proposed by Goldreich and Ron [17] with appropriately set parameters and some parts of our analysis follow the ideas from that paper. And so, already Goldreich and Ron proved that the algorithm accepts good expanders (though their arguments use the algebraic notion of expanders). However, the missing part of the analysis in [17] and also the most challenging part of our analysis is a proof that if a graph is far from an expander then the algorithm will reject it. The main reason why this task is non trivial is that while we quite well understand properties of expanders (see e.g., the survey exposition in [19]), we know much less about graphs that are not expanders, and even less about the graphs that are far from expanders.

The proof that our algorithm will reject graphs that are far from expanders is in Sections 4.1 – 4.3. We prove this property by first showing that graphs that are far from being an expander have a large set of vertices for which the expansion is small. With this property at hand, we can prove that for many vertices, the random walk starting at these vertices will not quickly converge to the

uniform distribution. This can be used to demonstrate that our algorithm will reject all graphs that are far from expanders. This will imply that our algorithm is a tester for expansion.

Let us finally notice that since the relation between α and α^* depends (weakly) on n, our result does not imply the conjecture of [17], which still remains open.

1.1.1 Further comments

Our property testing algorithm for vertex-expansions can be easily extended to *edge-expansions* (see [19]), and the quantitative results are essentially identical. Furthermore, we can also test the expansion defined in the algebraic sense using the notion of the spectral gap, or the second largest eigenvalue of the normalized adjacency matrix, however this result is weaker than those conjectured by Goldreich and Ron in [17].

2 Preliminaries

Let $G = (V, E)$ be an undirected graph with n vertices. Without loss of generality, we assume that $V = \{1, \ldots, n\}$. For a subset $U \subseteq V$ we use $G[U]$ to denote the subgraph induced by U. For a vertex v let $\deg(v)$ denote its degree. We assume that G is stored in the *adjacency list* model for bounded degree graphs with degree bound d, i.e., $\deg(v) \leq d$ for all $v \in V$. In this model, we have constant time access to a function $f_G : [n] \times [d] \to [n] \cup \{+\}$. The value $f_G(v, i)$ is the ith neighbor of v or a special symbol $+$ if v has less than i neighbors. In this paper we will assume $d \geq 4$. Note that this does not imply that the maximum degree of the graph must be 4.

2.1 Expanders and graphs that are far from expanders

In our paper, our main focus will be on the definition of expanders through the *vertex-expansion*. For any two sets of vertices $U, W \subseteq V$, we denote by $N_G(U, W)$ the *neighborhood of* U *in* W with respect to the graph G, that is, $N_G(U, W) = \{v \in W \setminus U : \exists u \in U \text{ such that } (v, u) \in E\}$. When G is clear from the context, we will skip the dependency of G and write $N(U, V)$ instead.

Definition 2.1 *A graph* $G = (V, E)$ *is a* ***α-expander*** *if for every* $U \subseteq V$ *with* $|U| \leq n/2$ *holds* $|N(U, V)| \geq \alpha \cdot |U|$.

Let us remind that our main goal is to study graphs with degree at most d. With this in mind, we have the following definition of graphs that are ϵ-far from expanders.

Definition 2.2 *A graph* G *is* ***ϵ-far from α-expander*** *if one has to modify more than* $\epsilon\, d\, n$ *entries in* f_G *to obtain an* α*-expander.*

Equivalently, G *is* ϵ*-far from* α*-expander if one has to modify more than* $\epsilon\, d\, n/2$ *edges to obtain a graph* G' *that has maximum degree at most* d *and that is an* α*-expander.*

2.1.1 Strong expanders and parameter β_d

We will also use the notion of strong expanders.

Definition 2.3 *A graph* $G = (V, E)$ *is called* β***-strong expander*** *if for every* $U \subseteq V$ *holds* $|N(U, V)| \geq \beta \cdot \min\{|U|, |V \setminus U|\}$.

It is not hard to see that if a graph G with maximum degree d^* is an α-expander then it is also α/d^*-strong expander. Indeed, if $G = (V, E)$ is an α-expander then for every subset $U \subseteq V$, if $|U| \leq |V|/2$ then $|N(U, V)| \geq \alpha \cdot |U|$. On the other hand, if $|U| > |V|/2$, then $|V \setminus U| \leq |V|/2$, and hence $|N(V \setminus U, V)| \geq \alpha \cdot |V \setminus U|$. This in turn, because each vertex in G has degree at most d^*, implies that $|N(U, V)| \geq \frac{|N(V \setminus U, V)|}{d^*} \geq \frac{\alpha}{d^*} |V \setminus U|$.

Let β_d^* be the largest number for which there is a d-regular β_d^*-strong expander. It is not difficult to prove that for every $d \geq 3$ we have $\beta_d^* \geq c > 0$ for certain constant c (e.g., by considering a random d-regular graph), and therefore, from now on, we will use $\beta_d = \min\{\beta_d^*, \frac{1}{3}\} = \Theta(1)$, with the meaning that for every $d \geq 3$ there is a β_d-strong expander of maximum degree d.

