LECTURE 15

Last time

- Testing triangle-freeness
- Regularity Lemma
- Triangle-removal lemma

Today

- Finish testing triangle-freeness
- Testing other properties of dense graphs
- Lower bound for triangle-freeness
- Behrend’s construction

Sign up for project meetings (if you haven’t yet)

Testing Triangle-Freeness

Input: parameters ε, n, access to undirected graph $G = (V, E)$ represented by $n \times n$ adjacency matrix.

Goal: **Accept** if G has no triangles;
reject w.p. $\geq \frac{2}{3}$ if G is ε-far from triangle-free
(at least $\varepsilon \binom{n}{2}$ edges need to be removed to get rid of all triangles).

- [Alon Fischer Krivelevich Szegedy 09]: Time that depends only on ε

Tester

Algorithm (**Input:** ε, n; query access to adjacency matrix of $G=(V,E)$)

1. Repeat ***s*** times:
2. Sample vertices v_1, v_2, v_3 uniformly at random
3. **Reject** if they form a triangle.
4. **Accept**.

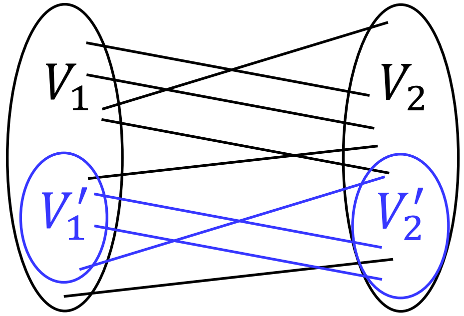
How many repetitions suffice?

Triangle-Removal Lemma

$\forall \varepsilon \exists \delta = \delta(\varepsilon)$ such that every n-node graph that is ε-far from triangle-free contains at least $\delta \cdot \binom{n}{3}$ triangles.

- By Witness Lemma, setting $s = 2/\delta$ yields a tester.

Definitions from Last Lecture



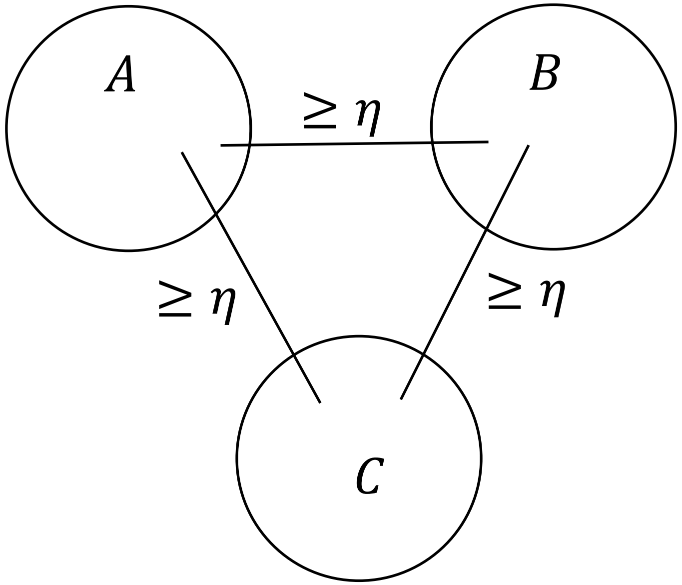
- The edge density of the pair (V_1, V_2), denoted $d(V_1, V_2)$, is $\frac{|e(V_1,V_2)|}{|V_1|\cdot|V_2|}$.

- A pair (V_1, V_2) of disjoint subsets of vertices is γ-regular if $\forall V_1' \subseteq V_1, V_2' \subseteq V_2$, such that $|V_1'| > \gamma|V_1|$ and $|V_2'| > \gamma|V_2|$,
  $|d(V_1, V_2) - d(V_1', V_2')| < \gamma$.

Triangles in a Graph with Three Regular Pairs

Lemma [Kolmos Simonovits]

$\forall \eta$>0, if A, B, C are disjoint subsets of V and each pair of them is γ^{Δ}-regular with density at least η then G contains at least $\delta^{\Delta}\ |A| \cdot |B| \cdot |C|$ triangles, where $\gamma^{\Delta} = \gamma^{\Delta}(\eta) = \frac{\eta}{2}$ and $\delta^{\Delta} = \delta^{\Delta}(\eta) = \frac{1}{8}(1-\eta)\eta^3$.



