*Combinatorics, Probability and Computing* (2010) 19, 693–709. © Cambridge University Press 2010
doi:10.1017/S096354831000012X

# Testing Expansion in Bounded-Degree Graphs

ARTUR CZUMAJ[1] and CHRISTIAN SOHLER[2†‡]

[1]Department of Computer Science and Centre for Discrete Mathematics and its Applications (DIMAP), University of Warwick, Coventry CV4 7AL, UK
(e-mail: A.Czumaj@warwick.ac.uk)

[2]Department of Computer Science, Technische Universität Dortmund, D-44221 Dortmund, Germany
(e-mail: christian.sohler@tu-dortmund.de)

*Received 29 May 2009; revised 19 March 2010; first published online 9 June 2010*

We consider the problem of *testing expansion in bounded-degree graphs*. We focus on the notion of *vertex expansion*: an $\alpha$-expander is a graph $G=(V,E)$ in which every subset $U \subseteq V$ of at most $|V|/2$ vertices has a neighbourhood 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 $\widetilde{O}(\sqrt{n})$. We prove that the property-testing algorithm proposed by Goldreich and Ron with appropriately set parameters accepts every $\alpha$-expander with probability at least $\frac{2}{3}$ and rejects every graph that is $\varepsilon$-far from any $\alpha^*$-expander with probability at least $\frac{2}{3}$, where $\alpha^* = \Theta(\frac{\alpha^2}{d^2 \log(n/\varepsilon)})$ 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 $O(\frac{d^2\sqrt{n}\log(n/\varepsilon)}{\alpha^2\varepsilon^3})$.

## 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 is 'far' from satisfying the property. The notion of being far from a property is parametrized by a distance parameter $\varepsilon$. An object is $\varepsilon$-far from having a property $\Pi$ if it differs in more than an $\varepsilon$-fraction of its description from any object having the property $\Pi$. For example, when the object is a (dense) graph represented by an

† Research supported in part by DFG grant Me 872/8-3, by EPSRC grant EP/G064679/1, by EPSRC Science and Innovation Award EP/D063191/1 and the Centre for Discrete Mathematics and its Applications (DIMAP), University of Warwick.

‡ A preliminary version of this paper appeared in *Proc. 48th Annual IEEE Symposium on Foundations of Computer Science* (FOCS'07), pp. 570–578 (2007).

https://doi.org/10.1017/S096354831000012X Published online by Cambridge University Press

adjacency matrix and the considered property is bipartiteness, then a graph is $\varepsilon$-far from bipartite if one has to change more than $\varepsilon n^2$ entries in the adjacency matrix to make it bipartite.

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 $\varepsilon$-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 references, see the surveys [12], [13], [21], [25]). In the area of graph properties, we have recently seen a vast number of works dealing with properties of *dense graphs* represented by adjacency matrices. Following a series of results for specific problems, very powerful general results characterizing testability of entire classes of graph properties have recently 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 [9]; see also [2], which gives a certain characterization for testability in the dense graphs model). These results imply that, in the adjacency matrix model, most 'natural' graph properties can be tested with a constant number of queries.

Property testing in sparse graphs is less well understood, although significant progress has been made recently. To study properties of sparse graphs one typically assumes that the input graph $G$ is given in the *bounded-degree graph model* introduced by Goldreich and Ron [17]. In this model, $G$ 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 $i$th neighbour of a given vertex $v$. The number of accesses to the adjacency list is the query complexity of the tester. The task of a property-testing algorithm is to distinguish between graphs that satisfy a predetermined graph property $\Pi$ and graphs that are *$\varepsilon$-far* from property $\Pi$; here we say that a graph $G$ is $\varepsilon$-far from $\Pi$ if one has to modify more than $\varepsilon dn$ edges in $G$ to obtain a graph having property $\Pi$.

The first results for the bounded-degree graph model were obtained by Goldreich and Ron [17], who showed that a number of fundamental graph properties, such as connectivity, $k$-connectivity, cycle-freeness and being Eulerian, are testable in constant time. In contrast, they also prove that testing bipartiteness and testing expansion require $\Omega(\sqrt{n})$ time. Numerous other properties are known to require more than a constant number of queries: for example, acyclicity requires $\Omega(n^{1/3})$ queries [6] and testing 3-colourability requires $\Omega(n)$ queries [8]. For some of these properties one can achieve a sublinear query complexity: for example, bipartiteness is testable in $\widetilde{O}(\sqrt{n})$ query complexity for constant $\varepsilon$ using a combination of random sampling and random walks on the input graph.

https://doi.org/10.1017/S096354831000012X Published online by Cambridge University Press

Another approach to property testing in bounded-degree graphs has been proposed recently in [10], where the authors assumed that the input graph comes from a hereditary class of graphs that does not contain expanders, *e.g.*, from the class of planar graphs. Under this assumption, it is possible to show that every hereditary graph property is testable with a constant number of queries. The results from [10] have been very recently extended by Benjamini, Schramm and Shapira [7], who proved that any minor-closed property (*e.g.*, the planarity) of bounded-degree graphs can be tested in constant time.

In many of these property-testing algorithms a crucial role is played by the expansion of the graphs at hand. For example, in [17], the expansion properties of the graph are critical to 'controlling' the analysis of the random walks, and in [10], 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 [17]. Goldreich and Ron [16] made a major attempt to design a sublinear-time algorithm for testing expansion; however, in contrast to our paper, they considered the *algebraic characterization of expanders*. In contrast to the 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 $\lambda$, $0 < \lambda < 1$, has query complexity of $\widetilde{O}(\sqrt{n})$ and (with probability at least $\frac{2}{3}$) accepts every graph whose (normalized) second-largest eigenvalue satisfies $\lambda(G) \leqslant \lambda$, and rejects every graph that is $\varepsilon$-far from any graph $G'$ with $\lambda(G') \leqslant \lambda^*$, where $\lambda^* \gg \lambda$ but possibly $\lambda^* \leqslant \lambda^{O(1)}$. Goldreich and Ron [16] give a property-testing algorithm, and show that, assuming a certain appealing conjecture, their algorithm is a property tester for expansion.