2.2 Random walks and Markov Chains

The main technical tool in our analysis will be the study of *random walks* on G. We describe a random walk by a transition matrix P. For two vertices $u, v \in V$, let $P(u, v)$ define the probability to move from vertex u to v.

We will first make sure that the random walk is on a regular graph and that the graph is not bipartite (this is because it is known that for such a graph, if it is connected then the random walk converges to a uniform distribution). This will be achieved by a standard trick of adding (*virtually*, without modifying the graph) to each vertex v, $2\,d - \deg(v)$ self-loops. In this way we obtain a $2d$-regular graph. For two vertices v, w, we define $P(v, w) = \frac{1}{2d}$, if $(v, w) \in E$ and $P(v, w) = 0$, otherwise. Furthermore, we define $P(v, v) = \frac{2d - \deg(v)}{2d} = 1 - \frac{\deg(v)}{2d}$.

To analyze the random walk we will use some standard tools from the analysis of Markov chains. We give here a brief introduction following the exposition in [21]. Let $\mathfrak{M}$ be a Markov chain with finite state space Ω and stationary distribution $\boldsymbol{\pi} = (\pi_x)_{x \in \Omega}$. Let $P : \Omega^2 \to [0, 1]$ denote the matrix of its transition probabilities. We use $P^t(x, \cdot)$ to denote the distribution at time t for given initial state x, i.e., $P^t(x, S)$ is the probability that the Markov chain with initial state x ends after t steps in a state $y \in S$. We assume that $\mathfrak{M}$ is *ergodic* (connected and aperiodic) and is *reversible*, i.e., that $\pi_x \cdot P(x, y) = \pi_y \cdot P(y, x)$ for all $x, y \in \Omega$. We define $Q(x, y) = \pi_x \cdot P(x, y)$ for all $x, y \in \Omega$.

Definition 2.4 *The* ***variation distance*** *with respect to the initial state* x *is defined as*

$$\Delta_x(t) \;=\; \max_{S \subseteq \Omega} \left| P^t(x, S) - \pi(S) \right| \;=\; \tfrac{1}{2} \sum_{y \in \Omega} \left| P^t(x, y) - \pi(y) \right| \;.$$

Definition 2.5 *The* ***rate of convergence*** *of* $\mathfrak{M}$ *with initial state* x *to the stationary distribution is defined as*

$$\tau_x(\zeta) = \min\{t : \Delta_x(t') \leq \zeta \;\; \textit{for all} \;\; t' \geq t\} \;.$$

*We also call $\tau_x(\zeta)$ the **mixing time** of the Markov chain.*

Definition 2.6 (Conductance [25]) *The **conductance** of a Markov chain $\mathfrak{M}$ is defined as*

$$\Phi \;=\; \Phi(\mathfrak{M}) \;=\; \min_{S\subseteq\Omega, 0<\pi(S)\leq\frac{1}{2}} \frac{Q(S,\overline{S})}{\pi(S)} \;,$$

where $Q(S,\overline{S}) = \sum_{(x,y)\in E, x\in S, y\in\overline{S}} Q(x,y)$ and $\overline{S} = \Omega \setminus S$.

Proposition 2.7 [24] *Let $\mathfrak{M}$ be a finite, reversible, ergodic Markov chain with loop probabilities $P(x,x) \geq \frac{1}{2}$ for all states x. Let Φ be the conductance of $\mathfrak{M}$. Then the mixing time of $\mathfrak{M}$ satisfies*

$$\tau_x(\zeta) \leq 2\Phi^{-2}(\ln(\pi_x^{-1}) + \ln(\zeta^{-1})) \;,$$

for any choice of initial state x.

The previously defined random walk can also be viewed as a Markov chain $\mathfrak{M}_G$ with state space $\Omega = V$ and transition matrix P. It is well-known (and easy to see) that if G is connected then $\mathfrak{M}_G$ is ergodic, reversible, has $P(x,x) \geq \frac{1}{2}$ for all $x \in V$, and the stationary distribution of $\mathfrak{M}_G$ is uniform. It is easy to see that the conductance Φ_G of $\mathfrak{M}_G$ is given by

$$\Phi_G = \Phi(\mathfrak{M}_G) = \min_{U\subseteq V, |U|\leq|V|/2} \frac{|E(U, V\setminus U)|}{2\,d\cdot|U|} \;,$$

where $E(U, V\setminus U)$ is the set of edges of edges between U and $V\setminus U$.

