Sublinear Algorithms	November 17, 19, 2015

Lecture 23-24

Lecturer: Sofya Raskhodnikova *Scribe(s): Hongyuan Zhan*

In this lecture, we will see the lower bound on query complexity for testing triangle-freeness in dense graph model. The note is based on lecture 23, 24 in Sublinear Algorithm course Fall 2015, lectured by Sofya Raskhodnikova. The original work is based on [1].

1 Introduction

1.1 Overview

The following theorem on query complexity for testing triangle-freeness in dense graph model is by Alon in [1].

Theorem 1 (Alon's Theorem)**.** *Every 1-sided-error ϵ-tester for triangle-freeness in the dense graph model makes at least $\left(\frac{c}{\epsilon}\right)^{c\log\left(\frac{c}{\epsilon}\right)}$ queries for some constant c.*

The idea of the proof is to construct graphs that are ϵ-far from triangle-free, but has at most $\left(\frac{c}{\epsilon}\right)^{c\log\left(\frac{c}{\epsilon}\right)} n^3$ triangles.

1.2 Notations

The following notations will be used in this note. $\mathbf{Z}$ denotes the ring of integers, and $\mathbf{Z}^+$ consists for only positive integers. For a set X, $|X|$ denotes its cardinality. $[m]$ denotes the set $\{1, 2, \cdots, m\}$.

2 Tools

We will use some results from number theory as the main tools in constructing the desired graph. The following theorem is usually called the *Behrend's theorem on 3-term arithmetic progression* (Behrend's 3AP) [2].

Theorem 2 (Behrend's 3AP)**.** *For all $m \in \mathbf{Z}$, there is a subset $X \in [m]$, whose cardinality satisfies $|X| \geq \frac{m}{e^{10\sqrt{\log m}}}$, and the only solution to*

$$x_1 + x_2 = 2x_3, \quad \textit{for } x_1, x_2, x_3 \in X,$$

is $x_1 = x_2 = x_3$.

We will construct a set parametrized by B, d. Let $B \in \mathbf{Z}^+$, and consider the set

$$X_B = \{\sum_{i=0}^{k} x_i d^i | 0 \leq x_i < \frac{d}{2}, \sum_{i=0}^{k} x_i^2 = B\},$$

where d is an integer parameter, $k = \lfloor \frac{\log m}{\log d} \rfloor - 1$, and x_i is non-negative. Conceptually, we could view the expression $\sum_{i=0}^{k} x_i d^i$ as a number in the d-nary system with coefficient x_i in the i-th digit. We will show that for some B, the set X_B satisfies the requirements in Behrend's theorem.

Claim 3. *$X_B \subseteq [m]$, for all $B \in \mathbf{Z}^+$.*

Proof. Clear that $\forall x \in X_B$, $x \geq 1$. Moreover, $x \leq d^{k+1} \leq d^{\frac{\log m}{\log d}} = m$. The first inequality follows from the definition of the set X_B (d-nary expression), the second inequality follows from the definition of k. ◻

Claim 4. *$\forall B \in \mathbf{Z}^+$, the only solution to*

$$x_1 + x_2 = 2x_3$$

such that $x_1, x_2, x_3 \in X_B$, is

$$x_1 = x_2 = x_3.$$

Proof. Let $x + y = 2z$, s.t. $x, y, z \in X_B$. We have

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

Since on each digit, $x_i, y_i, z_i < \frac{d}{2}$, there are no carries. This implies

$$x_i + y_i = 2z_i, \quad i = 1, 2, \cdots, k.$$

Consider the convex function $f(a) = a^2$. By Jensen's Inequality,

$$\sum_i \frac{f(a_i)}{n} \geq f(\frac{\sum_i a_i}{n}),$$

with equality if and only if all a_i's are same. Thus,

$$\frac{x_i^2 + y_i^2}{2} \geq (\frac{2z_i}{2})^2$$
$$x_i^2 + y_i^2 \geq 2z_i^2,$$

with the equality holds iff $x_i = y_i = z_i$. Summing over all i, we have

$$\sum_{i=0}^{k} x_i^2 + \sum_{i=0}^{k} y_i^2 \geq \sum_{i=0}^{k} 2z_i^2.$$

with equality holds when $x_i = y_i = z_i$. Since $x, y, z \in X_B$, by definition, $\sum_{i=0}^{k} x_i^2 = \sum_{i=0}^{k} y_i^2 = \sum_{i=0}^{k} z_i^2 = B$. Therefore the equality holds, and hence, $x = y = z$. □

