Testing expansion in bounded-degree graphs

Artur Czumaj and Christian Sohler

The 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'2007) 570–578.

Speaker: Joseph, Chuang-Chieh Lin
Supervisor: Professor Maw-Shang Chang

Computation Theory Laboratory
Department of Computer Science and Information Engineering
National Chung Cheng University, Taiwan

June 9, 2010

Outline


胸圍 :


baby


i ♥ daddy

Outline

Background on property testing

- Try to answer "yes" or "no" for the following *relaxed* decision problems by observing only a small fraction of the input.
  - Does the input **satisfy a designated property**, or
  - is ϵ**-far from satisfying the property**?

Background on property testing

- Try to answer "yes" or "no" for the following *relaxed* decision problems by observing only a small fraction of the input.
  - Does the input **satisfy a designated property**, or
  - is **ϵ-far from satisfying the property**?

Background on property testing (contd.)

- In property testing, we use ϵ-far to say that the input is far from a certain property.
- ϵ: the least fraction of the input needs to be modified.
- For example:
  - A sequence of integers $L = (0, 2, 3, 4, 1)$.
  - Allowed operations: integer deletions
  - L is 0.2-far from being monotonically nondecreasing.

Background on property testing (contd.)

- In property testing, we use ϵ-far to say that the input is far from a certain property.
- ϵ: the least fraction of the input needs to be modified.
- For example:
  - A sequence of integers $L = (0, 2, 3, 4, 1)$.
  - Allowed operations: integer deletions
  - L is 0.2-far from being monotonically nondecreasing.

The model for bounded-degree graphs

- Graph model: adjacency **list** for graphs with vertex-degree bounded by d.
  - It takes $O(1)$ time to access to a function $f_G : [n] \times [d] \mapsto [n] \times \{+\}$.
    - 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, $d \geq 4$.
- ϵ-far from satisfying a graph property $\mathbb{P}$:
  - one has to modify $> \epsilon dn$ entries in f_G (i.e., $> \epsilon dn/2$ edges) to make the input graph satisfy $\mathbb{P}$.

The model for bounded-degree graphs

- Graph model: adjacency **list** for graphs with vertex-degree bounded by d.
  - It takes $O(1)$ time to access to a function $f_G : [n] \times [d] \mapsto [n] \times \{+\}$.
    - 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, $d \geq 4$.
- ϵ-far from satisfying a graph property $\mathbb{P}$:
  - one has to modify $> \epsilon dn$ entries in f_G (i.e., $> \epsilon dn/2$ edges) to make the input graph satisfy $\mathbb{P}$.

Background on property testing (contd.)

- The complexity measure: queries.
- The query complexity (say $q(n, d, \epsilon)$) is asked to be sublinear in $|V| = n$.
  - $q(n, d, \epsilon) = o(f(n))$ if $\lim_{n \to \infty} \frac{q(n,d,\epsilon)}{f(n)} \to 0$, where ϵ and d are viewed as constants.

Background on property testing (contd.)

- The complexity measure: queries.
- The query complexity (say $q(n, d, \epsilon)$) is asked to be sublinear in $|V| = n$.
  - $q(n, d, \epsilon) = o(f(n))$ if $\lim_{n \to \infty} \frac{q(n,d,\epsilon)}{f(n)} \to 0$, where ϵ and d are viewed as constants.

Property testers

- A property tester for $\mathbb{P}$ is an algorithm utilizing sublinear queries such that:
  - if the input satisfies $\mathbb{P}$:
    answers "yes" with probability $\geq 2/3$ ($1 \rightarrow$ one-sided error);
  - if the input is ϵ-far from satisfying $\mathbb{P}$:
    answers "no" with probability $\geq 2/3$.

Background on property testing (contd.)

- Unlike testing graph properties in the adjacency matrix model, only a few, very simple graph properties are known to be testable (i.e., query complexity is independent of n).
- For most of nontrivial graph properties, super-constant lower bounds exist.
  - bipartiteness: $\Omega(\sqrt{n})$.
  - 3-colorability: $\Omega(n)$.
  - acyclicity (in directed graphs): $\Omega(n^{1/3})$.
  - ...
- The focus turned on property testers with sublinear query complexity.

