Lecture 7 (part II)

Testing dense graphs

- bipartiteness

Adjacency Matrix model

G represented by matrix A
s.t. can query A in
one step

$$A = \begin{bmatrix} & & \\ & A_{ij} & \\ & & \end{bmatrix}$$

Distance from property P:

<u>def</u> G is ε-far from P if must change $> \varepsilon \cdot n^2$
entries in A to turn G into member of P

Testing "sparse" properties:

	Graph type	max degree	natural representation	notion of distance
Previously	sparse			
Now	dense			

Bipartiteness:

- Can color nodes red/blue s.t. no edge monochromatic
- Can partition nodes into (V_1, V_2) s.t.

$$\nexists \underset{\substack{\parallel \\ (u,v)}}{e} \in E \quad \text{s.t.} \quad \begin{matrix} u, v \in V_1 \\ \text{or } u, v \in V_2 \end{matrix}$$

ε-far from bipartite: (definition)

- must remove $> \varepsilon \cdot n^2$ edges to make bipartite
- $\forall$ partitions $V = (V_1, V_2)$, $> \varepsilon \cdot n^2$ violating edges

Testing Algorithms:

- Testing <u>exact</u> bipartiteness;

- Proposed testing algorithm:

  - Pick sample of nodes of size $O\left(\frac{1}{\varepsilon^2}\log\frac{1}{\varepsilon}\right)$
  - Consider <u>induced</u> graph on sample
  - If bipartite, output PASS
    else output FAIL

A first attempt at a proof?

if G bipartite, induced graph is bipartite, so algorithm passes

if G ε-far from bipartite:

must remove εn^2 edges to make it bipartite

equivalently:

Lets try to use the "partition" defn of bipartiteness:

Algorithm 0

Pick $m = \Theta(?)$ random edge slots & query

$\forall$ partitions (V_1, V_2):

$\text{violating}_{(V_1,V_2)} \leftarrow$ # violating edges in sample wrt (V_1, V_2)

If $\forall (V_1, V_2)$ $\text{violating}_{(V_1,V_2)} > 0$ then output FAIL

else output PASS

Wait! How small should δ be?

Recall: All partitions are bad

But: if any partition "looks" good, the algorithm outputs PASS

Probability any partition "looks" good:

for one partition (V_1, V_2),

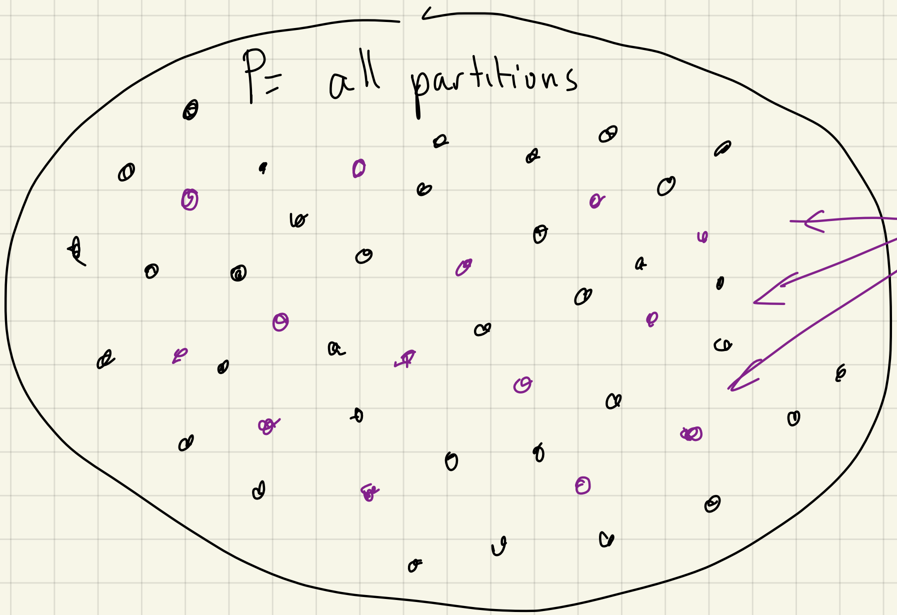
for all partitions (V_1, V_2),

Do we really need a union bnd?