Notice also that for an α-expander we have

$$\Phi_G \;=\; \min_{U\subseteq V, |U|\leq|V|/2} \frac{|E(U, V\setminus U)|}{2\,d\cdot|U|} \;\geq\; \min_{U\subseteq V, |U|\leq|V|/2} \frac{|N(U,V)|}{2\,d\cdot|U|} \;\geq\; \frac{\alpha}{2\,d} \;.$$

3 Property testing of expanders

In this section we present the property testing algorithm and state our main theorem which describes its performance. The algorithm is essentially the algorithm proposed by Goldreich and Ron [17] for testing expanders with the parameters set appropriately to facilitate our analysis.

```
EXPANSIONTESTER(G, ℓ, m, s)
 repeat s times
    Select vertex v uniformly at random from V
    Perform m random walks of length ℓ starting from v
    Count the number of pairwise collisions between
        the endpoints of these m random walks
    if the number of pairwise collisions is greater than (1+7ε)/n · C(m,2)
        then reject
 accept
```

The query complexity of the algorithm is $\mathcal{O}(\ell \cdot m \cdot s)$. In the next section we will prove the following theorem, which is the central result of this paper.

Theorem 1 *Let $0 \leq \epsilon \leq 1/100$. Algorithm* EXPANSIONTESTER *with $s \geq 16/\epsilon$, $m \geq \frac{12 \cdot s \cdot \sqrt{n}}{\epsilon^2}$, and $\ell \geq \frac{16 \cdot d^2 \cdot \ln(n/\epsilon)}{\alpha^2}$, accepts every α-expander with probability at least $\frac{2}{3}$ and rejects every graph that is ϵ-far from a $\frac{c \cdot \alpha^2}{d^2 \cdot \ln(n/\epsilon)}$-expander with probability at least $\frac{2}{3}$, where $c > 0$ is a large enough constant. The query complexity of this algorithm is $\mathcal{O}\left(\frac{d^2 \cdot \ln(n/\epsilon) \cdot \sqrt{n}}{\alpha^2 \cdot \epsilon^3}\right)$.*

In what follows we use $\alpha^* = \frac{c \cdot \alpha^2}{d^2 \cdot \ln(n/\epsilon)}$, and thus our goal is to show that algorithm EXPANSIONTESTER accepts every α-expander with probability at least $\frac{2}{3}$ and rejects every graph that is ϵ-far from an α^*-expander with probability at least $\frac{2}{3}$.

4 Analysis

The main idea of algorithm EXPANSIONTESTER is to exploit the relation between good expansion and fast convergence of random walks (in regular graphs) to the uniform distribution [17] (see also [19]). The length of the random walk is chosen in such a way that for an α-expander the distribution of the endpoints of the random walk will be close to the uniform distribution. The algorithm compares the number of collisions among the endpoints with the expected number of collision in the uniform distribution. If this number is too large then the algorithm rejects. The key and most challenging part of the analysis (which forms Sections 4.1 – 4.3) is to prove that if a graph is far away from an α^*-expander (with $\alpha^* = \frac{c \cdot \alpha^2}{d^2 \cdot \ln(n/\epsilon)}$) then the rejection will most likely happen.

It has already been proved in [17] (see the proof of Lemma 1 in [17] and also [5]) that the variance of this process is relatively small.

Lemma 4.1 (Goldreich, Ron [17]) *Let X_v be the random variable for the number of collisions among the endpoints of m random walk of length ℓ starting from vertex v. Let P_v^ℓ denote the distribution of the endpoint of this random walk. Then the following holds:*

$$\mathbf{E}[X_v] = \binom{m}{2} \cdot \|P_v^\ell\|_2^2 \textit{ and } \mathbf{Var}[X_v] \leq 2 \cdot (\mathbf{E}[X_v])^{3/2} \ .$$

With this result we can follow the arguments of Goldreich and Ron [17] to show that an α-expander is accepted provided that the random walks are sufficiently long. Let us consider an α-expander and a random walk of length $\ell = \frac{16\, d^2 \cdot \ln(n/\epsilon)}{\alpha^2} \geq \frac{8\, d^2 \cdot (\ln(n) + \ln(n/\epsilon))}{\alpha^2}$. By applying Proposition 2.7 we observe that for any starting vertex $v \in V$, the variation distance between P_v^ℓ and the uniform distribution is at most ϵ/n. This yields $\|P_v^\ell\|_2^2 \leq (1+\epsilon)^2/n$. By combining this bound with Lemma 4.1 and Chebyshev inequality, we can obtain the following lemma (a similar lemma was proven for algebraic notion of expansion in [17]).

Lemma 4.2 (Accepting expanders) *Let* $m \geq \frac{12 \cdot s \cdot \sqrt{n}}{\epsilon^2}$ *and* $\ell \geq \frac{8\, d^2 \cdot (\ln(n) + \ln(n/\epsilon))}{\alpha^2}$. *Then Algorithm* EXPANSIONTESTER *accepts every* α*-expander with probability at least* $\frac{2}{3}$.