Outline

Definition 2.1

Let $\alpha > 0$. A graph $G = (V, E)$ is an α-expander (The expansion of G is α) if for every $U \subseteq V$ with $|U| \leq n/2$, it holds that $N_G(U, V) \geq \alpha \cdot |U|$.

- For $U, W \subseteq V$, $N_G(U, W) = \{v \in W \setminus U : \exists u \in U \text{ such that } (v, u) \in E\}$.
- For example:
  - What is the expansion of K_n?
  - What is the expansion of C_n?
  - What is the lower bound on the expansion of a k-club with n vertices?
  - What is the lower bound on the expansion of an s-plex with n vertices?

Definition 2.1

Let $\alpha > 0$. A graph $G = (V, E)$ is an α-expander (The expansion of G is α) if for every $U \subseteq V$ with $|U| \leq n/2$, it holds that $N_G(U, V) \geq \alpha \cdot |U|$.

- For $U, W \subseteq V$,
  $N_G(U, W) = \{v \in W \setminus U : \exists u \in U \text{ such that } (v, u) \in E\}$.
- For example:
  - What is the expansion of K_n?
  - What is the expansion of C_n?
  - What is the lower bound on the expansion of a k-club with n vertices?
  - What is the lower bound on the expansion of an s-plex with n vertices?

Definition 2.1

Let $\alpha > 0$. A graph $G = (V, E)$ is an α-expander (The expansion of G is α) if for every $U \subseteq V$ with $|U| \leq n/2$, it holds that $N_G(U, V) \geq \alpha \cdot |U|$.

- For $U, W \subseteq V$,
  $N_G(U, W) = \{v \in W \setminus U : \exists u \in U \text{ such that } (v, u) \in E\}$.
- For example:
  - What is the expansion of K_n?
  - What is the expansion of C_n?
  - What is the lower bound on the expansion of a k-club with n vertices?
  - What is the lower bound on the expansion of an s-plex with n vertices?

Definition 2.1

Let $\alpha > 0$. A graph $G = (V, E)$ is an α-expander (The expansion of G is α) if for every $U \subseteq V$ with $|U| \leq n/2$, it holds that $N_G(U, V) \geq \alpha \cdot |U|$.

- For $U, W \subseteq V$, $N_G(U, W) = \{v \in W \setminus U : \exists u \in U \text{ such that } (v, u) \in E\}$.
- For example:
  - What is the expansion of K_n?
  - What is the expansion of C_n?
  - What is the lower bound on the expansion of a k-club with n vertices?
  - What is the lower bound on the expansion of an s-plex with n vertices?

α-expanders

Definition 2.1

Let $\alpha > 0$. A graph $G = (V, E)$ is an α-expander (The expansion of G is α) if for every $U \subseteq V$ with $|U| \leq n/2$, it holds that $N_G(U, V) \geq \alpha \cdot |U|$.

- For $U, W \subseteq V$,
  $N_G(U, W) = \{v \in W \setminus U : \exists u \in U \text{ such that } (v, u) \in E\}$.
- For example:
  - What is the expansion of K_n?
  - What is the expansion of C_n?
  - What is the lower bound on the expansion of a k-club with n vertices?
  - What is the lower bound on the expansion of an s-plex with n vertices?

A well-known fact

Theorem 2.2 (Planar Separator Theorem (Lipton & Tarjan 1979))

Every planar graph with n vertices (n is sufficiently large) has a subset of vertices A, where $\frac{1}{3}n \leq |A| \leq \frac{1}{2}n$*, such that* $N(A, V) \leq 4\sqrt{n}$.

- The expansion of a planar graph: $O(1/\sqrt{n})$.