### 1.1. Our result

In this paper we study the problem of testing expanders, but unlike Goldreich and Ron [16], we consider the combinatorial notion of *vertex expansion*. A graph $G = (V, E)$ is called an *$\alpha$-expander* (or $\alpha$-vertex-expander) if, for every $U \subseteq V$ with $|U| \leqslant n/2$, the neighbourhood of $U$ in $G$ has size at least $\alpha \cdot |U|$.[1] We reconsider the algorithm of Goldreich and Ron [16] and prove that when run with appropriate parameters, this algorithm distinguishes between good expanders and graphs far from weak expanders with $\widetilde{O}(\sqrt{n})$ queries. In particular, for any $\alpha > 0$, the algorithm has query complexity of $O(\frac{d^2 \cdot \ln(n/\varepsilon) \cdot \sqrt{n}}{\alpha^2 \cdot \varepsilon^3})$, and, with probability $\frac{2}{3}$, it distinguishes between an $\alpha$-expander and a graph that is $\varepsilon$-far from any $\alpha^*$-expander, where $\alpha^* = O(\frac{\alpha^2}{d^2 \log(n/\varepsilon)})$.

Our property tester uses the algorithm proposed by Goldreich and Ron [16] with appropriately set parameters, and some parts of our analysis follow the ideas from that

[1] It is known that the notions of vertex expansion and algebraic expansions are very closely related. In particular, it is known that if a $d$-regular graph $G$ is an $\alpha$-expander then $\lambda(G) \leqslant d - \frac{\alpha^2}{4+2\alpha^2}$ [1], and if $\lambda(G) = \lambda$ then $G$ is an $\alpha$-expander with $\alpha = \frac{d-\lambda}{2d}$ [3]. Observe that this relation is less tight when $\alpha = o(1)$. (See [18] for a more detailed discussion about this relationship.)

https://doi.org/10.1017/S096354831000012X Published online by Cambridge University Press

paper. Thus 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 [16], and also the most challenging part of our analysis, is a proof that if a graph is far from any (sufficiently good) expander then the algorithm will reject it. The main reason why this task is non-trivial is that, whereas we understand properties of expanders quite well (see, *e.g.*, the survey exposition in [18]), 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, a 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.

Our property-testing algorithm for vertex expansions can be easily extended to *edge expansions* (see [18]), 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 [16].

Let us notice that since the relation between $\alpha$ and $\alpha^*$ depends (weakly) on $n$, our result does not imply the conjecture of [16].

### 1.2. Subsequent works

There have been three subsequent developments after a preliminary version of this paper was made available. Independently of our paper, Kale and Seshadri [19] obtained an incomparable result for testing expansion using algebraic techniques. They proved that the algorithm of Goldreich and Ron in [16] rejects every graph that is $\Theta(\alpha^2/d)$-far from every expander with degree at most $2d$. While this result improves the bound for the expansion, it sacrifices the bound for the maximum degree from $d$ to $2d$. The revised version of that paper [20] then reduces the degree bound to $d$, leading to an improvement of the result presented in this paper. A similar improvement has been achieved independently by Nachmias and Shapira [23], who combine the techniques from [16], the conference version of this paper, and [19]. The two latter results prove the result conjectured in [16] in the sense that there is no dependence on $n$ required in the expansion gap of the tester.

## 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 of $G$ 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.*, $\deg(v) \leqslant d$ 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 $i$th neighbour of $v$ or a special symbol $+$ if $v$ has fewer than $i$ neighbours. In this paper we will assume $d \geqslant 4$.

https://doi.org/10.1017/S096354831000012X Published online by Cambridge University Press

### 2.1. Expanders and graphs that are far from expanders

For a detailed exposition of expanders, their properties, and their applications, we refer, *e.g.*, to a recent survey by Hoory, Linial and Wigderson [18].

In this paper, our main focus is on expanders defined through the notion of *vertex expansion*. For any two sets of vertices $U, W \subseteq V$, we denote by $N_G(U, W)$ the *neighbourhood 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 1.** Let $\alpha > 0$. A graph $G = (V, E)$ is an *$\alpha$-expander* if, for every $U \subseteq V$ with $|U| \leqslant n/2$, we have $|N(U, V)| \geqslant \alpha \cdot |U|$.

Let us recall 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 $\varepsilon$-far from expanders.

**Definition 2.** A graph $G$ is *$\varepsilon$-far from any $\alpha$-expander* if one has to modify more than $\varepsilon dn$ entries in $f_G$ to obtain an $\alpha$-expander.

Equivalently, $G$ is $\varepsilon$-far from any $\alpha$-expander if one has to modify more than $\varepsilon dn/2$ edges to obtain a graph $G'$ that has maximum degree at most $d$ and that is an $\alpha$-expander.

#### 2.1.1. Strong expanders and parameter $\beta_d$.

We will also use the notion of strong expanders.

**Definition 3.** A graph $G = (V, E)$ is called *$\beta$-strong expander* if, for every $U \subseteq V$, we have $|N(U, V)| \geqslant \beta \cdot \min\{|U|, |V \setminus U|\}$.

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

$$|N(U, V)| \geqslant \frac{|N(V \setminus U, V)|}{d^*} \geqslant \frac{\alpha}{d^*} |V \setminus U|.$$

Let $\beta_d^*$ be the largest number for which there is a $d$-regular $\beta_d^*$-strong expander. It is not difficult to prove that for every $d \geqslant 3$ we have $\beta_d^* \geqslant c > 0$ for a certain constant $c$ (*e.g.*, by considering a random $d$-regular graph [14, 18]), and therefore, from now on, we will use $\beta_d = \min\{\beta_d^*, \frac{1}{3}\} = \Theta(1)$, meaning that for every $d \geqslant 3$ there is a $\beta_d$-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 any two vertices $u, v \in V$, we let $P(u, v)$ denote the probability of moving 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 connected graph the random

https://doi.org/10.1017/S096354831000012X Published online by Cambridge University Press

walk converges to a uniform distribution). This will be achieved by a standard trick of adding (*virtually*, without modifying the graph) to each vertex $v$, $2d - \deg(v)$ 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 analyse 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 [22]. Let $\mathfrak{M}$ be a Markov chain with finite state space $\Omega$ 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 4.** The *variation distance* with respect to the initial state $x$ is defined by

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

In our analysis we will also frequently use an extension of Definition 4, which is the notion of the variation distance between two arbitrary probability distributions.