Proof : Let X_v be the random variable for the number of collisions of m random walks of length ℓ starting at vertex v. By combining Lemma 4.1 with Chebyshev inequality we obtain,

$$\mathbf{Pr}\big[\, |X_v - \mathbf{E}[X_v]| \geq \epsilon \cdot \mathbf{E}[X_v] \big] \;\leq\; \frac{\mathbf{Var}[X_v]}{\epsilon^2 \cdot \mathbf{E}[X_v]^2} \;\leq\; \frac{4}{\epsilon^2 \cdot m \cdot \|P_v^\ell\|_2} \;\leq\; \frac{4 \cdot \sqrt{n}}{\epsilon^2 \cdot m} \;,$$

where the last inequality follows from the fact that $\|P_v^\ell\|_2 \geq \frac{1}{\sqrt{n}}$, since P_v^ℓ is a probability distribution. Hence, by choosing

$$m \geq \frac{12 \cdot s \cdot \sqrt{n}}{\epsilon^2}$$

we obtain

$$\mathbf{Pr}\big[\, |X_v - \mathbf{E}[X_v]| \geq \epsilon \cdot \mathbf{E}[X_v] \big] \leq \frac{1}{3s} \;.$$

Therefore, by the union bound we obtain

$$\mathbf{Pr}\big[\forall v \in S : |X_v - \mathbf{E}[X_v]| \leq \epsilon \cdot \mathbf{E}[X_v] \big] \geq \frac{2}{3} \;.$$

Thus, using the fact that $\mathbf{E}[X_v] = \binom{m}{2} \cdot \|P_v^\ell\|_2^2$ by Lemma 4.1, with probability at least $\frac{2}{3}$ we obtain that there are at most

$$(1+\epsilon)\binom{m}{2}\frac{(1+\epsilon)^2}{n} = \frac{(1+\epsilon)^3}{n}\binom{m}{2} \leq \frac{1+7\epsilon}{n}\binom{m}{2}$$

collisions for each $v \in S$. Hence, the algorithm accepts with probability at least $\frac{2}{3}$. □

Lemma 4.3 (Rejections) *Let* $0 < \epsilon < \frac{1}{100}$. *Let* $G = (V, E)$ *a graph and for each* $v \in V$ *let* P_v^ℓ *be the distribution of the endpoints of an* ℓ*-step random walk starting at* v. *If there exists a set* $U \subseteq V$ *of cardinality at least* $\delta\, n$, $0 < \delta < \frac{1}{2}$, *such that for every* $u \in U$ *the variation distance of* P_u^ℓ *to the uniform distribution is at least* $20\,\epsilon$, *then algorithm* EXPANSIONTESTER *run with* $s \geq 2/\delta$ *rejects with probability at least* $\frac{2}{3}$.

Proof : We first show that any distribution with variation distance at least ϵ to the uniform distribution will result in a high expected number of collisions for our random walk. Let $\mathcal{U}$ denote the uniform distribution. It is easy to see that the variation distance between P_v^ℓ and $\mathcal{U}$ is $\frac{1}{2} \cdot \|P_v^\ell - \mathcal{U}\|_1$. By Lemma 4.1, the expected number of collisions is $\binom{m}{2} \cdot \|P_v^\ell\|_2^2$. The smaller this expected value is, the harder we can distinguish the corresponding distribution from the uniform distribution, which minimizes this expected value. Thus, we are looking for the vector P_v^ℓ with $\|P_v^\ell - \mathcal{U}\|_1 \geq 10\,\epsilon$ that minimizes $\|P_v^\ell\|_2^2$. This is attained by a vector with $10\,\epsilon\, n$ zeros and the remaining values set to $\frac{1}{n - 10\,\epsilon\, n}$. Then we get

$$\|P_v^\ell\|_2^2 \;\geq\; (n - 10\epsilon n) \cdot \left(\frac{1}{n - 10\epsilon n}\right)^2 \;\geq\; \frac{1}{1 - 10\epsilon} \cdot \frac{1}{n} \;\geq\; \frac{1 + 10\epsilon}{n} \;.$$

We know from the proof of Lemma 4.2 that the observed number of collision is with probability at least $1 - \frac{1}{3s}$ at least $(1-\epsilon)$ times its expectation. Thus, we get that with probability at least $1 - \frac{1}{3s}$ we have more than

$$(1-\epsilon) \cdot \binom{m}{2} \cdot \|P_v^\ell\|_2^2 \; > \; \frac{1+7\epsilon}{n} \cdot \binom{m}{2}$$

collisions.

The probability that algorithm EXPANSIONTESTER selects a vertex from U as one of the starting vertices for the random walk is $1-(1-\delta)^s$, and therefore the overall probability of the rejection is at least

$$(1-(1-\delta)^s) \cdot \left(1 - \frac{1}{3s}\right) \geq \frac{2}{3}$$

for $\delta < \frac{1}{2}$ and $s \geq 2/\delta$. □

4.1 Being far from expanders

The main contribution of this paper is a proof that a graph that is ϵ-far from an α^*-expander is indeed rejected by algorithm EXPANSIONTESTER with probability at least $\frac{2}{3}$. To prove this we first show that any graph that is ϵ-far from an α^*-expander has a small cut that separates a large set of vertices from the rest of the graph, i.e., it has a small ratio cut. (In this section we use the parameter β_d defined in Section 2.1.1.)