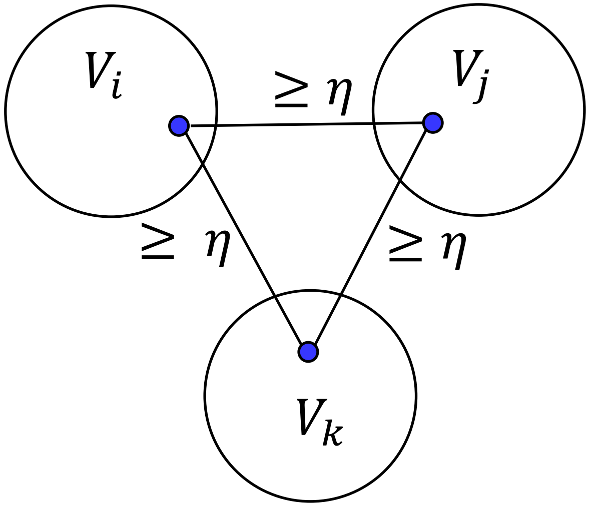
Proof of the Triangle-Removal Lemma: Idea

Triangle-Removal Lemma

$\forall \varepsilon \exists \delta = \delta(\varepsilon)$ such that every n-node graph that is ε-far from triangle-free contains at least $\delta \cdot \binom{n}{3}$ distinct triangles.

Main Idea: Consider a graph G which is ε-far from being triangle-free.

- We apply the Regularity Lemma to get a regular partition.
- We carefully remove fewer than $\varepsilon \binom{n}{2}$ edges, and show that there remains a triangle consisting of edges between regular dense pairs.
- We apply [Kolmos Simonovits] to get many triangles.



Proof of the Triangle-Removal Lemma

Triangle-Removal Lemma

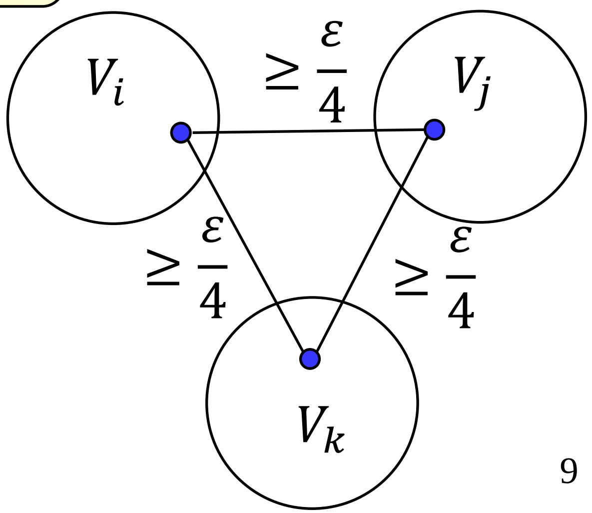
$\forall \varepsilon \exists \delta = \delta(\varepsilon)$ such that every n-node graph that is ε-far from triangle-free contains at least $\delta \cdot \binom{n}{3}$ distinct triangles.

Proof: Consider a graph G which is ε-far from being triangle-free.

- An edge (u, v), where $u \in V_i$ and $v \in V_j$ is useful if it satisfies:
  1. $i \neq j$
  2. (V_i, V_j) is $\varepsilon/8$-regular
  3. the density $d(V_i, V_j) \geq \varepsilon/4$

Claim. Graph G has less than $\varepsilon \binom{n}{2}$ non-useful edges.

Triangle of useful edges



- When we remove all non-useful edges, there is still a triangle!
- By [Kolmos Simonovits], there are at least

$$\delta^{\Delta}\left(\frac{\varepsilon}{4}\right) \cdot |V_i| \cdot |V_j| \cdot |V_k| \geq \frac{1}{8}\left(1 - \frac{\varepsilon}{4}\right)\left(\frac{\varepsilon}{4}\right)^3 \cdot \frac{n^3}{T^3}$$

triangles.