**Definition 5.** The *variation distance* between two probability distributions $\mathcal{X}$ and $\mathcal{Y}$ over the same finite domain $\Omega$ is defined by

$$d_{TV}(\mathcal{X},\mathcal{Y}) = \max_{S\subseteq\Omega} |\mathbf{Pr}_{\mathcal{X}}[S] - \mathbf{Pr}_{\mathcal{Y}}[S]| = \tfrac{1}{2} \sum_{\omega\in\Omega} |\mathbf{Pr}_{\mathcal{X}}[\omega] - \mathbf{Pr}_{\mathcal{Y}}[\omega]|.$$

**Definition 6.** The *rate of convergence* of $\mathfrak{M}$ with initial state $x$ to the stationary distribution is defined by

$$\tau_x(\zeta) = \min\{t : \Delta_x(t') \leqslant \zeta \quad \text{for all } t' \geqslant t\}.$$

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

**Definition 7 (Conductance [27]).** The *conductance* of a Markov chain $\mathfrak{M}$ is defined by

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

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 2.1 ([26]).** *Let $\mathfrak{M}$ be a finite, reversible, ergodic Markov chain with loop probabilities $P(x,x) \geqslant \frac{1}{2}$ for all states $x$. Let $\Phi$ be the conductance of $\mathfrak{M}$. Then the mixing time*

https://doi.org/10.1017/S096354831000012X Published online by Cambridge University Press

*of $\mathfrak{M}$ satisfies*

$$\tau_x(\zeta) \leqslant 2\Phi^{-2}\big(\ln\left(\pi_x^{-1}\right) + \ln(\zeta^{-1})\big),$$

*for any choice of initial state $x$.*

The previously defined random walk on $G$ can 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 $P(x,x) \geqslant \frac{1}{2}$ for all $x \in V$, and the stationary distribution of $\mathfrak{M}_G$ is uniform. Further, the conductance $\Phi_G$ of $\mathfrak{M}_G$ is given by

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

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

Notice also that for an $\alpha$-expander we have

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

## 3. Testing expanders

In this section we present a property-testing algorithm for expanders and state our main theorem which describes its performance. The algorithm is essentially the algorithm proposed by Goldreich and Ron [16] for testing expanders with the parameters set appropriately to facilitate our analysis.

```
ExpansionTester(G, ℓ, m, s)
repeat s times
    Select a vertex v uniformly at random from V
    Perform m independent 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 · (m choose 2)
        then reject
accept
```

The query complexity of the algorithm is $O(\ell ms)$. In the next section, we will prove the following main result of this paper.

**Theorem 3.1.** *Let $0 \leqslant \varepsilon \leqslant 0.0017$. Algorithm* ExpansionTester *with $s \geqslant \frac{48}{\varepsilon}$, $m \geqslant \frac{12 \cdot s \cdot \sqrt{n}}{\varepsilon^2}$, and $\ell \geqslant \frac{16 \cdot d^2 \cdot \ln(n/\varepsilon)}{\alpha^2}$ accepts every $\alpha$-expander with probability at least $\frac{2}{3}$ and rejects every graph that is $\varepsilon$-far from any $\frac{c \cdot \alpha^2}{d^2 \cdot \ln(n/\varepsilon)}$-expander with probability at least $\frac{2}{3}$, where $c > 0$ is a sufficiently large constant. The query complexity of this algorithm is $O(\frac{d^2 \cdot \ln(n/\varepsilon) \cdot \sqrt{n}}{\alpha^2 \cdot \varepsilon^3})$.*

https://doi.org/10.1017/S096354831000012X Published online by Cambridge University Press

In what follows we use $\alpha^* = \frac{c\cdot\alpha^2}{d^2\cdot\ln(n/\varepsilon)}$, and thus our goal is to show that algorithm ExpansionTester accepts every $\alpha$-expander with probability at least $\frac{2}{3}$ and rejects every graph that is $\varepsilon$-far from any $\alpha^*$-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 [16]. The length of the random walk is chosen in such a way that, for an $\alpha$-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 every $\alpha^*$-expander (with $\alpha^* = \frac{c\cdot\alpha^2}{d^2\cdot\ln(n/\varepsilon)}$) then the rejection will most likely occur.

It has already been proved in [16] (see the proof of Lemma 1 in [16] and also [5]) that the variance of this process is relatively small.

**Lemma 4.1 (Goldreich and Ron [16]).** *Let $X_v$ be the random variable for the number of collisions among the endpoints of $m$ random walks of length $\ell$ starting from vertex $v$. Let $P_v^\ell$ denote the distribution of the endpoint of this random walk. Then the following holds:*

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

With this result, we can follow the arguments of Goldreich and Ron [16] to show that an $\alpha$-expander is accepted provided that the random walks are sufficiently long. Let us consider an $\alpha$-expander and a random walk of length

$$\ell = \frac{16d^2 \cdot \ln(n/\varepsilon)}{\alpha^2} \geqslant \frac{8d^2 \cdot (\ln(n) + \ln(n/\varepsilon))}{\alpha^2}.$$

By applying Proposition 2.1 we observe that for any starting vertex $v \in V$, the variation distance between $P_v^\ell$ and the uniform distribution is at most $\varepsilon/n$. This yields $\|P_v^\ell\|_2^2 \leqslant (1+\varepsilon)^2/n$. By combining this bound with Lemma 4.1 and Chebyshev's inequality, we can obtain the following lemma (a similar lemma is proved for the algebraic notion of expansion in [16]).