Lemma 4.4 *Let $\alpha^* \leq \frac{1}{3}$. If $G = (V, E)$ has a subset of vertices $A \subseteq V$ with $|A| \leq \epsilon\, n/4$ such that $G[V \setminus A]$ is an $4\,\alpha^*/\beta_{d-1}$-expander, then G is not ϵ-far from α^*-expander.*

Proof : To prove the lemma, we show that if there is $A \subseteq V$ with $|A| \leq \epsilon\, n/4$ such that $G[V \setminus A]$ is an $4\,\alpha^*/\beta_{d-1}$-expander, then we can modify $\epsilon\, d\, n/2$ edges in G to obtain a graph G^* that has maximum degree at most d and is an α^*-expander.

For simplicity of presentation we assume that $|A|$ is even; extension to the general case is straightforward.

We will construct G^* from G using the following procedure:

1. Remove all edges incident to A.
2. Add a $(d-1)$-regular β_{d-1}-strong expander on A.
3. Remove an arbitrary matching M of size $|A|/2$ in $G[V \setminus A]$; let V_M denote the set of endpoints of the edges in M.
4. Add an arbitrary matching between A and V_M.

It is easy to see that if G has maximum degree at most d then so has G^*, and that G^* is obtained from G by modifying at most $d\,|A| + (d-1)\,|A|/2 + |A|/2 + |A| = (d+1+d/2)\,|A| \leq 2\,d\,|A| \leq \epsilon\, d\, n/2$ edges in G. Therefore now, we only have to prove that G^* is an α^*-expander.

To show that G^* is an α^*-expander we have to show that for any set $X \subseteq V$ with $|X| \leq n/2$, $|N_{G^*}(X, V)| \geq \alpha^* |X|$. To do this, let us fix any set $X \subseteq V$ with $|X| \leq n/2$ and let $X_A = X \cap A$ and $X_B = X \cap (V \setminus A)$. We begin with three useful inequalities for the sizes of the neighborhoods of X_A and X_B. Let $B = V \setminus A$.

Claim 4.5 *The following three inequalities hold for* G^**:*

(1) $|N(X_A, V \setminus X) \cap B| \geq |X_A| - |X_B|$*;*

(2) $|N(X_A, V \setminus X) \cap A| \geq \beta_{d-1} \cdot \min\{|X_A|, |A \setminus X_A|\}$*;*

(3) $|N(X, V)| \geq \frac{4\,\alpha^*\,|X_B|}{\beta_{d-1}} - \min\{|X_A|, |A \setminus X_A|\}$.

Proof of Claim 4.5: Since $N(X_A, V \setminus X) \cap B$ is the set of the neighbors of the vertices in X_A that are contained in $B \setminus X_B$, the first inequality follows from the fact that each vertex in A (and hence in X_A) is connected in G^* to a unique vertex in B.

Since $N(X_A, V \setminus X) \cap A$ is the set of the neighbors of the vertices in X_A that are contained in $A \setminus X_A$, the second inequality follows from the fact that the vertices in A are connected by a β_{d-1}-strong expander in G^*.

To see the third inequality, let us first observe that in G, since $G[V \setminus A]$ is an $4\,\alpha^*/\beta_{d-1}$-expander, X_B has at least $\frac{4\,\alpha^*\,|X_B|}{\beta_{d-1}}$ neighbors in $V \setminus (A \cup X_B)$. Next, in our construction of G^*, we may first remove from G some of the edges on the cut $(X_B, V \setminus (A \cup X_B))$ and then add some of the edges in the matching between A and V_M. Each time we remove an edge $(x, y) \in M$ with $x \in X_B$ and $y \notin X_B$, then we will connect x and y to two vertices in A. By removing (x, y), we may decrease the size of $N(X_B, V)$ by one, but inserting an edge (x, z) with $z \in A$ may potentially increase the size of $N(X_B, V)$ by one. Therefore, after the operation of removing (x, y) and inserting the edges (x, z) and (y, w) with $z, w \in A$, the only case when the size of $N(X, V)$ decreases by one is when $z \in X_A$ and $w \in A \setminus X_A$. (Indeed, if $z \in A \setminus X_A$, then removal of (x, y) decrease the size of $N(X, V)$ by one but insertion of (x, z) increase it back by one. Similarly, if $w \in X_A$, then removal of (x, y) decrease the size of $N(X, V)$ by one but insertion of (y, w) increase it back by one.) Now, each such a decrease requires that one of the vertices from X_A is matched and one of the vertices from $A \setminus X_A$ is matched, and therefore the maximum decrease is at most $\min\{|X_A|, |A \setminus X_A|\}$. Hence, after these operations, we started with $|N(X, V)| \geq \frac{4\,\alpha^*\,|X_B|}{\beta_{d-1}}$, and then we are able to decrease this value by at most $\min\{|X_A|, |A \setminus X_A|\}$; hence $|N(X, V)| \geq \frac{4\,\alpha^*\,|X_B|}{\beta_{d-1}} - \min\{|X_A|, |A \setminus X_A|\}$. □

Now, once we have Claim 4.5, we are ready to proceed with the proof of Lemma 4.4.