Algebraic notion of graph expansion

- Let $A(G)$ be an $n \times n$ adjacency *matrix of a d-regular graph G.*
  - *Each entry (u, v) contains the number of edges in G between u and v.*
- *Since $A(G)$ is symmetric, $A(G)$ has n eigenvalues* $\mu_0 \geq \mu_1 \geq \ldots \geq \mu_{n-1}$.
- **Theorem:** *Let α be the expansion of G. Then $\mu_0 = d$ and*

$$\frac{d - \mu_1}{2} \leq \alpha \leq \sqrt{2d(d - \mu_1)}.$$

Related work on testing expansion

- Testing whether G is an α-expander: It's still OPEN.
  - Lower bound for testing expansion: $\Omega(\sqrt{n})$ [Goldreich & Ron 2002].

Conjecture (Goldreich & Ron 2000)

In the bounded-degree model, a property tester for testing if a graph G is an α-expander exists.

- The focus turned to the relaxed goal: distinguish between α-expanders and graphs that are ϵ-far from being an α'-expander ($\alpha' < \alpha$).

Related work on testing expansion (contd.)

To be concise, here we omit the factors of ϵ and d.

- Distinguishing between α-expanders and graphs far from being $\Theta(\frac{\alpha^2}{\log n})$-expanders (Czumaj & Sohler; FOCS'2007).
- Distinguishing between α-expanders and graphs far from being $\Omega(\alpha^2)$-expanders (Nachmias & Shapira; Information and Computation 2010)

Related work on testing expansion (contd.)

To be concise, here we omit the factors of ϵ and d.

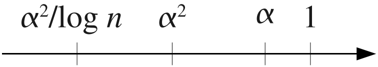
- Distinguishing between α-expanders and graphs far from being $\Theta(\frac{\alpha^2}{\log n})$-expanders (Czumaj & Sohler; FOCS'2007).

- Distinguishing between α-expanders and graphs far from being $\Omega(\alpha^2)$-expanders (Nachmias & Shapira; Information and Computation 2010)

Related work on testing expansion (contd.)

To be concise, here we omit the factors of ϵ and d.

- Distinguishing between α-expanders and graphs far from being $\Theta(\frac{\alpha^2}{\log n})$-expanders (Czumaj & Sohler; FOCS'2007).
- Distinguishing between α-expanders and graphs far from being $\Omega(\alpha^2)$-expanders (Nachmias & Shapira; Information and Computation 2010)

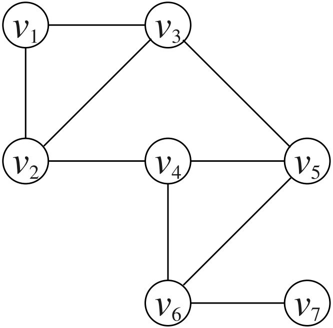


Outline

Before we proceed with the tester ...

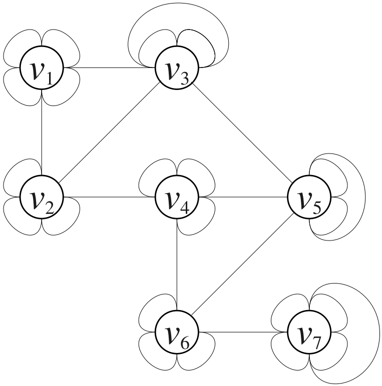
- For each vertex $v \in V$, we add $2d - \deg(v)$ self-loops.
  - In this way, we obtain a $(2d)$-regualr graph.
- And then, we study random walks on G.
  - For $v, w \in V$, we define $P(v, w) = \frac{1}{2d}$ if $(v, w) \in E$ and $P(v, w) = 0$ o.w.;
  - We define $P(v, v) = \frac{2d - \deg(v)}{2d} = 1 - \frac{\deg(v)}{2d}$ for each $v \in V$.
    - Obviously, $P(v, v) \geq 1/2$.

$2d - \deg(v)$ self-loops are added for each $v \in V$.



$P(v, w) = 1/6$ if $(v, w) \in E$ and 0 otherwise.
$P(v, v) = 1 - \deg(v)/6$ for each $v \in V$.

$2d - \deg(v)$ self-loops are added for each $v \in V$.



$P(v, w) = 1/6$ if $(v, w) \in E$ and 0 otherwise.
$P(v, v) = 1 - \deg(v)/6$ for each $v \in V$.

The expansion tester by Goldreich & Ron

A property tester of two-sided error.

