Introduction to Property Testing
Lecture 7

Lecture by Dr. Reut Levi
Typeset by Steven Karas

2020-06-03
Last edited 00:07:05 2020-06-29

Disclaimer These notes are based on the lectures for the course Introduction to Property Testing, taught by Dr. Reut Levi at IDC Herzliyah in the spring semester of 2019/2020. Sections may be based on the lecture slides prepared by Dr. Reut Levi.

1 Agenda

- Guidelines for homework 3
- Recap of last week

An extension was granted for homework #3 until next Sunday (June 14th). Homework 4 will be published on June 10th.

2 Guidelines for homework 3

2.1 Exercise 8.5

In exercise 8.5, we are asked to prove that bipartiteness does not have a POT. In other words, there is no tester with a constant number of queries.

Since we want to prove a lower bound, we may assume w.l.o.g. that the input graph is a directed cycle as a ring (in degree 1 and out degree 1). In this reduced model, we query the graph in the following ways:

- $in(i)$ - returns the incoming edges of vertex i
- $out(i)$ - returns the outgoing edges of vertex i

Assume towards contradiction that there exists a POT[1]A that makes q queries, where q is a constant. By definition of a POT this implies that $\Pr[A$ accepts a YES instance$] \geq t$ and $\Pr[A$ accepts a NO instance$] < t$.

Consider the following families of graphs: [2]

No-instances (family 1) wherein each graph is a group of k/l subgraphs each of which is a directed cycle of length $l \overset{\text{def}}{=} 2 \cdot \lceil q/2 \rceil + 1$ (odd length). Note that the graphs in this family are isomorphic and cover all possible vertex labels, but the topology stays the same.

> If q is odd, $l = q$, otherwise $l = q+1$

Yes-instances (family 2) wherein each graph is a group of $k/2l$ subgraphs each of which is a directed cycle of length $2l$.

To prove the lower bound we show 2 random processes: one that queries according to a random graph in family 1 and one that queries according to a random graph in family 2 and show that the answers of those processes are indistinguishable so long as we make no more than q queries. Formally, for any history of $k < q$ query-answer pairs:

$$((\underbrace{i_1}_{\text{label of a vertex}} , \underbrace{b_1}_{\text{bit for in/out}}), \underbrace{a_1}_{\text{answer}}), \ \ldots$$

---

[1]This is a two-sided POT. Proving that a one-sided POT cannot exist is much simpler as it's sufficient to show that it can't distinguish unclosed cycles.

[2]We construct graphs as topologies (comprised of all permutations of a particular graph's vertices), as we need to show that there does not exist any algorithm for a single instance and for a given instance an algorithm could potentially "guess" based on vertex labels without regard for the actual topology. This is explained further by footnote 36 on page 34 in the book.

The answer to query (i_{k+1}, b_{k+1}) is distributed the same in both families. Up to some point, the queries must be indistinguishable meaning that the knowledge graph that each process has built thus far is identical and therefore we don't know if we're looking at a yes or no instance.

This formulation can also be used with a single cycle, but the multiple-cycle version is slightly stronger as it proves the claim for other types of graphs.

2.2 Exercise 8.7

Prove that if $G = ([k], E)$ is ε-far from being bipartite then with probability $\geq 2/3$, the subgraph induced on $\widetilde{\Theta}(1/\varepsilon^2)$ vertices picked u.a.r. of G is $l(\varepsilon)$-far from being bipartite.

Recall the analysis we saw for testing bipartiteness in the dense model. Here too we consider a set R of $\widetilde{\Theta}(1/\varepsilon^2)$ vertices (picked u.a.r.) and partition R into two sets U and S. We would like to show that with probability $\geq 2/3$, for every 2-partition of U, the subgraph $G([S])$ has $\Omega(\varepsilon|S|^2)$ disturbing edges. The intuition is that if we find disturbing edges in every random partition then we are by definition that far away from being bipartite, so we can saw with some probability that if we sample at random and find a disturbing edge then we're likely some distance from being bipartite.

Denote $v_1, \ldots, v_s$, $u_1, \ldots, u_s$ as the vertices in S where $s = |S|/2$. Consider the following edge-disjoint matchings[3]:

Match each vertex v_i to $u_{(i+j) \bmod s}$ where j is the offset. As such, we have a family of matchings. Note that our matching is random pairs of vertices from the graph and can be modeled as pairs selected independently from $V \times V$.