**Lemma 4.2 (Accepting expanders).** *Let*

$$0 < \varepsilon < 1, \quad m \geqslant \frac{12 \cdot s \cdot \sqrt{n}}{\varepsilon^2} \quad \textit{and} \quad \ell \geqslant \frac{8d^2 \cdot (\ln(n) + \ln(n/\varepsilon))}{\alpha^2}.$$

*Then Algorithm* ExpansionTester *accepts every $\alpha$-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 $\ell$ starting at vertex $v$. By combining Lemma 4.1 with Chebyshev's inequality, we

https://doi.org/10.1017/S096354831000012X Published online by Cambridge University Press

obtain

$$\mathbf{Pr}[|X_v - \mathbf{E}[X_v]| \geqslant \varepsilon \cdot \mathbf{E}[X_v]] \leqslant \frac{\mathbf{Var}[X_v]}{\varepsilon^2 \cdot \mathbf{E}[X_v]^2} \leqslant \frac{4}{\varepsilon^2 \cdot m \cdot \|P_v^\ell\|_2} \leqslant \frac{4 \cdot \sqrt{n}}{\varepsilon^2 \cdot m},$$

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

$$m \geqslant \frac{12 \cdot s \cdot \sqrt{n}}{\varepsilon^2},$$

we obtain

$$\mathbf{Pr}[|X_v - \mathbf{E}[X_v]| \geqslant \varepsilon \cdot \mathbf{E}[X_v]] \leqslant \tfrac{1}{3s}.$$

Therefore, by the union bound we obtain

$$\mathbf{Pr}[\forall v \in S : |X_v - \mathbf{E}[X_v]| \leqslant \varepsilon \cdot \mathbf{E}[X_v]] \geqslant \tfrac{2}{3}.$$

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

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

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

**Lemma 4.3 (Rejections).** *Let $0 < \varepsilon < 0.1$. Let $G = (V, E)$ be a graph, and for each $v \in V$ let $P_v^\ell$ be the distribution of the endpoint of an $\ell$-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 $P_u^\ell$ to the uniform distribution is at least $6\sqrt{\varepsilon}$, then algorithm* ExpansionTester *run with $s \geqslant \frac{2}{\delta}$ rejects with probability at least $\frac{2}{3}$.*

**Proof.** We first show that any distribution with variation distance at least $6\sqrt{\varepsilon}$ to the uniform distribution will result in a high expected number of collisions for our random walks. If $\mathcal{U}$ denotes the uniform distribution, then the variation distance between $P_v^\ell$ 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 it is to distinguish the corresponding distribution from the uniform distribution, which minimizes this expected value. Thus, we are looking for a probability vector $P_v^\ell$ with $\|P_v^\ell - \mathcal{U}\|_1 \geqslant 3\sqrt{\varepsilon}$ that minimizes $\|P_v^\ell\|_2^2$. We claim that this is attained by a vector with $\frac{n}{2}$ elements with values $\frac{1+3\sqrt{\varepsilon}}{n}$ and $\frac{n}{2}$ elements with values $\frac{1-3\sqrt{\varepsilon}}{n}$, in which case $\|P_v^\ell\|_2^2 = \frac{1+9\varepsilon}{n}$.

To see this, let us first observe that any $P_v^\ell$ with $\|P_v^\ell - \mathcal{U}\|_1 \geqslant 3\sqrt{\varepsilon}$ that minimizes $\|P_v^\ell\|_2^2$ will satisfy $\|P_v^\ell - \mathcal{U}\|_1 = 3\sqrt{\varepsilon}$. Next, let us consider any vector $P_v^\ell = (p_1, \ldots, p_n)$ with $p_i \geqslant p_{i+1} \geqslant 0$ for every $1 \leqslant i < n$, and $\sum_i p_i = 1$. Let $t$ be such that $p_t \geqslant \frac{1}{n} > p_{t+1}$. Since we want to consider $P_v^\ell$ with $\|P_v^\ell - \mathcal{U}\|_1 = 3\sqrt{\varepsilon}$, we have

$$\sum_{i=1}^{t}\left(p_i - \frac{1}{n}\right) = \sum_{i=t+1}^{n}\left(\frac{1}{n} - p_i\right) = 1.5\sqrt{\varepsilon}.$$

https://doi.org/10.1017/S096354831000012X Published online by Cambridge University Press

Conditioned on $\sum_{i=1}^{t}(p_i - \frac{1}{n}) = 1.5\sqrt{\varepsilon}$, $\sum_{i=1}^{t} p_i^2$ is minimized when all $p_i$ (with $1 \leqslant i \leqslant t$) are identical, that is, when $p_i = \frac{1}{n} + \frac{1.5\sqrt{\varepsilon}}{t}$ for $i \leqslant t$. Similarly, conditioned on $\sum_{i=t+1}^{n}(\frac{1}{n} - p_i) = 1.5\sqrt{\varepsilon}$, $\sum_{i=t+1}^{n} p_i^2$ is minimized when $p_i = \frac{1}{n} - \frac{1.5\sqrt{\varepsilon}}{n-t}$ for all $i > t$. Hence, assuming that vector $P_v^\ell$ has $t$ elements greater than or equal to $\frac{1}{n}$, under the conditions above $\|P_v^\ell\|_2^2$ is minimized when $t$ elements have value $\frac{1}{n} + \frac{1.5\sqrt{\varepsilon}}{t}$ and the other $n-t$ elements have value $\frac{1}{n} - \frac{1.5\sqrt{\varepsilon}}{n-t}$, in which case

$$\|P_v^\ell\|_2^2 = \frac{1}{n} + \frac{9\varepsilon}{4}\left(\frac{1}{t} + \frac{1}{n-t}\right).$$

Since the last term is minimized when $t = n - t$, we conclude that

$$\|P_v^\ell\|_2^2 \geqslant \frac{1+9\varepsilon}{n}.$$

Next, we know from the proof of Lemma 4.2 that the observed number of collisions is, with probability at least $1 - \frac{1}{3s}$, at least $(1-\varepsilon)$ times its expectation. Thus, with probability at least $1 - \frac{1}{3s}$, we have that the number of collisions is greater than

$$(1-\varepsilon)\cdot\binom{m}{2}\cdot\|P_v^\ell\|_2^2 \geqslant \frac{(1-\varepsilon)(1+9\varepsilon)}{n}\cdot\binom{m}{2} > \frac{1+7\varepsilon}{n}\cdot\binom{m}{2},$$

where in the last inequality we used the fact that $\varepsilon \leqslant 0.1$.

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, by our analysis above, the overall probability of the rejection is at least

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

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

### 4.1. Being far from expanders

The main contribution of this paper is a proof that a graph that is $\varepsilon$-far from any $\alpha^*$-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 $\varepsilon$-far from any $\alpha^*$-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 $\beta_d$ defined in Section 2.1.1.)