Expansion-Tester(G, ℓ, m, s)

repeat s times;
    Select a vertex $v \in V$ uniformly at random;
    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 $> \frac{1+7\epsilon}{n}\binom{m}{2}$
        then reject;
accept;

Theorem 3.1 (Main Theorem)

Let $0 \leq \epsilon \leq 0.025$*. With*

$$s \geq \frac{48}{\epsilon}, m \geq \frac{12 \cdot s \cdot \sqrt{n}}{\epsilon^2}, \ell \geq \frac{16 \cdot d^2 \cdot \ln(n/\epsilon)}{\alpha^2},$$

Algorithm Expansion-Tester

- *accepts every* α*-expander with probability* $\geq \frac{2}{3}$*, and*
- *rejects with probability* $\geq \frac{2}{3}$ *every graph that is* ϵ*-far from any* $\frac{c \cdot \alpha^2}{d^2 \cdot \ln(n/\epsilon)}$*-expander with probability* $\geq \frac{2}{3}$*, where* $c > 0$ *is a large enough constant.*

The query complexity of this algorithm is $O(\ell \cdot m \cdot s) = O\left(\frac{d^2 \cdot \ln(n/\epsilon) \cdot \sqrt{n}}{\alpha^2 \cdot \epsilon^3}\right)$.

The general idea of how the tester works

- The graph is **regular** and **non-bipartite**, so the distribution of the endpoint of a random walk converges to a uniform distribution.
  - For peopele who are familiar with Markov chains, the above distribution is called a **stationary distribution**.
- The key point is how fast (i.e., the **mixing time** of the corresponding Markov chain) the distribution of the endpoints of the random walk converges to a uniform distribution.
- A graph with high expansion is believed to have fast mixing time.

The general idea of how the tester works

- The graph is **regular** and **non-bipartite**, so the distribution of the endpoint of a random walk converges to a uniform distribution.
  - For peopele who are familiar with Markov chains, the above distribution is called a **stationary distribution**.
- The key point is how fast (i.e., the **mixing time** of the corresponding Markov chain) the distribution of the endpoints of the random walk converges to a uniform distribution.
- A graph with high expansion is believed to have fast mixing time.

The general idea of how the tester works

- The graph is **regular** and **non-bipartite**, so the distribution of the endpoint of a random walk converges to a uniform distribution.
  - For peopele who are familiar with Markov chains, the above distribution is called a **stationary distribution**.
- The key point is how fast (i.e., the **mixing time** of the corresponding Markov chain) the distribution of the endpoints of the random walk converges to a uniform distribution.
- A graph with high expansion is believed to have fast mixing time.

The general idea of how the tester works (contd.)

- How to know that the distribution of the endpoints of the random walk is close to the uniform distribution or not?
  - Repeatedly perform the random walk and count the number of collisions.
  - We say that two random walks have a collision: *their endpoints are the same*.
- If a graph is an α-expander, then the expected number of collisions should be small.

The general idea of how the tester works (contd.)

- How to know that the distribution of the endpoints of the random walk is close to the uniform distribution or not?
  - Repeatedly perform the random walk and count the number of collisions.
  - We say that two random walks have a collision: *their endpoints are the same*.
- If a graph is an α-expander, then the expected number of collisions should be small.

The general idea of how the tester works (contd.)

- How to know that the distribution of the endpoints of the random walk is close to the uniform distribution or not?
  - Repeatedly perform the random walk and count the number of collisions.
  - We say that two random walks have a collision: *their endpoints are the same*.
- If a graph is an α-expander, then the expected number of collisions should be small.

The general idea of how the tester works (contd.)

- For graphs far from α^*-expanders, the author showed that:
  - There exists a subset $U \subseteq V$ with $|U| < n/2$ such that the random walks starting from any $u \in U$ requires much longer mixing time.
  - When the random walks do not proceed long enough, the **variation distance** between the uniform distribution and the distribution of the endpoints of the random walk starting from any $u \in U$ is large.
  - The above fact implies that the expected number of collisions of the random walks is high.

The general idea of how the tester works (contd.)