Testing Other Properties

Testing Subgraph-Freeness [Alon 02]

Let H be a fixed graph on h nodes.

Let $\mathcal{P}_H$ be the property that G does not contain a copy of H as a subgraph.

1. If H is bipartite:
   - There is a 2-sided error tester for $\mathcal{P}_H$ with $O\left(\frac{1}{\varepsilon}\right)$ queries.
   - There is a 1-sided error tester for $\mathcal{P}_H$ with $O\left(h^2\left(\frac{1}{2\varepsilon}\right)^{h^2/4}\right)$ queries.

Polynomial in $1/\varepsilon$ for fixed H.

2. If H is not bipartite, then there exists c > 0, such that every 1-sided error tester for $\mathcal{P}_H$ makes $\Omega(\left(\frac{c}{\varepsilon}\right)^{c\log\frac{c}{\varepsilon}})$ queries.

Super-polynomial in $1/\varepsilon$.

- We will prove part (2) for triangles.

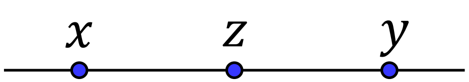
Main Combinatorial Tool for Proving the Lower Bound: Behrend's Construction

Dense Sets of Integers with no Arithmetic Progression

Behrend's Theorem

For all integer $m \geq 1$, there exists a set $S \subseteq [m]$ such that $|S| \geq \frac{m}{2^{3\sqrt{\log_2 m}}}$ and the only solution to $x + y = 2z$ for $x, y, z \in S$ is $x = y = z$.

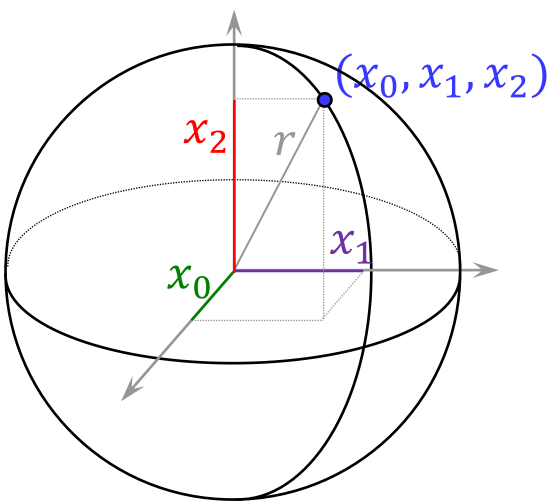
- Behrend's bound [Behrend 46] is slightly better.



- The best known is $\Omega\left(\frac{m}{2^{2\sqrt{2}\sqrt{\log_2 m}}} \log_2^{1/4} m\right)$ [Elkin 10]

Proof idea: Represent integers in $[m]$ as k-digit numbers base d, where k and d are parameters.

- For a number x, view its digits as coordinates of a point $(x_0, x_1, \ldots, x_{k-1})$
- Pick points that lie on the same sphere: i.e., with fixed $x_0^2 + x_1^2 + \cdots + x_{k-1}^2$
- Then no three of them lie on the same line, which ensures that no point is the average of two other points.



Proof of Behrend's Theorem

Behrend's Theorem

For all integer $m \geq 1$, there exists a set $S \subseteq [m]$ such that $|S| \geq \frac{m}{2^{3\sqrt{\log_2 m}}}$ and the only solution to $x + y = 2z$ for $x, y, z \in S$ is $x = y = z$.

Proof: For an integer $B > 0$, define a set

$$S_B = \left\{ \sum_{i=0}^{k-1} x_i d^i : \text{each } x_i \in \left\{0, \dots, \frac{d}{2} - 1\right\} \text{ and } B = \sum_{i=0}^{k-1} x_i^2 \right\}$$

- All numbers in sets S_B are less than d^k.

  We set $d^k = m$ to ensure $S_B \subseteq [m]\ \forall B$.

Claim

For all B, the only solution to $x + y = 2z$ for $x, y, z \in S_B$ is $x = y = z$.