**Lemma 4.4.** *Let $0 < \varepsilon < 1$ and $\alpha^* \leqslant \frac{1}{10}$. If $G=(V,E)$ has a subset of vertices $A \subseteq V$ with $|A| \leqslant \frac{1}{12}\varepsilon n$ such that $G[V \setminus A]$ is an $\frac{4\alpha^*}{\beta_{d-1}}$-expander, then $G$ is not $\varepsilon$-far from any $\alpha^*$-expander.*

**Proof.** To prove the lemma, we show that if there exists $A \subseteq V$ with $|A| \leqslant \frac{1}{12}\varepsilon n$ such that $G[V \setminus A]$ is an $\frac{4\alpha^*}{\beta_{d-1}}$-expander, then we can modify $\frac{1}{2}\varepsilon dn$ edges in $G$ to obtain a graph $G^*$ that has maximum degree at most $d$ and is an $\alpha^*$-expander. Since each edge appears in the adjacency list of two vertices, we can obtain $G^*$ by changing at most $\varepsilon dn$ entries in the functional representation of $G$. Hence, $G$ is not $\varepsilon$-far from any $\alpha^*$-expander.

We first argue that we can restrict our attention to the case $\varepsilon \leqslant \frac{9}{d}$. To see this, let us recall that there are 3-regular expanders (*e.g.*, Ramanujan graphs) where the

https://doi.org/10.1017/S096354831000012X Published online by Cambridge University Press

second-largest eigenvalue $\lambda$ is at most $2\sqrt{2}+\varepsilon$ for every positive $\varepsilon$ (see, *e.g.*, [14] or [18]). Then, the relation between $\lambda$ and vertex expansion [3] implies that such graphs have vertex expansion at least 0.1. With this, we can modify any graph $G$ of maximum degree at most $d$, $d \geqslant 3$, into a $\frac{1}{10}$-expander of maximum degree at most $d$ by first removing at most $3n$ edges (to ensure that every vertex is of maximum degree at most $d-3$) and then adding at most $\frac{3}{2}n$ edges of a 3-regular $\frac{1}{3}$-expander. Therefore, since $\alpha^* \leqslant \frac{1}{10}$, if $\varepsilon > \frac{9}{d}$, then any graph of maximum degree at most $d$ can be transformed into an $\alpha^*$-expander by modifying at most $\frac{1}{2}\varepsilon dn$ edges. Thus, if $\varepsilon > \frac{9}{d}$ then no graph of maximum degree at most $d$ is $\varepsilon$-far from any $\alpha^*$-expander. Hence, we have to consider only the case $\varepsilon \leqslant \frac{9}{d}$.

Let us now focus on the case $\varepsilon \leqslant \frac{9}{d}$. For simplicity of the 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 $\beta_{d-1}$-strong expander on $A$.
(3) Remove an arbitrary matching $M$ of size $|A|/2$ from $G[V \setminus A]$; let $V_M$ denote the set of endpoints of the edges in $M$.
(4) Add an arbitrary matching between $A$ and $V_M$.

Before we proceed, let us first observe that $G[V \setminus A]$ always has a matching of size at least $|A|/2$, and hence step (3) in our construction can always be performed. Indeed, since $G[V \setminus A]$ is a $\frac{4\alpha^*}{\beta_{d-1}}$-expander, $G[V \setminus A]$ must have at least $|V \setminus A| - 1$ edges. Since $G[V \setminus A]$ is of maximum degree at most $d$, a greedy matching algorithm (choose a spanning tree $T$ of $G[V \setminus A]$, choose an arbitrary edge $e$ incident to a leave of $T$, add $e$ to the matching, and remove all edges adjacent to $e$) will find in $G[V \setminus A]$ a matching of size at least $\frac{|V \setminus A|-1}{d}$.

Now, basic algebraic manipulations using the assumption that $|A| \leqslant \frac{1}{12}\varepsilon n$, $d \geqslant 4$, and $\varepsilon \leqslant \frac{9}{d}$, give us the following lower bound for the size of the greedy matching:

$$\max\left\{1, \frac{|V \setminus A|-1}{d}\right\} \geqslant \frac{|V \setminus A|}{2d} \geqslant \frac{n(1-\frac{1}{12}\varepsilon)}{2d} \geqslant \tfrac{1}{24}\varepsilon n \geqslant |A|/2.$$

Now, once we see that the construction above can always be performed, let us observe that if $G$ has maximum degree at most $d$ then so has $G^*$, and that $G^*$ is obtained from $G$ by modifying at most $d|A| + (d-1)|A|/2 + |A|/2 + |A| = (d+1+d/2)|A| \leqslant 2d|A| \leqslant \frac{1}{2}\varepsilon dn$ edges in $G$. Therefore, now we only have to prove that $G^*$ is an $\alpha^*$-expander.

To show that $G^*$ is an $\alpha^*$-expander we have to prove that for any set $X \subseteq V$ with $|X| \leqslant n/2$, we have $|N_{G^*}(X, V)| \geqslant \alpha^*|X|$. Let us define $B = V \setminus A$ and fix an arbitrary set $X \subseteq V$ with $|X| \leqslant n/2$. Let $X_A = X \cap A$ and $X_B = X \cap (V \setminus A)$. The situation is depicted in Figure 1.

We will distinguish between two cases depending on whether $|X_A| \geqslant 2|X_B|$ or $|X_A| < 2|X_B|$.

**Case I:** $|X_A| \geqslant 2|X_B|$. We have

$$|N_{G^*}(X, V)| \geqslant |N_{G^*}(X_A, V \setminus X)| \geqslant |N_{G^*}(X_A, V \setminus X) \cap B|.$$

https://doi.org/10.1017/S096354831000012X Published online by Cambridge University Press

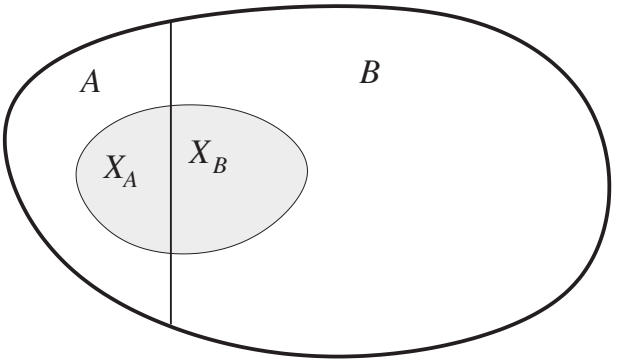



*Figure 1.* We consider an arbitrary set $X = X_A \cup X_B$ and prove that it is well connected to the rest of the graph.

