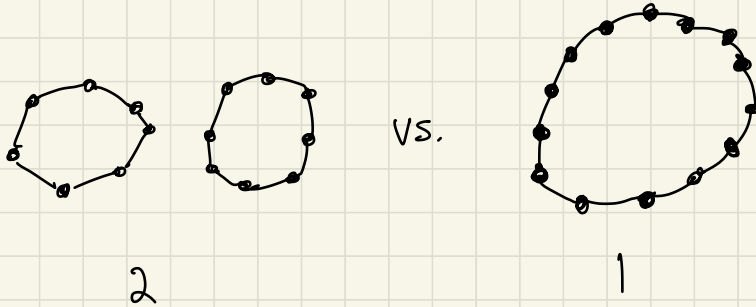


Lecture 2

Topics:

- finish estimate ave degree (see slides)
- Estimate number of connected components
- Estimate minimum spanning tree wt.

Estimating the number of connected components



need $\Omega(n)$ time

$\Rightarrow$ multiplicative approx needs
linear time.

what about additive approx?

Additively estimate # of Connected Components

Given: Graph G , max degree Δ ← Adjacency list representation
parameter ε

$|V| = n$
 $|E| = m$

Output: let $C = \#$ conn comp in G

output y st.

$$C - \varepsilon n \leq y \leq C + \varepsilon n$$

← additive εn -approximation



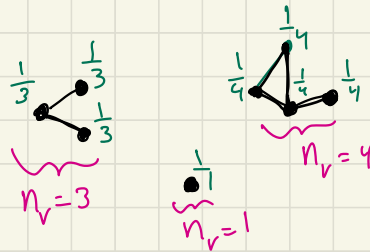
$C = 3$

Model: Adjacency list data structure

Main Insight:

A different characterization of # conn comp:

notation $\forall v$, let $n_v \equiv$ # nodes in v 's conn comp



observation: $\forall$ conn comp $A \subseteq V$

$$\sum_{u \in A} \frac{1}{n_u} = \sum_{u \in A} \frac{1}{|A|} = 1$$

So $\# \text{ conn comp } C = \sum_{u \in V} \frac{1}{n_u} = n \cdot (\text{ave val of } \frac{1}{n_u})$

Why better?

maybe not?

compute n_u

sum n

$O(n)$ time?

terms?

estimate

estimate

estimate

$\vdots$

Goal Estimate $C = \sum_{u \in V} \frac{1}{n_u}$

1) estimate $\frac{1}{n_u}$ quickly $\leftarrow$ additive error

2) estimate sum via sampling $\leftarrow$ additive error but implies multiplicative error
bnds

First:

Estimate $\frac{1}{n_u}$: (if it gets really small, we don't need to know more)

def $\hat{n}_u \equiv \min \{ n_u, 2/\epsilon \} \Rightarrow \frac{1}{\hat{n}_u} = \max \{ \frac{1}{n_u}, \frac{\epsilon}{2} \}$

$$\hat{C} = \sum_{u \in V} \frac{1}{\hat{n}_u}$$

why $\epsilon/2$?
why not 0?



Lemma $\forall u \quad \left| \frac{1}{\hat{n}_u} - \frac{1}{n_u} \right| \leq \epsilon/2$

Corr $|C - \hat{C}| \leq \frac{\epsilon n}{2}$

we'll see later !!

idea: if n_u is really big, can be hard to compute $\frac{1}{n_u}$ is small $\leftarrow \leq \epsilon/2$

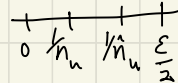
$\Rightarrow$ never need $> 2/\epsilon$ BFS steps to compute $\hat{n}_u$ to within $\epsilon/2$

Proof of lemma:

if $n_u < 2/\varepsilon$ then $\hat{n}_u = n_u$ so $\frac{1}{\hat{n}_u} - \frac{1}{n_u} = 0$

if $n_u > 2/\varepsilon$ then $\hat{n}_u = 2/\varepsilon$ so

$$0 < \frac{1}{n_u} < \frac{1}{\hat{n}_u} = \frac{\varepsilon}{2}$$



$$\Rightarrow \left| \frac{1}{n_u} - \frac{1}{\hat{n}_u} \right| \leq \frac{\varepsilon}{2}$$

□

How do we compute $\hat{n}_u$?

Algorithm:

Do BFS from u until

- OR
- visit whole component of u
 - visit $2/\varepsilon$ distinct nodes

↖ $\hat{n}_u \leftarrow n_u$

↖ $\hat{n}_u \leftarrow 2/\varepsilon$

Output $\#$ visited nodes

runtime:

$$O(\Delta \cdot \frac{1}{\varepsilon}) \leftarrow \begin{matrix} \uparrow \\ \text{time per step} \\ \text{of BFS} \end{matrix} \quad \text{max } \# \text{ of nodes visited}$$

WAIT: actually, $O(1/\varepsilon^2)$ suffices

How to estimate $\sum_u \frac{1}{\hat{n}_u}$?