- For graphs far from α^*-expanders, the author showed that:
  - There exists a subset $U \subseteq V$ with $|U| < n/2$ such that the random walks starting from any $u \in U$ requires much longer mixing time.
  - When the random walks do not proceed long enough, the **variation distance** between the uniform distribution and the distribution of the endpoints of the random walk starting from any $u \in U$ is large.
  - The above fact implies that the expected number of collisions of the random walks is high.

The general idea of how the tester works (contd.)

- For graphs far from α^*-expanders, the author showed that:
  - There exists a subset $U \subseteq V$ with $|U| < n/2$ such that the random walks starting from any $u \in U$ requires much longer mixing time.
  - When the random walks do not proceed long enough, the **variation distance** between the uniform distribution and the distribution of the endpoints of the random walk starting from any $u \in U$ is large.
  - The above fact implies that the expected number of collisions of the random walks is high.

The general idea of how the tester works (contd.)

- For graphs far from α^*-expanders, the author showed that:
  - There exists a subset $U \subseteq V$ with $|U| < n/2$ such that the random walks starting from any $u \in U$ requires much longer mixing time.
  - When the random walks do not proceed long enough, the **variation distance** between the uniform distribution and the distribution of the endpoints of the random walk starting from any $u \in U$ is large.
  - The above fact implies that the expected number of collisions of the random walks is high.

Outline

Markov chains

- Markov chain: a sequence of random variables $X_0, X_1, X_2, \ldots,$ (stochastic process) with the Markov property:
  - $\mathbf{Pr}[X_{n+1} = x \mid X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n]$ $= \mathbf{Pr}[X_{n+1} = x \mid X_n = x_n]$.
- For all i, $X_i \in \Omega$, where Ω is a finite state space.
- $P : \Omega^2 \mapsto [0, 1]$ denote the matrix of the transition probabilities.
  - There is an underlying graph corresponding to P.

Markov chains

- Markov chain: a sequence of random variables $X_0, X_1, X_2, \ldots,$ (stochastic process) with the Markov property:
  - $\mathbf{Pr}[X_{n+1} = x \mid X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n]$
    $= \mathbf{Pr}[X_{n+1} = x \mid X_n = x_n]$.
- For all i, $X_i \in \Omega$, where Ω is a finite state space.
- $P : \Omega^2 \mapsto [0, 1]$ denote the matrix of the transition probabilities.
  - There is an underlying graph corresponding to P.

Markov chains (contd.)

- When the underlying undirected graph is regular, connected and non-bipartite, the Markov chain $\mathcal{M}$ has a stationary distribution $\boldsymbol{\pi}$, which is a uniform distribution $\mathscr{U}$.
  - $\boldsymbol{\pi} = (\pi_x)_{x \in \Omega}$ is a stationary distribution of $\mathcal{M}$ if $\sum_{j \in \Omega} \pi_j = 1$ and $\pi_j = \sum_{i \in \Omega} = \pi_i \cdot P(i, j)$ for each $j \in \Omega$.
  - That is, $\boldsymbol{\pi} = \boldsymbol{\pi} \cdot P$
- A Markov chain $\mathcal{M}$ is reversible if $\pi_x \cdot P(x, y) = \pi_y \cdot P(y, x)$.
- In this paper, the random walk can be viewed as a Markov chain $\mathcal{M}_G$ with state space $\Omega = V$.
  - It is easy to see that $\mathcal{M}_G$ is reversible and has a uniform stationary distribution.

Markov chains (contd.)

- When the underlying undirected graph is regular, connected and non-bipartite, the Markov chain $\mathcal{M}$ has a stationary distribution $\boldsymbol{\pi}$, which is a uniform distribution $\mathscr{U}$.
  - $\boldsymbol{\pi} = (\pi_x)_{x \in \Omega}$ is a stationary distribution of $\mathcal{M}$ if $\sum_{j \in \Omega} \pi_j = 1$ and $\pi_j = \sum_{i \in \Omega} = \pi_i \cdot P(i, j)$ for each $j \in \Omega$.
  - That is, $\boldsymbol{\pi} = \boldsymbol{\pi} \cdot P$
- A Markov chain $\mathcal{M}$ is reversible if $\pi_x \cdot P(x, y) = \pi_y \cdot P(y, x)$.
- In this paper, the random walk can be viewed as a Markov chain $\mathcal{M}_G$ with state space $\Omega = V$.
  - It is easy to see that $\mathcal{M}_G$ is reversible and has a uniform stationary distribution.