Observe that $N_{G^*}(X_A, V \setminus X) \cap B$ is the set of neighbours of the vertices in $X_A$ that are contained in $B \setminus X_B$. Consider step (4) of our construction of graph $G^*$. In step (4) we insert into $G^*$ a matching between $A$ and $V_M$. Consider the edges that match the vertices of $X_A \subseteq A$ with $V_M \subseteq B$. By the matching property, every vertex from $V_M$ is matched with exactly one vertex in $A$. Therefore, there are at least $|X_A| - |X_B|$ such matching edges that connect $X_A$ with $V_M \setminus X_B$. This implies that

$$|N_{G^*}(X_A, V \setminus X) \cap B| \geqslant |X_A| - |X_B|.$$

Since we assumed that $|X_A| \geqslant 2|X_B|$, the inequality above yields the following:

$$|N_{G^*}(X_A, V \setminus X) \cap B| \geqslant |X_A| - |X_B| \geqslant \tfrac{1}{2}|X_A| \geqslant \tfrac{1}{3}(|X_A| + |X_B|) = \tfrac{1}{3}|X| \geqslant \alpha^*|X|.$$

This proves the first case.

**Case II:** $|X_A| < 2|X_B|$. To deal with the case $|X_A| < 2|X_B|$, we distinguish between the subcase that

$$\min\{|X_A|, |A \setminus X_A|\} > \frac{3\alpha^*|X_B|}{\beta_{d-1}} \quad \text{and} \quad \min\{|X_A|, |A \setminus X_A|\} \leqslant \frac{3\alpha^*|X_B|}{\beta_{d-1}}.$$

If $\min\{|X_A|, |A \setminus X_A|\} > \frac{3\alpha^*|X_B|}{\beta_{d-1}}$, then we will exploit the fact that $X$ is well connected to the remainder of the graph. Since in step (2) of our construction of $G^*$ we add a $\beta_{d-1}$-strong expander on the set $A$, we know that every subset $Z \subseteq A$ has at least $\beta_{d-1} \cdot \min\{|Z|, |A \setminus Z|\}$ neighbours in $A$. Hence,

$$|N_{G^*}(X_A, V \setminus X) \cap A| \geqslant \beta_{d-1} \cdot \min\{|X_A|, |A \setminus X_A|\}.$$

Now we use our assumption that $\min\{|X_A|, |A \setminus X_A|\} > \frac{3\alpha^*|X_B|}{\beta_{d-1}}$ to obtain

$$\beta_{d-1} \cdot \min\{|X_A|, |A \setminus X_A|\} \geqslant \beta_{d-1} \cdot \frac{3\alpha^*|X_B|}{\beta_{d-1}} = 3\alpha^*|X_B| \geqslant \alpha^*|X|,$$

https://doi.org/10.1017/S096354831000012X Published online by Cambridge University Press

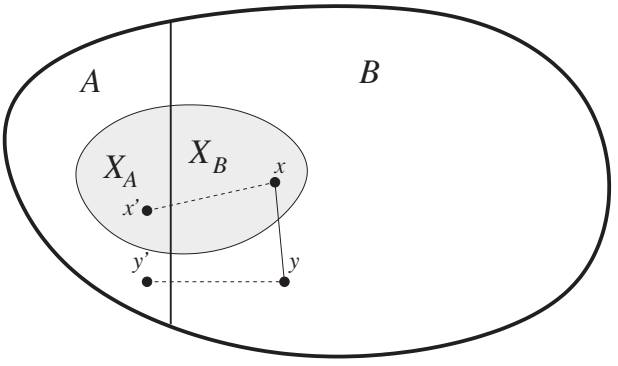



*Figure 2.* Let $(x, y) \in M$ be an edge with $x \in X_B$ and $y \in B \setminus X_B$. The only case in which we do not have a substitution for $(x, y)$ after its removal in step (3) and the addition of an arbitrary matching in step (4) is when $x' \in X_A$ and $y \in A \setminus X_A$.

where the last inequality follows from our assumption that $|X_A| < 2|X_B|$. Now we combine these two bounds to obtain the required inequality:

$$|N_{G^*}(X, V)| \geqslant |N_{G^*}(X_A, V \setminus X) \cap A| \geqslant \beta_{d-1} \cdot \min\{|X_A|, |A \setminus X_A|\} \geqslant \alpha^* |X|.$$

To analyse the other subcase, where $\min\{|X_A|, |A \setminus X_A|\} \leqslant \frac{3\alpha^*|X_B|}{\beta_{d-1}}$, we first establish the following claim.

**Claim 4.5.**

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

**Proof.** Let us first observe that in $G$, since $G[V \setminus A]$ is an $\frac{4\alpha^*}{\beta_{d-1}}$-expander, $X_B$ has at least $\frac{4\alpha^*|X_B|}{\beta_{d-1}}$ neighbours in $V \setminus (A \cup X_B)$. Unfortunately, step (3) of our construction of $G^*$ removes a matching from $X_B$. We have to argue that after the addition of an arbitrary matching in step (4) there are still enough edges connecting $X$ to $V \setminus X$.