- If $|X_A| \geq 2\,|X_B|$ then by Claim 4.5 (1) and our assumption that $\alpha^* \leq \frac{1}{3}$:

$$\begin{aligned} |N(X, V)| &\geq |N(X_A, V \setminus X) \cap B| \geq |X_A| - |X_B| \\ &\geq \tfrac{1}{2}\,|X_A| \geq \tfrac{1}{3}\,(|X_A| + |X_B|) \\ &= \tfrac{1}{3}\,|X| \geq \alpha^*\,|X| \ . \end{aligned}$$

- If $|X_A| < 2\,|X_B|$ then we consider two cases.
  - If $\min\{|X_A|, |A \setminus X_A|\} > \frac{3\,\alpha^*\,|X_B|}{\beta_{d-1}}$ then by Claim 4.5 (2):

$$\begin{aligned} |N(X,V)| &\geq |N(X_A, V \setminus X) \cap A| \\ &\geq \beta_{d-1} \cdot \min\{|X_A|, |A \setminus X_A|\} \\ &> \beta_{d-1} \cdot \frac{3\,\alpha^*\,|X_B|}{\beta_{d-1}} \\ &= 3\,\alpha^*\,|X_B| \;\geq\; \alpha^*\,|X| \;. \end{aligned}$$

  - If $\min\{|X_A|, |A \setminus X_A|\} \leq \frac{3\,\alpha^*\,|X_B|}{\beta_{d-1}}$ then by Claim 4.5 (3):

$$\begin{aligned} |N(X,V)| &\geq \frac{4\,\alpha^*\,|X_B|}{\beta_{d-1}} - \min\{|X_A|, |A \setminus X_A|\} \\ &\geq \frac{4\,\alpha^*\,|X_B|}{\beta_{d-1}} - \frac{3\,\alpha^*\,|X_B|}{\beta_{d-1}} \\ &= \frac{\alpha^*\,|X_B|}{\beta_{d-1}} \;\geq\; \alpha^*\,|X| \;, \end{aligned}$$

    where in the last inequality we used the fact that $\beta_{d-1} \leq \frac{1}{3}$. □

Now we can apply Lemma 4.4 to obtain that there is a small cut that separates a large set of vertices from the rest of the graph.

Corollary 4.6 *Let $G = (V, E)$ be ϵ-far from an α^*-expander with $\alpha^* \leq \frac{1}{3}$. Then there is a subset of vertices $A \subseteq V$ with $\epsilon n/4 \leq |A| \leq (1+\epsilon) \cdot n/2$ such that $|N_G(A, V)| < \frac{4\,\alpha^*\,|A|}{\beta_{d-1}}$.*

Proof : Let G be a graph that is ϵ-far from an α^*-expander. Let $A_1 = \emptyset$. We apply Lemma 4.4 with $A = A_1$, from which it follows that $G[V \setminus A]$ is not a $\frac{4\,\alpha^*}{\beta_{d-1}}$-expander. Thus there must be a set $A_2 \subseteq V \setminus A$ of size at most $n/2$ with $|N_G(A_2, V \setminus A)| \leq \frac{4\,\alpha^*}{\beta_{d-1}} \cdot |A_2|$. If $|A_1 \cup A_2| < \epsilon\, n/4$ then we repeat this approach. We set $A = A_1 \cup A_2$ and apply Lemma 4.4 to obtain a new set A_3 of at most $n/2$ vertices with small neighborhood. We can repeat this process until $|\bigcup_i A_i| \geq \epsilon\, n/4$. Now, let us write $A = |\bigcup_i A_i|$. By our construction we have $|A| \leq (1+\epsilon)\, n/2$. It is also easy to verify that $|N_G(A, V)| \leq \frac{4\,\alpha^*}{\beta_{d-1}} \cdot |A|$. □

4.2 A small ratio cut implies bad mixing time

With Corollary 4.6 at hand, to complete our analysis it remains to show that the existence of a small cut separating a large fraction of vertices from the rest of the graph implies that for many starting vertices a random walk converges slowly to the uniform distribution.

Let A be a set of vertices with $|A| \leq (1+\epsilon) \cdot n/2$ and $|N_G(A, V)| \leq \gamma \cdot |A|$ for $\gamma \leq \frac{1}{10(\ell+1)}$. Let us define the graph G_A to be the graph induced by $A \cup N_G(A, V)$, that is, $G_A = G[A \cup N_G(A, V)]$.

We consider a random walk on $G_A = (V_A, E_A)$, i.e., for every vertex we add self-loops so that we obtain a $(2d)$-regular graph and then we consider a random walk on this graph.