Proof of Claim

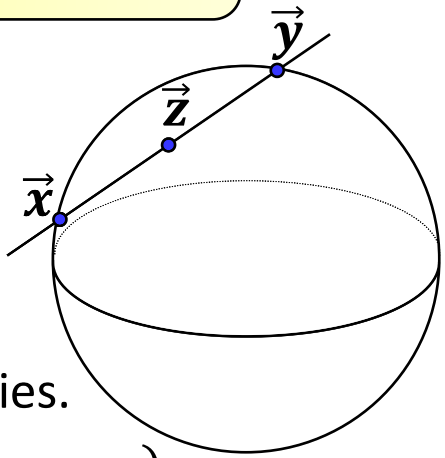
For an integer $B > 0$, define a set

$$S_B = \left\{ \sum_{i=0}^{k-1} x_i d^i : \text{each } x_i \in \left\{0, \ldots, \frac{d}{2} - 1\right\} \text{ and } B = \sum_{i=0}^{k-1} x_i^2 \right\}$$

Claim

For all B, the only solution to $x + y = 2z$ for $x, y, z \in S_B$ is $x = y = z$.

Proof: Suppose $x + y = 2z$ for some $x, y, z \in S_B$.



- Representing x, y, z base d, we get

$$\sum_{i=0}^{k-1} x_i d^i + \sum_{i=0}^{k-1} y_i d^i = 2 \sum_{i=0}^{k-1} z_i d^i$$

- Since x_i, y_i, z_i are less than $d/2$ for all i, there are no carries.

  That is, $(x_0, x_1, \ldots, x_{k-1}) + (y_0, y_1, \ldots, y_{k-1}) = 2(z_0, z_1, \ldots, z_{k-1})$

  But these three points are on a sphere,

  so one can be the average of the other two only if they are identical.

Proof of Behrend's Theorem: Setting Parameters

Behrend's Theorem

For all integer $m \geq 1$, there exists a set $S \subseteq [m]$ such that $|S| \geq \frac{m}{2^{3\sqrt{\log_2 m}}}$ and the only solution to $x + y = 2z$ for $x, y, z \in S$ is $x = y = z$.

Proof: For an integer $B > 0$, define a set

$$S_B = \left\{ \sum_{i=0}^{k-1} x_i d^i : \text{each } x_i \in \left\{0, \dots, \frac{d}{2} - 1\right\} \text{ and } B = \sum_{i=0}^{k-1} x_i^2 \right\}$$

- Set $d^k = m$ and $d = 2^{\sqrt{1/2 \cdot \log m}}$. Then $k =$
- How many possibilities for B?
- How many numbers are in all sets S_B?
- By an averaging argument, at least one of the sets has size at least

Testing Triangle-Freeness

Input: parameters ε, n, access to undirected graph $G = (V, E)$ represented by $n \times n$ adjacency matrix.

Goal: **Accept** if G has no triangles;
reject w.p. $\geq \frac{2}{3}$ if G is ε-far from triangle-free
(at least $\varepsilon \binom{n}{2}$ edges need to be removed to get rid of all triangles).

- [Alon Fischer Krivelevich Szegedy 09]: Time that depends only on ε
- Goal

Lower Bound for Testing Triangle-Freeness [Alon 02]

Testing triangle-freeness with 1-sided error requires super-polynomial dependence on $1/\varepsilon$.

$\Omega\left(\left(\frac{c}{\varepsilon}\right)^{c \log \frac{c}{\varepsilon}}\right)$ queries for some $c > 0$

Canonical Tester for Dense Graphs

Canonical Tester (**Input:** ε, n; query access to adjacency matrix of $G=(V,E)$)

1. Sample s nodes uniformly at random.
2. Query all pairs of sampled nodes.
3. **Accept** or **reject** based on available information.

- Consider any property $\mathcal{P}$ of graphs that does not depend on the names of the nodes. That is, if $G \in \mathcal{P}$ and G' is isomorphic to G then $G' \in \mathcal{P}$.