Algorithm estimate $\hat{c}$

b is a constant

$$r \leftarrow b/\varepsilon^3$$

Choose $U = \{u_1, \dots, u_r\}$ random nodes

$$\forall u_i \in U$$

compute $\hat{n}_{u_i}$ via above algorithm

$$\text{Output } \tilde{c} = \frac{n}{r} \sum_{u \in U} \frac{1}{\hat{n}_u}$$

average value
of sample U
where $|U|=r$

runtime:

$$O\left(\frac{1}{\varepsilon^2} \cdot \frac{1}{\varepsilon^3}\right) = O\left(\frac{1}{\varepsilon^5}\right)$$

$\sim$
algorithm
to compute $\hat{n}_u$

Theorem:

$$\Pr [|\tilde{C} - \hat{C}| \leq \frac{\epsilon n}{2}] \geq 3/4$$

Proof

Chernoff Bound:

Given $X_1, \dots, X_r$ iid $X_i \in [0, 1]$

Let $S = \sum_{i=1}^r X_i$ $p = E[X_i] = E[S]/r$

Then: $\Pr [|\frac{S}{r} - p| \geq \delta p] \leq e^{-n(r\delta^2)}$

here, let $X_i = \frac{1}{\hat{n}_{u_i}}$ for each sampled $u_i \in \mathcal{U}$

$$\text{so } \frac{S}{r} = \frac{\sum_{i=1}^r X_i}{r} = \frac{1}{r} \sum_{i=1}^r \frac{1}{\hat{n}_{u_i}} \quad \text{recall } \tilde{C}/n - \text{Output} = \tilde{C}$$

$$\text{let } p = E[X_i] = \frac{1}{n} \sum_{u \in V} \frac{1}{\hat{n}_{u_i}} = \frac{\hat{C}}{n}$$

$$\text{let } \delta = \frac{\epsilon}{2}$$

recall

$$r = b/\varepsilon^3$$

$$1 \leq \hat{n}_u \leq 2/\varepsilon \quad (\text{from definition})$$

$$\text{so} \quad 1 \geq \frac{1}{\hat{n}_u} \geq \varepsilon/2$$

$$\Rightarrow n \geq \sum_{u \in V} \frac{1}{\hat{n}_u} \geq \frac{\varepsilon n}{2}$$

$\underbrace{\hspace{1cm}}_{\equiv \hat{c}}$

$$\text{so} \quad \left(\frac{\varepsilon}{2} \leq \frac{\hat{c}}{n} \right) \leq 1$$

$$\Pr[|\tilde{c} - \hat{c}| \geq \frac{\varepsilon n}{2}] \leq \Pr[|\tilde{c} - \hat{c}| \geq \frac{\varepsilon}{2} \cdot \hat{c}]$$
$$= \Pr\left[\left|\frac{\tilde{c}}{n} - \frac{\hat{c}}{n}\right| \geq \frac{\varepsilon}{2} \cdot \frac{\hat{c}}{n}\right]$$

$$\leq e^{-\Omega(rp\delta^2)}$$

$$= e^{-\Omega(b/\varepsilon^3 \cdot \frac{\hat{c}}{n} \cdot \frac{\varepsilon^2}{4})} \leq e^{-\Omega(b)}$$

this is
why we
set $\frac{1}{\hat{n}_u}$ to $\frac{\varepsilon}{2}$
instead of 0

$$\geq \frac{\varepsilon}{2}$$

↑
pick
const
 b big
enough
so that
 $e^{-\Omega(b)} \leq 1/4$



Finishing up:

We had

$$|C - \hat{C}| \leq \frac{\varepsilon}{2} \cdot n \quad \text{by earlier corollary}$$

$$\text{now we have } |\hat{C} - \tilde{C}| \leq \frac{\varepsilon}{2} n$$

$\Downarrow$

$$|C - \tilde{C}| \leq \varepsilon n \quad \text{by } \Delta \neq$$

$\uparrow$ $\uparrow$
output
conn
comp



Approximate Min Spanning Tree (MST)

Input (1) $G = (V, E)$ adjacency list representation

$$n = |V|$$

max degree Δ

each edge has weight

$$w_{uv} \in \{1..w\} \cup \{\infty\}$$

Connected

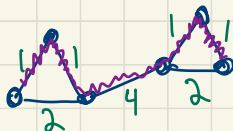
$\uparrow$ ie, $w_{uv} \notin E$
assumption that
edges integral
not necessary