Our random walk will start in a vertex chosen uniformly at random from V_A and it will take ℓ steps. Let Y_i denote the indicator random variable for the event that the ith vertex of the walk is in $N_G(A, V)$. Since the starting vertex is chosen uniformly at random and since the stationary distribution is uniform, we know that the ith vertex is distributed uniformly at random in V. Hence,

$$\mathbf{Pr}[Y_i = 1] = \frac{|N_G(A, V)|}{|V_A|} \ .$$

By linearity of expectation, the expected number of vertices from $N_G(A, V)$ that are visited by our random walk is

$$\mathbf{E}\Big[\sum_{i=0}^{\ell} Y_i\Big] \ = \ (\ell + 1) \cdot \frac{|N_G(A, V)|}{|V_A|} \ \leq \ \gamma \cdot (\ell + 1) \ .$$

Therefore the probability that any vertex from $N_G(A, V)$ is visited during the random walk is

$$\mathbf{Pr}[\exists i : Y_i = 1] \leq \mathbf{E}\Big[\sum_{i=0}^{\ell} Y_i\Big] \leq \gamma \cdot (\ell + 1) \ .$$

Let us now move to the random walk in G. The claim above implies that the probability that an ℓ-step random walk in G starting at a vertex chosen uniformly at random from A will remain in A with probability at least $\max\{0, 1 - \gamma \cdot (\ell + 1)\}$. Now, by setting $\gamma \leq \frac{1}{10(\ell+1)}$, we stay in A with probability at least $\frac{9}{10}$. Therefore, there must be a set $U \subseteq A$ of at least $|A|/2$ vertices such that a random walk starting from a vertex in U remains in A with probability at least $\frac{3}{4}$. Next, let us observe that since $|V \setminus A| \geq (1 - \epsilon) \cdot n/2$, in the uniform distribution, a vertex from outside A must be chosen with the probability $\frac{|V \setminus A|}{|V|} \geq \frac{1-\epsilon}{2}$, in contrast to the distribution of the endpoints of an ℓ-step random walk in G starting at a vertex from U, where this probability is at most $\frac{1}{4}$. This implies that the variation distance of the distribution of the endpoints of an ℓ-step random walk starting in U to the uniform distribution is at least $\frac{1-\epsilon}{4}$. We conclude:

Lemma 4.7 *Let A be a set of points with $|A| \leq (1 + \epsilon) \cdot n/2$ and $|N_G(A, V)| \leq \frac{|A|}{10(\ell+1)}$. Let P_v^ℓ be the distribution of the endpoints of an ℓ-step random walk starting in v. Then there exists a set U, $|U| \geq |A|/2$, such that for every $v \in U$, the variation distance from P_v^ℓ to the uniform distribution is at least $\frac{1-\epsilon}{4}$.*

4.3 Putting everything together: proof of Theorem 1

Now, we are ready to conclude the proof of Theorem 1. Let us define $\alpha^* = \frac{\beta_{d-1}}{40(\ell+1)} = \Theta(\frac{\alpha^2}{d^2 \cdot \ln(n/\epsilon)})$, where we use the fact that $\beta_{d-1} = \Theta(1)$. (Notice that $\alpha^* \leq \frac{1}{3}$.) Let us consider a graph $G = (V, E)$ that is ϵ-far from an α^*-expander. By Corollary 4.6, G has a set of vertices A with $\epsilon\, n/4 \leq |A| \leq (1 + \epsilon)\, n/2$ and $|N(A, V)| \leq \frac{4\,\alpha^*\,|A|}{\beta_{d-1}}$. Hence, by our setting of α^*, we have $|N(A, V)| \leq \frac{|A|}{10\,(\ell+1)}$. Next, by Lemma 4.7, there exists a set U of cardinality at least $|A|/2 \geq \epsilon\, n/8$ such that for each

vertex $v \in U$, P_v^ℓ has variation distance to the uniform distribution of at least $\frac{1-\epsilon}{4}$. Finally, we apply Lemma 4.3 to conclude that G is rejected with probability at least $\frac{2}{3}$. From Lemma 4.2, it follows that every α-expander is accepted with probability $\frac{2}{3}$ and so Theorem 1 follows. □

5 Conclusions

So far, most results on property testing in sparse graphs involve some notion of expansion. Therefore the understanding of the effect of expansion on property testing seems to be one of the key challenges on the way to understand property testing in sparse graphs. In this paper we make a major step toward the understanding this relation by analyzing a property testing algorithm for expansion of graphs that has been proposed in [17]. It was conjectured in [17] that the algorithm is indeed a property tester for expansion and in this paper, we show that (under a different, combinatorial notion of expansion and with a slightly larger gap in the expansion between the graphs that are to be accepted and those to be rejected) the proposed algorithm is indeed a property tester for expansion.

Our result does not settle the conjecture from [17] and the status of this conjecture is still open. However, if one could remove the $\mathcal{O}(\log n)$-factor in the gap between α and α^* then we believe that a similar result as that conjectured by Goldreich and Ron could be proven.