To prove this, we will argue that during the construction of $G^*$ most of the edges either remain or are replaced by other edges connecting $X_B$ with $V \setminus X$. Let us consider an arbitrary edge $(x, y) \in M$ with $x \in X_B$ and $y \notin X_B$. Note that by our construction we have $y \in B \setminus X_B$. The situation is depicted in Figure 2. In step (3) of the construction of $G^*$, after removing edge $(x, y)$, in step (4) we match $x$ and $y$ with some vertices $x', y'$ from $A$, respectively. We observe that if $x' \in A \setminus X_A$ then we get a new edge from $X$ to $V \setminus X$ in $G^*$ that compensates for the removal of $(x, y)$. Similarly, if $y' \in X_A$ we get a new edge between $X$ and $V \setminus X$. It is only if $x' \in X_A$ and $y' \in A \setminus X_A$ that we do not have a substitution for the edge $(x, y)$. But in this case, we match one vertex from $X_A$ and one vertex from $A \setminus X_A$. Hence, this case can occur at most $\min\{|X_A|, |A \setminus X_A|\}$ times. This implies

$$|N_{G^*}(X, V)| \geqslant \frac{4\alpha^*|X_B|}{\beta_{d-1}} - \min\{|X_A|, |A \setminus X_A|\},$$

which proves the claim. □

https://doi.org/10.1017/S096354831000012X Published online by Cambridge University Press

Now, we prove the remaining subcase, when $|X_A| < 2|X_B|$ and $\min\{|X_A|, |A \setminus X_A|\} \leqslant \frac{3\alpha^*|X_B|}{\beta_{d-1}}$. By combining Claim 4.5 with the bound

$$\min\{|X_A|, |A \setminus X_A|\} \leqslant \frac{3\alpha^*|X_B|}{\beta_{d-1}},$$

we obtain

$$|N_{G^*}(X,V)| \geqslant \frac{4\alpha^*|X_B|}{\beta_{d-1}} - \min\{|X_A|, |A \setminus X_A|\} \geqslant \frac{4\alpha^*|X_B|}{\beta_{d-1}} - \frac{3\alpha^*|X_B|}{\beta_{d-1}} = \frac{\alpha^*|X_B|}{\beta_{d-1}} \geqslant \alpha^*|X|,$$

where the last inequality uses the fact that $\beta_{d-1} \leqslant \frac{1}{3}$. This completes the proof of Lemma 4.4. □

Once we have Lemma 4.4, we can show that if a graph is far from any expander then there exists a small cut that separates a large set of vertices from the rest of the graph.

**Corollary 4.6.** *Let $G=(V,E)$ be $\varepsilon$-far from any $\alpha^*$-expander with $\alpha^* \leqslant \frac{1}{10}$. Then there is a subset of vertices $A \subseteq V$ with $\frac{1}{12}\varepsilon n \leqslant |A| \leqslant \frac{1}{2}(1+\varepsilon)n$ such that $|N_G(A,V)| < \frac{4\alpha^*}{\beta_{d-1}}|A|$.*

**Proof.** Let $G$ be a graph that is $\varepsilon$-far from any $\alpha^*$-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 $|N_G(A_2, V \setminus A)| \leqslant \frac{4\alpha^*}{\beta_{d-1}} \cdot |A_2|$. If $|A_1 \cup A_2| < \frac{1}{12}\varepsilon n$ 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 neighbourhood. We can repeat this process until $|\bigcup_i A_i| \geqslant \frac{1}{12}\varepsilon n$. Now, let us write $A = |\bigcup_i A_i|$. By our construction we have $|A| \leqslant \frac{1}{2}(1+\varepsilon)n$.

To see that $|N_G(A,V)| \leqslant \frac{4\alpha^*}{\beta_{d-1}} \cdot |A|$, we first observe that our construction ensures that for every $i$, we have $|N_G(A_i, V \setminus \bigcup_{j<i} A_j)| \leqslant \frac{4\alpha^*}{\beta_{d-1}} \cdot |A_i|$. Then we obtain

$$|N_G(A,V)| \leqslant \sum_i \left| N_G\left(A_i, V \setminus \bigcup_{j<i} A_j\right)\right| \leqslant \sum_i \frac{4\alpha^*}{\beta_{d-1}} \cdot |A_i| = \frac{4\alpha^*}{\beta_{d-1}} \cdot |A|. \qquad \square$$

### 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 $|A| \leqslant \frac{1}{2}(1+\varepsilon)n$ and $|N_G(A,V)| \leqslant \gamma|A|$ for $\gamma \leqslant \frac{1}{10(\ell+1)}$. 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 $G_A = (V_A, E_A)$ (recall that the random walk is defined on a $(2d)$-regular graph: for every vertex we add self-loops to obtain a $(2d)$-regular graph and then 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 $\ell$ steps. Let $Y_i$ denote the indicator random variable for the event that the $i$th

https://doi.org/10.1017/S096354831000012X Published online by Cambridge University Press

vertex of the random 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 $i$th 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

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

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

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

Let us now move to the random walk in $G$. The claim above implies that the probability that an $\ell$-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 $\gamma \leqslant \frac{1}{10(\ell+1)}$, 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 $|V \setminus A| \geqslant \frac{1}{2}(1 - \varepsilon)n$, in the uniform distribution, a vertex from the outside of $A$ must be chosen with the probability $\frac{|V \setminus A|}{|V|} \geqslant \frac{1-\varepsilon}{2}$, in contrast to the distribution of the endpoints of an $\ell$-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 $\ell$-step random walk starting in $U$ to the uniform distribution is at least $\frac{1-2\varepsilon}{4}$. We conclude the following.

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

### 4.3. Putting everything together: proof of Theorem 3.1

Now we are ready to conclude the proof of Theorem 3.1.

Let us define

$$\alpha^* = \frac{\beta_{d-1}}{40(\ell + 1)} = \Theta\left(\frac{\alpha^2}{d^2 \cdot \ln(n/\varepsilon)}\right),$$

where we use the fact that $\beta_{d-1} = \Theta(1)$. Notice that this ensures that $\alpha^* \leqslant \frac{1}{10}$. We will prove that if $G$ is an $\alpha$-expander then Algorithm ExpansionTester will accept $G$ with probability at least $\frac{2}{3}$ and if $G$ is $\varepsilon$-far from any $\alpha^*$-expander then it will be rejected with probability at least $\frac{2}{3}$.

https://doi.org/10.1017/S096354831000012X Published online by Cambridge University Press

The first claim follows directly from Lemma 4.2 and hence we only have to prove the latter claim. Let us consider a graph $G=(V,E)$ of maximum degree at most $d$ that is $\varepsilon$-far from any $\alpha^*$-expander. By Corollary 4.6, $G$ has a set of vertices $A$ with $\frac{1}{12}\varepsilon n \leqslant |A| \leqslant \frac{1}{2}(1+\varepsilon)n$ and $|N(A,V)| \leqslant \frac{4\alpha^*|A|}{\beta_{d-1}}$. Hence, by our setting of $\alpha^*$, we have $|N(A,V)| \leqslant \frac{|A|}{10(\ell+1)}$. Next, by Lemma 4.7, there exists a set $U$ of cardinality at least $|A|/2 \geqslant \frac{1}{24}\varepsilon n$ such that, for each vertex $v \in U$, $P_v^\ell$ has variation distance to the uniform distribution of at least $\frac{1-2\varepsilon}{4}$, which, in turn, is at least $6\sqrt{\varepsilon}$ for $\varepsilon < .0017$. Finally, we apply Lemma 4.3 to conclude that $G$ is rejected with probability at least $\frac{2}{3}$.