Fix j and consider matching j. Assume that U is good, and recall from Claim 2 that for any good U and any 2-partition of U, at least $\frac{\varepsilon k^2}{6}$ edges disturb (U_1, U_2). From here, fix (U_1, U_2) - can you bound the probability that the number of edges of the matching j that disturb U_1, U_2 is $\leq \frac{\varepsilon s}{1000}$? hint: use Chernoff's Bound

We say that a matching is good if $\forall (U_1, U_2)$ the number of edges of the matching that disturb (U_1, U_2) is $> \frac{\varepsilon s}{1000}$.

Can you bound the probability that a matching is not good? (hint: union bound)

Assume the probability that a matching is not good is $\leq 1/100$. Can you bound the probability that more than half the matchings are not good? (hint: Markov's Inequality)

Show that if more than half the matchings are good then the subgraph induced by R is $\Omega(\varepsilon)$-far from being bipartite

---

[3]Edge disjoint matchings are collections of edges such that no vertex has more than one edge

3 External Graph Theory

This theory concerns itself with the effects of local properties on global properties of graphs.

3.1 Testing subgraph freeness

We will focus on 3-clique freeness (triangles).

3.1.1 Szemerédi's regularity lemma

Notation:

If A, B are disjoint we denote $E(A, B)$ as the edges with one endpoint in A and the other in B.

If $A, B \subseteq V$ are disjoint and nonempty, we define the edge-density of (A, B) as $d(A, B) \stackrel{\text{def}}{=} \frac{E(A,B)}{|A| \cdot |B|}$.

We say that (A, B) is γ-regular if for every $A' \subseteq A; B' \subseteq B$ such that $|A'| \geq \gamma|A|$ and $|B'| \geq \gamma|B|$ it holds that $|d(A', B') - d(A, B)| \leq \gamma$.

For every $l \in \mathbb{N}$ and $\gamma > 0$, there exists $T = T(l, \gamma)$ such that for every sufficiently large $G = (V, E)$, there exists $t = [l, T]$ and t-partition of V, denoted $V_1, \ldots, V_t$ such that:

1. for all $i \in [t]$, each of the groups V_i are the same size (plus minus 1):

$$\left\lfloor \frac{|V|}{t} \right\rfloor \leq |V_i| \leq \left\lceil \frac{|V|}{t} \right\rceil$$

2. for all but at most γ-fraction of $\{i, j\} \in \binom{[t]}{2}$, it holds that v_i, v_j is γ-regular.

3.1.2 Triangle Freeness

There are at least $\rho(\varepsilon) \cdot k^3$ triangles in a graph which is ε-far from being triangle free, where k is the number of vertices in the graph. Trivially, there must be $\geq \frac{\varepsilon k^2}{l}$ triangles because otherwise we could remove one edge from each triangle and obtain a graph which is triangle free, but we know that we're ε-far.

Let $G = ([k], E)$ be a graph that is ε-far from being triangle free. Apply the regularity lemma with $\gamma = (0.1\varepsilon)$ and $l = 10/\varepsilon$. Let $V_1, \ldots, V_t$ denote the resulting partition.

First, we need to remove some of the edges from this partition. Omit all edges internal to V_i $\forall i \in [t]$. The number of edges we delete is bound:

$$\leq t \binom{\lceil k/t \rceil}{2} \leq \frac{k^2}{t} \leq \frac{k^2}{l} = 0.1k^2$$

Omit all the edges between pairs of sets which are not γ-regular. The number of edges thus deleted is also bound:

$$\leq \gamma \binom{t}{2} \left\lceil \frac{k}{t} \right\rceil^2 \cdot 0.2\varepsilon < 0.1\varepsilon k^2$$

Omit all edges between pairs of sets with edge density $< 0.2\varepsilon$. The number of edges we delete here is yet again bound:

$$\leq \binom{t}{2} \left\lceil \frac{k}{t} \right\rceil^2 \cdot 0.2\varepsilon$$

The resulting graph G' is still not triangle free.

A tester for triangles As per the theorem above, if we randomly sample $\Theta(1/\rho(\varepsilon))$ 3-tuples from a graph that is ε-far from being triangle free, then with probability $> 2/3$ one of them will be a triangle. As proof, note that the probability of sampling a random triangle is $\geq \rho(\varepsilon)k^3/k^3$ because there are fewer than k^3 possible 3-tuples. The probability that none of the $\Theta(1/\rho(\varepsilon))$ samples is a triangle is very very small:

$$(1 - \frac{1}{x})^x < \frac{1}{e}$$

References

[1] Oded Goldreich. *Introduction to Property Testing*. Cambridge University Press, 2017.