5.1 Acknowledgements

We thank Noga Alon and Mark Jerrum for their helpful comments.

References

[1] N. Alon. Eigenvalues and expanders. *Combinatorica*, 6(2): 83–96, 1986.

[2] N. Alon, E. Fischer, I. Newman, and A. Shapira. A combinatorial characterization of the testable graph properties: it's all about regularity. *Proceedings of the 38th Annual ACM Symposium on Theory of Computing (STOC)*, pp. 251–260, 2006.

[3] N. Alon and V. D. Milman. λ_1, isoperimetric inequalities for graphs, and superconcentrators. *Journal of Combinatorial Theory, Series B*, 38(1): 73–88, 1985.

[4] N. Alon and A. Shapira. A characterization of the (natural) graph properties testable with one-sided error. *Proceedings of the 46th IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 429–438, 2005.

[5] T. Batu, L. Fortnow, R. Rubinfeld, W. Smith, and P. White. Testing that distributions are close. *Proceedings of the 41st IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 259-269, 2000.

[6] M. A. Bender and D. Ron. Testing properties of directed graphs: acyclicity and connectivity. *Random Structures and Algorithms*, 20(2): 184–205, 2002.

[7] A. Bogdanov, K. Obata, and L. Trevisan. A lower bound for testing 3-colorability in bounded-degree graphs. *Proceedings of the 43rd IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 93–102, 2002.

[8] C. Borgs, J. Chayes, L. Lovász, V. T. Sos, B. Szegedy, and K. Vesztergombi. Graph limits and parameter testing. *Proceedings of the 38th Annual ACM Symposium on Theory of Computing (STOC)*, pp. 261–270, 2006.

[9] A. Czumaj and C. Sohler. Abstract combinatorial programs and efficient property testers. *SIAM Journal on Computing*, 34(3): 580–615, 2005.

[10] A. Czumaj and C. Sohler. Sublinear-time algorithms. *Bulletin of the EATCS*, 89: 23–47, June 2006.

[11] A. Czumaj, A. Shapira, and C. Sohler. Testing hereditary properties of non-expanding bounded-degree graphs. Submitted, 2007. This is a full version of [12].

[12] A. Czumaj and C. Sohler. On testable properties in bounded degree graphs. *Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 494–501, 2007. Full version appeared in [11].

[13] E. Fischer. The art of uninformed decisions: A primer to property testing. *Bulletin of the EATCS*, 75: 97–126, October 2001.

[14] O. Goldreich. Combinatorial property testing (a survey). In P. Pardalos, S. Rajasekaran, and J. Rolim, editors, *Proceedings of the DIMACS Workshop on Randomization Methods in Algorithm Design*, volume 43 of *DIMACS, Series in Discrete Mathetaics and Theoretical Computer Science*, pp. 45–59, 1997. American Mathematical Society, Providence, RI, 1999.

[15] O. Goldreich. Property testing in massive graphs. In J. Abello, P. M. Pardalos, and M. G. C. Resende, editors, *Handbook of massive data sets*, pp. 123–147. Kluwer Academic Publishers, 2002.

[16] O. Goldreich and D. Ron. A sublinear bipartiteness tester for bounded degree graphs. *Combinatorica*, 19(3):335–373, 1999.

[17] O. Goldreich and D. Ron. On testing expansion in bounded-degree graphs. *Electronic Colloquium on Computational Complexity (ECCC)*, Report No. 7, 2000.

[18] O. Goldreich and D. Ron. Property testing in bounded degree graphs. *Algorithmica*, 32(2): 302–343, 2002.

[19] S. Hoory, N. Linial, and A. Wigderson. Expander graphs and their applications. *Bulletin of the American Mathematical Society*, 43(4): 439–561, October 2006.

[20] R. Kumar and R. Rubinfeld. Sublinear time algorithms. *SIGACT News*, 34: 57–67, 2003.

[21] M. Jerrum and A. Sinclair. The Markov chain Monte Carlo method: An approach to approximate counting and integration. Chapter 12 in *Approximation Algorithms for NP-hard Problems*, edited by D. Hochbaum, PWS Publishing, Boston, 1996.

[22] M. Parnas and D. Ron. Testing the diameter of graphs. *Random Structures and Algorithms*, 20(2): 165–183, 2002.

[23] D. Ron. Property testing. In P. M. Pardalos, S. Rajasekaran, J. Reif, and J. D. P. Rolim, editors, *Handobook of Randomized Algorithms*, volume II, pp. 597–649. Kluwer Academic Publishers, 2001.

[24] A. Sinclair. Improved bounds for mixing rates of Markov chains and multicommodity flow. *Combinatorics, Probability and Computing*, 1: 351-370, 1992.

[25] A. Sinclair and M. Jerrum. Approximate counting, uniform generation and rapidly mixing Markov chains. *Information and Computation*, 82(1): 93–133, July 1989.