Claim 5. *For some $B \in \mathbf{Z}^+$,$|X_B| \geq \frac{m}{e^{10\sqrt{\log m}}}$.*

Proof. By definition, $B = \sum_{i=0}^{k} x_i^2$, and $x_i < \frac{d}{2}$, we have $1 \leq B \leq (k+1)(\frac{d}{2})^2$. Hence there are at most $(k+1)(\frac{d}{2})^2$ possible B. On each digit, there are $(d/2)$ possible coefficients, hence

$$|\cup_B X_B| = (\frac{d}{2})^{k+1} - 1 \geq (\frac{d}{2})^k,$$

where the -1 comes from the definition that $B \in \mathbf{Z}^+$. Therefore, there exists B^* such that $|X_{B^*}| \geq \frac{(d/2)^k}{(k+1)(\frac{d}{2})^2}$. Let $d = e^{10\sqrt{\log m}}$, we have

$$|X_{B^*}| \geq \frac{m}{e^{10\sqrt{\log m}}}.$$

Now, combine claim 3, 4, 5: for all $m \in \mathbf{Z}$, we constructed a set X_{B^*} satisfies all the requirements in theorem 2. Hence theorem 2 is proved. □

3 Constructing "Hard" Graphs

In this section, we will illustrate the construction of the desired "hard" graph: a graph that are ϵ-far from triangle-free, but has at most $\left(\frac{c}{e}\right)^{c\log\left(\frac{c}{e}\right)} n^3$ triangles. We demonstrate two constructions, a first attempt that does not satisfy our goal, and a modified construction that resolves the issues.

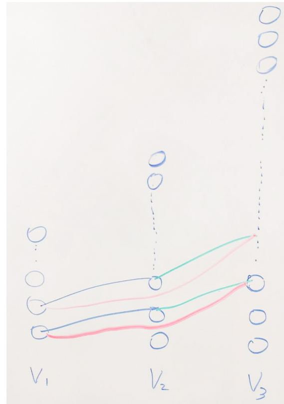

Figure 1: first construction of the "hard" graph

3.1 First Attempt (Fail)

Consider the following graph with three sets of nodes: V_1, V_2, V_3, such that $|V_1| = m, |V_2| = 2m, |V_3| = 3m$. Label $\{a_i\}_{i=1}^m$ the nodes in V_1, $\{b_i\}_{i=1}^{2m}$ nodes in V_2, $\{c_i\}_{i=1}^{3m}$ nodes in V_3. Let X be a subset of integers in $[m]$. Edges are added to the graph according to the following rule (see figure 1):

- for all $a_i \in V_1$, add the edge $(a_i, a_i + x)$, where $x \in X$ and $a_i + x \in V_2$.
- for all $b_i \in V_2$, add the edge $(b_i, b_i + x)$, where $x \in X$ and $b_i + x \in V_3$.
- for all $a_i \in V_1$, add the edge $(a_i, a_i + 2x)$, where $x \in X$ and $a_i + 2x \in V_3$.