(2) ε

Output

$$\text{let } M = \min_{\substack{T \text{ spans } G \\ \text{tree}}} \{ \underbrace{wt(T)}_{\sum_{(i,j) \in T} w_{ij}} \}$$

$$\text{output } \hat{M} \text{ st. } (1-\varepsilon) \cdot M \leq \hat{M} \leq (1+\varepsilon) \cdot M$$



assumption on wts $\Rightarrow n-1 \leq wt(T) \leq w(n-1)$

A different characterization of MST:

def $E^{(i)} = \{ (u,v) \mid w_{u,v} \in \{1..i\} \}$ edges of $w \leq i$

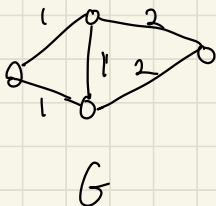
$$G^{(i)} = (V, E^{(i)})$$

$$C^{(i)} = \# \text{ conn comp of } G^{(i)}$$

Some examples:

1) $w=1$ only size 1 wts + connected
(by assumption)
here $M = n-1$

2) $w=2$ weights $\in \{1,2\}$



$$G^{(1)}$$
$$C^{(1)} = 2$$

idea for MST:

Use as many wt 1 edges as you can
then only need wt 2 edges to connect
current components

$\Rightarrow$ need $c^{(1)} - 1$ wt 2 edges

Total MST wt:

$$M = 1 \cdot (\# \text{ wt 1 edges}) + 2 \cdot (\# \text{ wt 2 edges})$$

$$= (n-1) + \underbrace{\# \text{ wt 2 edges}}_{c^{(1)} - 1}$$

$$= n-2 + c^{(1)}$$

let $\alpha_i = \#$ edges of wt i in any MST of G

Fact $\forall G$, α_i 's are the same in
any MST of G

Proof omitted.

(follows from correctness of Kruskal's
algorithm)

Claim $M = n - w + \sum_{i=1}^{w-1} C^{(i)}$

Pf.

$$\sum_{i=1}^w \alpha_i = n-1 \quad \text{since } \# \text{ edges in any MST} = n-1$$

$$\sum_{i \geq l} \alpha_i = \# \text{ conn comp of } G^{(l)} - 1$$

$$= C^{(l)} - 1$$

$$C^{(0)} = n$$

$$M = \sum_{i=1}^w i \cdot \alpha_i$$

$$= \sum_{i=1}^w \alpha_i + \sum_{i=2}^w \alpha_i + \dots + \sum_{i=w}^w \alpha_i$$

$\underbrace{\alpha_w}_{\alpha_w}$

$$= \underbrace{C^{(0)} - 1}_{n-1} + C^{(1)} - 1 + C^{(2)} - 1 + \dots + C^{(w-1)} - 1$$

$$= n - w + \sum_{i=1}^{w-1} C^{(i)}$$



Approximation algorithm:

For $i=1$ to $w-1$

$\hat{C}^{(i)} \leftarrow$ approx # conn comp of $G^{(i)}$ to within

$\frac{\varepsilon}{2w} \cdot n$ additive error

Output $\hat{M} = n - w + \sum_{i=1}^{w-1} \hat{C}^{(i)}$ let $\varepsilon' = \frac{\varepsilon}{2w}$

Runtime:

$\tilde{O}\left(\frac{1}{(\varepsilon')^5}\right) = O\left(\text{poly}\left(\frac{w}{\varepsilon}\right)\right)$ per call to approx-CC

Total: $\tilde{O}\left(\text{poly}\left(\frac{w}{\varepsilon}\right) \cdot \Delta\right)$

how do you work with $G^{(i)}$?

since calling approx-CC on subgraphs of G , can't sure factor of Δ in approx-CC

Can improve to $O\left(\frac{\Delta w}{\varepsilon^2} \log \frac{\Delta w}{\varepsilon}\right)$
need $\Omega\left(\frac{\Delta w}{\varepsilon^2}\right)$

No dependence on n !!!

Approximation guarantee:

Call approx #cc with "failure" prob $\delta \leq \frac{1}{4 \cdot w}$

how?
HW 0

$$\Pr[\text{all calls give } \varepsilon' n \text{ approx}] \geq 1 - \frac{w}{4 \cdot w}$$

union bnd

If $\uparrow$ happens, $|M - \hat{M}| \leq w \cdot \frac{\varepsilon}{2w} \cdot n = \frac{\varepsilon n}{2}$

additive error bnd

since $M \geq n-1 \geq \frac{n}{2}$

$$|M - \hat{M}| \leq \varepsilon M$$

mult error bnd

Further questions:

Which edges are on MST?

What if w is big?