Conductance of a Markov chain

Conductance is used to control the speed of convergence of a Markov chain. Here we adapt the original definition to $\mathcal{M}_G$.

- The conductance of $\mathcal{M}_G$:

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

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

- If G is an α-expander, then $\Phi_G \geq \frac{\alpha}{2d}$.

Variation distance

Definition 4.1 (Variation distance)

The variation distance between two probability distributions $\mathcal{X}$ and $\mathcal{Y}$ over the same finite domain Ω is

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

- Let $P_x^t(y)$ be the probability that the Markov chain with the initial state x ends after t steps in a state y. We define that

$$\Delta_x(t) = \frac{1}{2} \sum_{y \in \Omega} |P_x^t(y) - \pi_y|.$$

to be the variation distance w.r.t. the initial state x between $P_x^t(\cdot)$ and $\boldsymbol{\pi}$.

The rate of convergence & mixing time

Definition 4.2 (Rate of convergence)

The rate of convergence of a Markov chain $\mathcal{M}$ with initial state x to the stationary distribution is defined as

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

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

The mixing time w.r.t. the conductance & the stationary distribution

Proposition (Sinclair 1992)

$\mathcal{M}$: a finite, reversible, ergodic Markov chain and $P(x,x) \geq 1/2$ for all states x;
Φ: the conductance of $\mathcal{M}$.
Then the mixing time of $\mathcal{M}$ satisfies

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

- **Note:** The Markov chain $\mathcal{M}_G$ is "ergodic", though we do not introduce this term since it involves quite many concepts so that we just ignore its definition in this talk.

Outline

A result by Goldreich & Ron 2000

Lemma 5.1 (Goldreich & Ron 2000)

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

- $C_{i,j;v}$: indicator random variable; $C_{i,j;v} = 1$ iff the ith and the jth random walks starting from v have a collision.
- X_v: the number of collisions among the m random walks of length ℓ starting from v.
  - $X_v = \sum_{1 \leq i < j \leq m} C_{i,j;v}$.
- P_v^ℓ: the distribution of the endpoint of the random walk of length ℓ starting from v.
  - $||P_v^\ell||_2 = \sqrt{\sum_{w \in V} (P_v^\ell(w))^2}$ (i.e., 2-norm).
  - $(P_v^\ell(w))^2$: The probability that two random walks of length ℓ starting from v end at the same vertex w.

(*) By setting $\ell = \frac{16d^2 \cdot \ln(n/\epsilon)}{\alpha^2}$ and Sinclair's proposition, we have $||P_v^\ell||_2^2 \leq (1+\epsilon)^2/n$.

(**) Moreover, by Cauchy–Schwarz inequality $\Rightarrow ||P_v^\ell||_2^2 \geq 1/n$.

- Using (*) and Chebyshev's inequality, we have the following lemma.

Lemma 5.2 (Accepting expanders)

Let $m \geq \frac{12 \cdot s \cdot \sqrt{n}}{\epsilon^2}$ *and* $\ell \geq \frac{16d^2 \cdot \ln(n/\epsilon)}{\alpha^2}$. *Then Expansion-Tester accepts every* α*-expander with probability at least* $\frac{2}{3}$.

As to the rejections

Lemma 5.3 (Rejections)

Let $0 < \epsilon < 0.1$, $0 < \delta < 1/2$, and $s \geq 2/\delta$. If there exists $U \subseteq V$ with $|U| \geq \delta n$, such that for every $u \in U$, $d_{TV}(P_u^\ell, \mathscr{U}) \geq 1.5\sqrt{\epsilon}$, then Expansion-Tester rejects with probability at least $\frac{2}{3}$.

Ideas of the proof.