Do we need to try all partitions?

(or Can we find few "representative" partitions that are close to all partitions?)

Plan: Consider small set of representatives



$|P| = 2^n$

$R \equiv$ purple points are "representatives"

Useful R satisfies:

$$\forall \; p \in P \quad \exists \; r \in R \quad \text{s.t.} \quad \text{dist}(p, r) \le \varepsilon$$

Hope: (1) testing R tells you something about P

(2) $|R| \ll |P|$

(so union bound isn't as bad)

Plan :

Find "representative" partitions s.t. all partitions are $\frac{\varepsilon}{2}$-close to some representative.

- if G ε-far from bipartite then $\forall$ partitions $\geq \varepsilon n^2$ violating edges

- if G bipartite then $\exists$ partition with 0 violating edges

Algorithm 1

1. pick U, U' randomly from V

U: $\Theta(\frac{1}{\varepsilon}\log\frac{1}{\varepsilon})$ nodes

U': $\Theta(\frac{1}{\varepsilon^2}\log\frac{1}{\varepsilon})$ nodes

If U not bipartite, FAIL

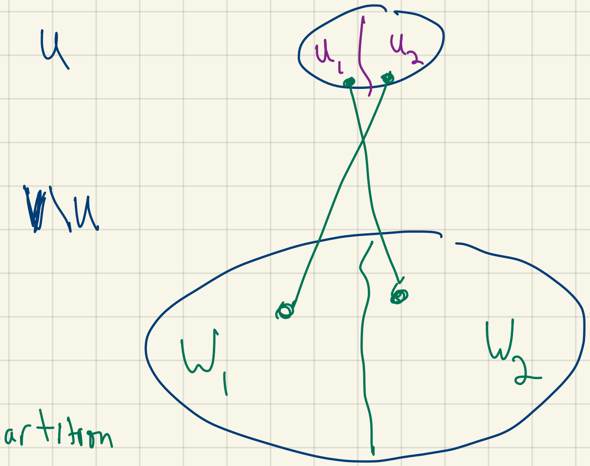
2. $\forall$ partitions of U into U_1, U_2:

   - induce partition $\overbrace{(U_1 \cup W_1, U_2 \cup W_2)}^{(Z_1, Z_2)}$ on whole graph:

     Partition: $\forall v \in V$ (including $v \in U$)
     - if v has nbr in U_2, put in W_1
     - if v " " " U_1, " " W_2
     - " " " " " both $\Rightarrow$ bad partition! continue to next partition
     - " " " " " neither, put in W_1

   - count $\#\{(u,v) \in P$ s.t. violate $(Z_1, Z_2)\}$
     if fraction $\leq 3/4\,\varepsilon$ PASS ~~+~~ halt
     else continue to next partition

3. FAIL



Algorithm 1

1. pick U, U' randomly from V

$\Theta(\frac{1}{\varepsilon}\log\frac{1}{\varepsilon})$ nodes — used to define set of partitions

$\Theta(\frac{1}{\varepsilon^2}\log\frac{1}{\varepsilon})$ nodes — used to define random edges: pair off $U' = \{u_1, v_1, u_2, v_2, \ldots\}$ to $P = \{(u_1,v_1), (u_2,v_2), \ldots\}$ pairs

If U not bipartite, FAIL

2. $\forall$ partitions of U into U_1, U_2: ← consider only $2^{|U|}$ partitions. enough?