By costruction, all triples of the form $(a, a + x, a + 2x)$ are triangles on the graph. In addition by Behrend's theorem (theorem 2), we know that there is a set $X \subseteq [m]$, all triangles are triples of the form $(a, a+x, a+2x)$. If not, say there are a triangle $(a, a+x_1, 2+2x_2)$, $x_1, x_2 \in X$, by Behrend's theorem, we must have $x_1 = x_2$. Therefore, this also implies all triangles are edge-disjoint. Notice that there are $6m$ edges on the graph and $6m$ nodes. Now we could do a summary of the graph and count the fraction of triangles (denote $|\Delta|$ the number of triangles, $|\Delta^{ED}|$ number of edge-disjoint triangles, $d(G, \Delta^{free})$ the distance to triangle-free graphs from our constructed graph).

$$\begin{aligned}
|V| &: 6m \\
|E| &: 6m \\
|\Delta| &: m|X| \\
|\Delta^{ED}| &: m|X| \\
d(G, \Delta^{free}) &: \tfrac{m|X|}{(6m)^2} = \Theta(\tfrac{1}{e^{10\sqrt{\log m}}}) \overset{\text{def}}{=} \epsilon_G, \text{ by Behrend's theorem} \\
\text{fraction of triangles} &: \tfrac{m|X|}{(6m)^3} = \Theta(\tfrac{\epsilon_G}{m})
\end{aligned}$$

To prove Alon's theorem (theorem 1), we want to construct a graph such that the fraction of triangles (among all triples) is a constant. If we had constructed a graph with constant fraction of triangles, that implies we need to at least make certain amount of queries. Therefore, this construction does not work because of the dependence on m, hence the dependence on the input size.

3.2 Modification

The previous construction suffers from the dependence on the number of nodes, to solve this issue, we do the following modification:

- Every node is replaced by S nodes. Denote $u(S)$ the set of nodes in lieu of node u in the first construction.
- Every edge (u, v) is replaced by edges on the complete bipartite graph $K_{u(S),v(S)}$.

Now let's do the graph summary on the modified graph:

$$
\begin{aligned}
|V| &: 6mS \\
|E| &: 6mS \\
|\Delta| &: m|X|S^3 \\
|\Delta^{ED}| &: m|X|S^2 \\
d(G, \Delta^{free}) &: \tfrac{m|X|S^2}{(6ms)^2} = \tfrac{|X|}{36m} = \tfrac{1}{36e^{10\sqrt{\log m}}}
\end{aligned}
$$

To make the modified graph ϵ-far from triangle-free, we need

$$
\begin{aligned}
\frac{1}{36e^{10\sqrt{\log m}}} &\geq \epsilon \\
\log m &\leq (\frac{1}{10}\log(\frac{1}{36\epsilon}))^2 \\
m &\leq (\frac{c}{\epsilon})^{c\log(\frac{c}{\epsilon})}
\end{aligned}
$$

for some constant c. We pick the largest m such that the modified graph is ϵ-far from triangle free. Denote n the total number of nodes, by construction, $S = \frac{n}{6m}$. The number of triangles in the modfied graph become

$$
\begin{aligned}
m|X|S^3 &= \frac{m^2}{\exp(10\sqrt{\log m})}(\frac{n}{6m})^3 \\
&\geq \frac{1}{6m}\epsilon n^3 \\
&\geq \frac{1}{6}(\frac{c}{\epsilon})^{c\log(\frac{c}{\epsilon})}\epsilon n^3 \\
&= \frac{1}{6}(\frac{c'}{\epsilon})^{c'\log(\frac{c'}{\epsilon})}n^3
\end{aligned}
$$

for some c'. Hence the fraction of triangles on the modified grapph is $\frac{1}{6}(\frac{c'}{\epsilon})^{c'\log(\frac{c'}{\epsilon})}$. Therefore, we constructed graphs which are ϵ-far from triangle-free, but has a constant fraction of triangles. Therefore, $(\frac{c'}{\epsilon})^{c'\log(\frac{c'}{\epsilon})}$ queries are needed. Hence, theorem 1 is proved.

References

[1] Noga Alon. Testing subgraphs in large graphs. *Random Structures & Algorithms*, 21(3-4):359–370, 2002.

[2] Felix A Behrend. On sets of integers which contain no three terms in arithmetical progression. *Proceedings of the National Academy of Sciences of the United States of America*, 32(12):331, 1946.