- $d_{TV}(P_u^\ell, \mathscr{U}) \geq 6\sqrt{\epsilon} \Rightarrow$ high expected number of collisions for the random walks.
  - The expected number of collisions: $\binom{m}{2} \cdot ||P_u^\ell||_2^2$.
  - We look for a probability vector P_u^ℓ **with the variation distance constraint** that minimizes $||P_u^\ell||_2^2$.
- Next, by the proof of Lemma 5.1, the observed number of collisions is $\geq (1-\epsilon)\binom{m}{2} \cdot ||P_u^\ell||_2^2$ with probability $\geq 1 - \frac{1}{3s}$.

As to the rejections (contd.)

The probability vector P_u^ℓ:

$$(\underbrace{\frac{1+3\sqrt{\epsilon}}{n},\ldots,\frac{1+3\sqrt{\epsilon}}{n}}_{n/2\text{ times}},\underbrace{\frac{1-3\sqrt{\epsilon}}{n},\ldots,\frac{1-3\sqrt{\epsilon}}{n}}_{n/2\text{ times}})$$

The vector of the uniform distribution $\mathscr{U}$:

$$(\underbrace{\frac{1}{n},\ \frac{1}{n},\ \ldots,\ \frac{1}{n}}_{n\text{ times}})$$

We have $\frac{1}{2}\cdot\sum_{w\in V}|P_u^\ell(w)-1/n| = 1.5\sqrt{\epsilon}$ and $||P_u^\ell||_2^2 = \frac{1+9\epsilon}{n}$.

So $(1-\epsilon)\cdot\binom{m}{2}\cdot||P_u^\ell||_2^2 \geq \frac{(1-\epsilon)(1+9\epsilon)}{n}\cdot\binom{m}{2} > \frac{1+7\epsilon}{n}\cdot\binom{m}{2}$.

Being far from α^*-expanders

Any graph that is ϵ-far from any α^*-expander has a small cut that separates a large set of vertices from the rest of the graph.

Lemma 5.4

Let $0 < \epsilon < 1$ and $\alpha^* \leq 0.1$. If G has a subset of vertices $A \subseteq V$ with $|A| \leq \frac{1}{12}\epsilon n$ such that $G[V \setminus A]$ is an $\frac{4\alpha^*}{\beta}$-expander, then G is not ϵ-far from any α^*-expander.

- Note that $\beta = \Theta(1)$ is a constant concerning *strong expansion*, which is ignored for this talk.

Corollary 5.5

Let G be ϵ-far from any α^*-expander with $\alpha^* \leq 0.1$. Then there exists $A \subseteq V$ with $\frac{1}{12}\epsilon n \leq |A| \leq \frac{1}{2}(1+\epsilon)n$ such that $|N_G(A, V)| < \frac{4\alpha^*}{\beta}|A|$.

Being far from α^*-expanders

Any graph that is ϵ-far from any α^*-expander has a small cut that separates a large set of vertices from the rest of the graph.

Lemma 5.4

Let $0 < \epsilon < 1$ and $\alpha^* \leq 0.1$. If G has a subset of vertices $A \subseteq V$ with $|A| \leq \frac{1}{12}\epsilon n$ such that $G[V \setminus A]$ is an $\frac{4\alpha^*}{\beta}$-expander, then G is not ϵ-far from any α^*-expander.

- Note that $\beta = \Theta(1)$ is a constant concerning *strong expansion*, which is ignored for this talk.

Corollary 5.5

Let G be ϵ-far from any α^*-expander with $\alpha^* \leq 0.1$. Then there exists $A \subseteq V$ with $\frac{1}{12}\epsilon n \leq |A| \leq \frac{1}{2}(1+\epsilon)n$ such that $|N_G(A, V)| < \frac{4\alpha^*}{\beta}|A|$.

Being far from α^*-expanders (contd.)