   • induce partition $\overset{(Z_1, Z_2)}{(U_1 \cup W_1, U_2 \cup W_2)}$ on whole graph:

   Compute in time $O(|U|)$ → Partition: $\forall v \in V$ (including $v \in U$)

   don't do this for all $v \in V$, just those in U'

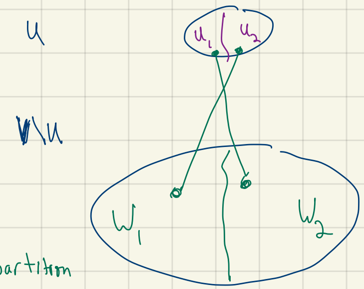
   - if v has nbr in U_2, put in W_1
   - if v " " " U_1, " " W_2
   - " " " " " both $\Rightarrow$ bad partition! continue to next partition
   - " " " " " neither, put in W_1

   • Count $\#\{(u,v) \in P$ s.t. violate $(Z_1, Z_2)\}$

   if fraction $\le 3/4\,\varepsilon$ PASS & halt
   else continue to next partition

   } why pass if see any violations? we aren't checking all partitions

3. FAIL



Query Complexity:

$$O\left(\frac{1}{\varepsilon^2}\log\frac{1}{\varepsilon}\right)$$

Time Complexity:

$$O\left(2^{\frac{1}{\varepsilon}\log 1/\varepsilon}\right)$$

← no dependence on n

can improve dependence on ε

Behavior: need to show that if G bipartite, likely to pass
& if G ε-far from bipartite, likely to Fail

if G is ε-far:

<u>all</u> partitions Z_1, Z_2 have $> \varepsilon \cdot n^2$ violating edges (so all those generated by algorithm do too!)

$\forall Z_1, Z_2$ $\Pr[$fraction of violating edges in P is $\le 3/4\,\varepsilon n^2] << \frac{1}{8 \cdot 2^{|U|}}$

$\Pr[\text{PASS}] = \Pr[$any partition of U generates partition that passes$] \le 2^{|U|} \cdot \frac{1}{8 \cdot 2^{|U|}} = \frac{1}{8}$

Algorithm 1

1. pick U, U' randomly from V

$U = \Theta(\frac{1}{\varepsilon}\log\frac{1}{\varepsilon})$ nodes, used to define set of partitions

$U' = \Theta(\frac{1}{\varepsilon^2}\log\frac{1}{\varepsilon})$ nodes, used to define random edges: pair off $U' = \{u_1, v_1, u_2, v_2, \ldots\}$ to $P = \{(u_1,v_1), (u_2,v_2), \ldots\}$ pairs

If U not bipartite, FAIL

2. $\forall$ partitions of U into U_1, U_2: ← consider only $2^{|U|}$ partitions. enough?

   - induce partition $\underbrace{(U_1 \cup W_1, U_2 \cup W_2)}$ on whole graph: with (Z_1, Z_2) written above it

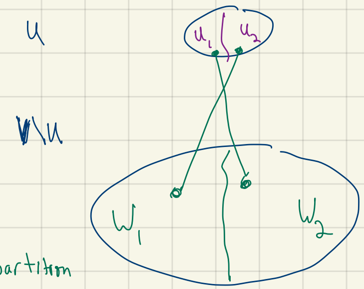
     Compute in time $O(|U|)$ → Partition: $\forall v \in V$ (including $v \in U$)
     - if v has nbr in U_2, put in W_1
     - if v " " " U_1, " " W_2
     - " " " " " both ⇒ bad partition! continue to next partition
     - " " " " " neither, put in W_1

     don't do this for all $v \in V$, just those in U'

   - Count $\#\{(u,v) \in P$ s.t. violate $(Z_1, Z_2)\}$
     if fraction $\le 3/4\,\varepsilon$ PASS ~~& halt~~
     else continue to next partition

   } why pass if see any violations? we aren't checking all partitions

3. FAIL



if G is bipartite:

does it pass?

Let (Y_1, Y_2) be bipartite partition.

$\#$ violating edges $= 0$

For sample U, partition according to Y_1, Y_2:

$U_1 \leftarrow U \cap Y_1$
$U_2 \leftarrow U \cap Y_2$

} partition of U

Comments

Can improve runtime to $\mathrm{poly}(1/\varepsilon)$

Proposed testing algorithm actually works

In adjacency list model (sparse graphs) need

$\Omega(\sqrt{n})$ queries,

Why more?

finer grain distinction

dense model: bipartite vs. $\varepsilon \cdot n^2$ edges need to be removed

sparse model: bipartite vs. $\varepsilon \cdot \Delta \cdot n$ edges need to be r