Exercise: Show that if there is an ε-tester T for $\mathcal{P}$ with query complexity $q(\varepsilon,n)$, then there is a canonical ε-tester T' for $\mathcal{P}$ with query complexity $O(q^2(\varepsilon,n))$. Moreover, if T has 1-sided error, so does T'.

A lower bound q for canonical tester implies a lower bound $\sqrt{q}$ for every tester

Sufficient to prove our lower bound $\Omega\left(\left(\frac{c}{\varepsilon}\right)^{c \log \frac{c}{\varepsilon}}\right)$ for 1-sided error canonical testers.

Goal for Proving the Lower bound

- A 1-sided error tester can reject only if it finds a triangle.
- Suppose we construct a graph G that is ε-far from being tringle free, where p fraction of triples are triangles for some small p.
- Consider a canonical tester T that samples q vertices.
- Let X be the number of triangles the tester catches.

$$\mathbb{E}[X] = p\binom{q}{3} = \Theta(p \cdot q^3)$$

- Suppose q is set so that $\mathbb{E}[X] \leq 1/2$
- By Markov, $\Pr[T \text{ rejects } G] \leq \Pr[X \geq 1] \leq \mathbb{E}[X] \leq \frac{1}{2} < \frac{2}{3}$
- So, for T to reject with high enough probability, $q = \Omega\left(p^{-\frac{1}{3}}\right)$

Sufficient to ensure $p = O\left(\left(\frac{\varepsilon}{c}\right)^{c \log \frac{c}{\varepsilon}}\right)$

Recall: Arithmetic-Progression-Free Sets

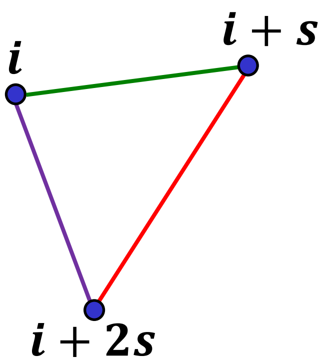
Behrend's Theorem

For all integer $m \geq 1$, there exists a set $S \subseteq [m]$ such that $|S| \geq \frac{m}{8^{\sqrt{\log_2 m}}}$ and the only solution to $x + y = 2z$ for $x, y, z \in S$ is $x = y = z$.

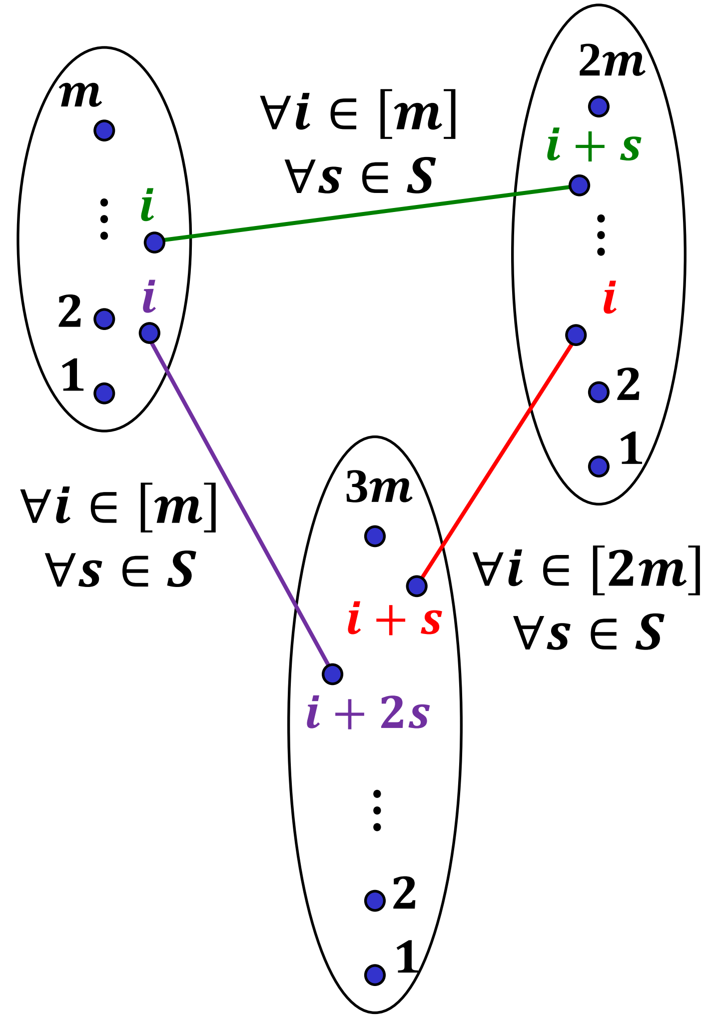
- We will use such a set S to construct a graph that is
  - far from triangle free
  - has relatively few triangles