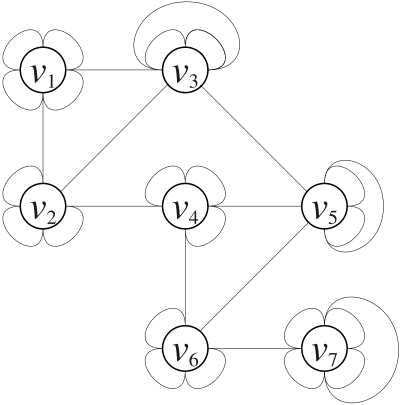
Lemma 5.6

Let A be a subset of V with $|A| \le \frac{1}{2}(1+\epsilon)n$ and $|N_G(A,V)| \le \frac{|A|}{10(\ell+1)}$. Then there exists a set U with $|U| \ge |A|/2$ such that for every $u \in U$,

$$d_{TV}(P_v^\ell, \mathscr{U}) \ge \frac{1-2\epsilon}{4}.$$

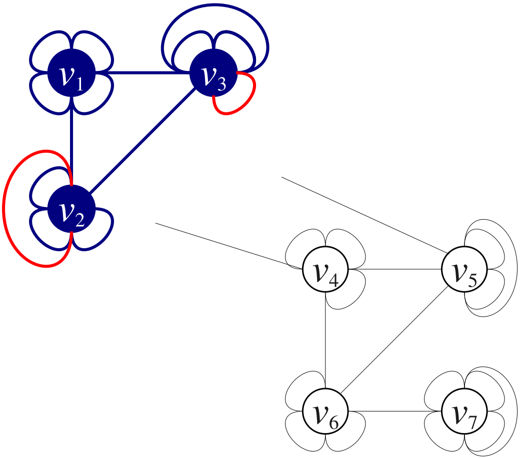
Note that $\frac{1-2\epsilon}{4} \ge 1.5\sqrt{\epsilon}$ for $\epsilon < 0.025$.

Being far from α^*-expanders (contd.)



$A = \{v_1\}$, $N_G(A, V) = \{v_2, v_3\}$.

Being far from α^*-expanders (contd.)



$A = \{v_1\}$, $N_G(A, V) = \{v_2, v_3\}$.

Being far from α^*-expanders (contd.)

- Let $G_A = G[A \cup N_G(A, V)]$. Consider a random walk on G_A.
- Y_i: the indicator random variable for the event that the ith vertex of the random walk is in $N_G(A, V)$.
  - $\mathbf{Pr}[Y_i = 1] = \frac{|N_G(A,V)|}{|V(G_A)|}$
  - **The reason:** the starting vertex is chosen uniformly at random & the stationary distribution is uniform.
- We can show that $\mathbf{Pr}\left[\exists i \in \{0, 1, \ldots, \ell\}, Y_i = 1\right] \leq \frac{1}{10(\ell+1)}$.

Being far from α^*-expanders (contd.)

- The probability that an ℓ-step random walk in G starting at a vertex chosen uniformly from A will remain in A is at least $1 - \frac{1}{10(\ell+1)} \geq \frac{9}{10}$.
  - Thus, there must be $U \subseteq A$ of size $\geq |A|/2$ such that a random walk starting from a vertex in U remains in A with probability $\geq \frac{3}{4}$.
  - Thus, there must be $U \subseteq A$ of size $\geq |A|/2$ such that a random walk starting from a vertex in U does **NOT** in A with probability $\leq \frac{1}{4}$.
  - In contrast to the uniform distribution: $\frac{|V \setminus A|}{|V|} \geq \frac{1-\epsilon}{2}$.

Being far from α^*-expanders (contd.)

- The probability that an ℓ-step random walk in G starting at a vertex chosen uniformly from A will remain in A is at least $1 - \frac{1}{10(\ell+1)} \geq \frac{9}{10}$.
  - Thus, there must be $U \subseteq A$ of size $\geq |A|/2$ such that a random walk starting from a vertex in U remains in A with probability $\geq \frac{3}{4}$.
  - Thus, there must be $U \subseteq A$ of size $\geq |A|/2$ such that a random walk starting from a vertex in U does **NOT** in A with probability $\leq \frac{1}{4}$.
  - In contrast to the uniform distribution: $\frac{|V \setminus A|}{|V|} \geq \frac{1-\epsilon}{2}$.

Putting everything together you will derive the proof of the main theorem.

Thank you!