This implies the proof of Theorem 3.1. □

## 5. Conclusions

Since quite a few results on property testing in sparse graphs involve some notion of expansion, the understanding of the effect of expansion on property testing seems to be one of the key challenges on the path to understanding property testing in sparse graphs. In this paper, we make a major step toward the understanding of this relation by analysing a property-testing algorithm for expansion of graphs, which was proposed by Goldreich and Ron [16]. Goldreich and Ron [16] conjectured that their algorithm is indeed a property tester for expansion, and in this paper we have shown that (under a different, combinatorial notion of expansion and with a somewhat 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.

## Acknowledgements

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

## References

[1] Alon, N. (1986) Eigenvalues and expanders. *Combinatorica* **6** 83–96.
[2] Alon, N., Fischer, E., Newman, I. and Shapira, A. (2009) A combinatorial characterization of the testable graph properties: It's all about regularity. *SIAM J. Comput.* **39** 143–167.
[3] Alon, N. and Milman, V. D. (1985) $\lambda_1$, isoperimetric inequalities for graphs, and superconcentrators. *J. Combin. Theory Ser. B* **38** 73–88.
[4] Alon, N. and Shapira, A. (2008) A characterization of the (natural) graph properties testable with one-sided error. *SIAM J. Comput.* **37** 1703–1727.
[5] Batu, T., Fortnow, L., Rubinfeld, R., Smith, W. and White, P. (2000) Testing that distributions are close. In *Proc. 41st IEEE Symposium on Foundations of Computer Science* (FOCS), pp. 259–269.
[6] Bender, M. A. and Ron, D. (2002) Testing properties of directed graphs: Acyclicity and connectivity. *Random Struct. Alg.* **20** 184–205.
[7] Benjamini, I., Schramm, O. and Shapira, A. (2008) Every minor-closed property of sparse graphs is testable. In *Proc. 40th Annual ACM Symposium on Theory of Computing* (STOC), pp. 393–402.
[8] Bogdanov, A., Obata, K. and Trevisan, L. (2002) A lower bound for testing 3-colorability in bounded-degree graphs. In *Proc. 43rd IEEE Symposium on Foundations of Computer Science* (FOCS), pp. 93–102.

https://doi.org/10.1017/S096354831000012X Published online by Cambridge University Press

[9] Borgs, C., Chayes, J., Lovász, L., Sos, V. T., Szegedy, B. and Vesztergombi, K. (2006) Graph limits and parameter testing. In *Proc. 38th Annual ACM Symposium on Theory of Computing* (STOC), pp. 261–270.
[10] Czumaj, A., Shapira, A. and Sohler, C. (2009) Testing hereditary properties of non-expanding bounded-degree graphs. *SIAM J. Comput.* **38** 2499–2510.
[11] Czumaj, A. and Sohler, C. (2005) Abstract combinatorial programs and efficient property testers. *SIAM J. Comput.* **34** 580–615.
[12] Czumaj, A. and Sohler, C. (2006) Sublinear-time algorithms. *Bull. EATCS* **89** 23–47.
[13] Fischer, E. (2001) The art of uninformed decisions: A primer to property testing. *Bull. EATCS* **75** 97–126.
[14] Friedman, J. (2004) A proof of Alon's second eigenvalue conjecture. *Mem. Amer. Math. Soc.*, to appear.
[15] Goldreich, O. and Ron, D. (1999) A sublinear bipartiteness tester for bounded degree graphs. *Combinatorica* **19** 335–373.
[16] Goldreich, O. and Ron, D. (2000) On testing expansion in bounded-degree graphs. *Electronic Colloquium on Computational Complexity* (ECCC), Report no. 7.
[17] Goldreich, O. and Ron, D. (2002) Property testing in bounded degree graphs. *Algorithmica* **32** 302–343.
[18] Hoory, S., Linial, N. and Wigderson, A. (2006) Expander graphs and their applications. *Bull. Amer. Math. Soc.* **43** 439–561.
[19] Kale, S. and Seshadhri, C. (2007) Testing expansion in bounded degree graphs. *Electronic Colloquium on Computational Complexity* (ECCC) TR-07-076.
[20] Kale, S. and Seshadhri, C. (2008) An expansion tester for bounded degree graphs. *Proc. 35th Annual International Colloquium on Automata, Languages and Programming* (ICALP), pp. 527–538.
[21] Kumar, R. and Rubinfeld, R. (2003) Sublinear time algorithms. *SIGACT News* **34** 57–67.
[22] Jerrum, M. and Sinclair, A. (1996) The Markov chain Monte Carlo method: An approach to approximate counting and integration. Chapter 12 in *Approximation Algorithms for NP-hard Problems* (D. Hochbaum, ed.), PWS Publishing, Boston.
[23] Nachmias, A. and Shapira, A. (2007) Testing the expansion of a graph. *Electronic Colloquium on Computational Complexity* (ECCC) TR-07-118.
[24] Parnas, M. and Ron, D. (2002) Testing the diameter of graphs. *Random Struct. Alg.* **20** 165–183.
[25] Ron, D. (2001) Property testing. In *Handbook of Randomized Algorithms* (P. M. Pardalos, S. Rajasekaran, J. Reif, and J. D. P. Rolim, eds), Vol. II, Kluwer, pp. 597–649.
[26] Sinclair, A. (1992) Improved bounds for mixing rates of Markov chains and multicommodity flow. *Combin. Probab. Comput.* **1** 351–370.
[27] Sinclair, A. and Jerrum, M. (1989) Approximate counting, uniform generation and rapidly mixing Markov chains. *Inform. Comput.* **82** 93–133.

https://doi.org/10.1017/S096354831000012X Published online by Cambridge University Press