Initial Graph Construction

- Let $S \subset [m]$ be a set from Behrend's Thm
- We construct a tripartite graph with m, $2m$, and $3m$ nodes in the three parts
- Intended triangles



- No other triangles:
  If $(i, i+x, i+x+y)$ is a triangle, then
  $x \in S, y \in S$, and $x+y = 2z$ for $z \in S$
  But then $x = y = z$ by construction of S
- All triangles are edge disjoint: each edge participates in exactly one triangle.



Parameters of the Initial Construction

- Number of nodes, n

  $6m$

- Number of edges

  $3m \cdot |S|$

- Number of (edge-disjoint) triangles, T

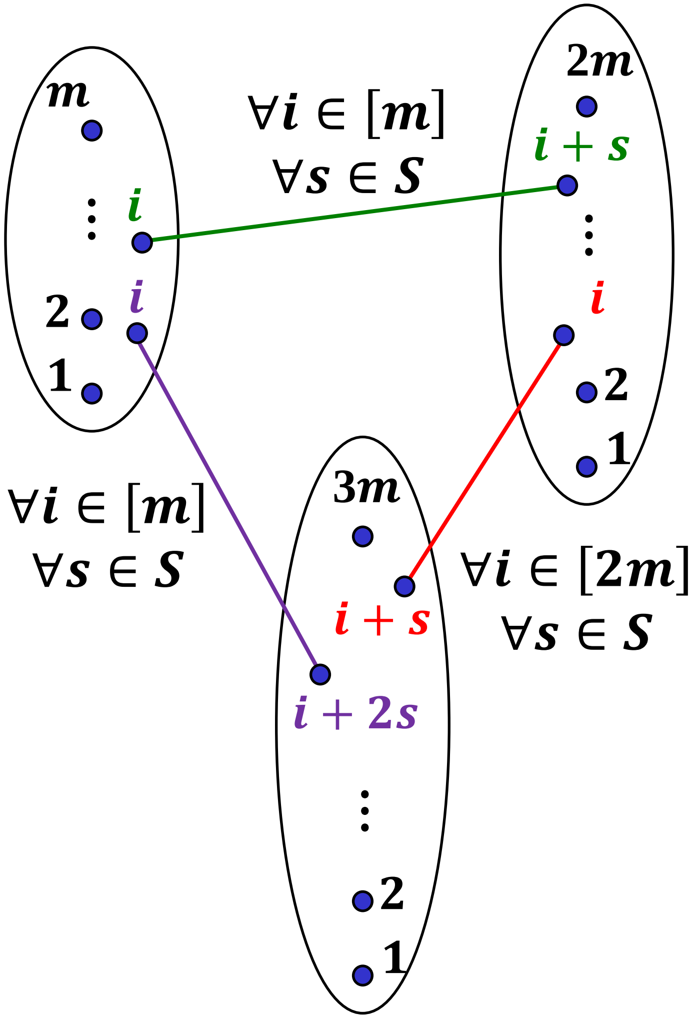
  $m \cdot |S|$

- Distance to triangle-freeness

Necessary and sufficient to remove one edge from each triangle, because they are edge-disjoint.

$$\frac{T}{\binom{n}{2}} = \Theta\left(\frac{m \cdot |S|}{m^2}\right) = \Theta\left(\frac{|S|}{m}\right) = \Theta\left(\frac{1}{8^{\sqrt{\log m}}}\right)$$

Not constant!

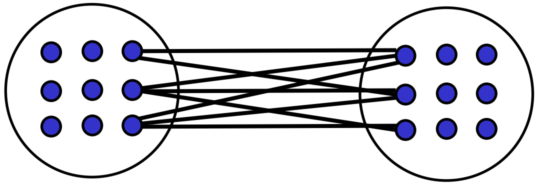


Blowup of a Graph

To construct a b-blowup of a graph,

- make b copies of each node;
- make two copies (of different nodes) adjacent

  iff their originals are adjacent.



Parameters of the Blowup Construction

- Number of nodes, n

  $6mb$
- Number of edges

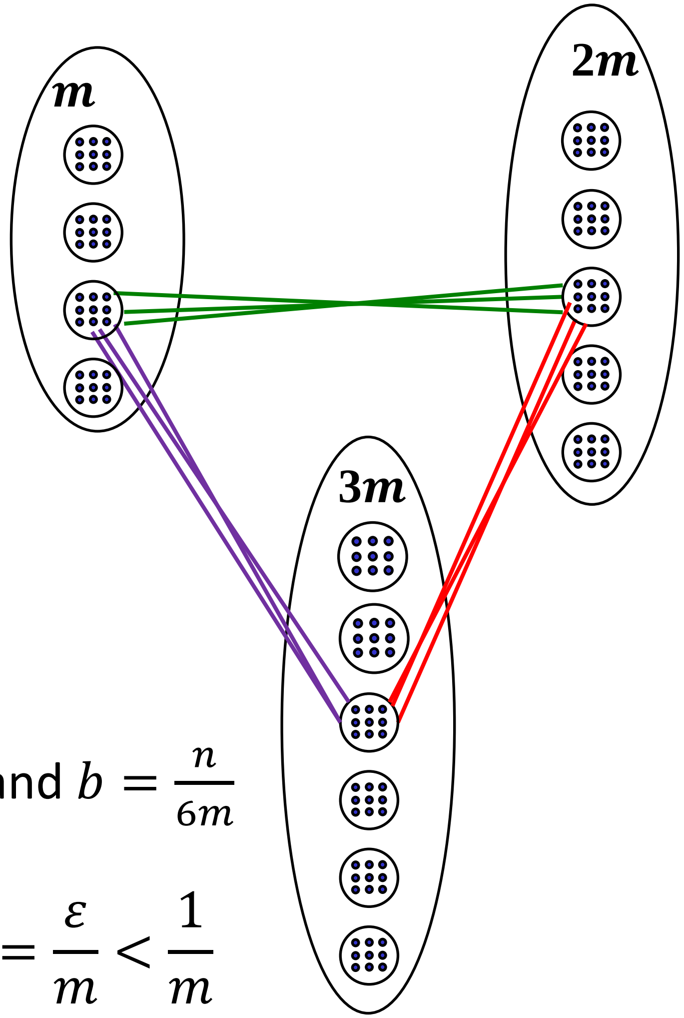
  $3m \cdot |S| \cdot b^2$
- Number of triangles

  $m \cdot |S| \cdot b^3$
- Number of (edge-disjoint) triangles, T

  $m \cdot |S| \cdot b^2$
- Distance to triangle-freeness

  $$\frac{T}{\binom{n}{2}} = \Theta\left(\frac{m \cdot |S| \cdot b^2}{(m \cdot b)^2}\right) = \Theta\left(\frac{|S|}{m}\right) = \Theta\left(\frac{1}{8^{\sqrt{\log m}}}\right)$$
- Given ε and n, pick m so that $\varepsilon = \Theta\left(\frac{1}{8^{\sqrt{\log m}}}\right)$ and $b = \frac{n}{6m}$
- Fraction of triples that are triangles:

  $$\approx \frac{m \cdot |S| \cdot b^3}{n^3} \approx \frac{m \cdot |S|}{m^3} = \frac{|S|}{m^2} = \frac{\varepsilon}{m} < \frac{1}{m}$$



Conclusion: Triangle-Freeness

- The query complexity of testing triangle-freeness with 1-sided error depends only on ε
(and is independent of the size of the graph).

- However, the dependence is super-polynomial in $1/\varepsilon$