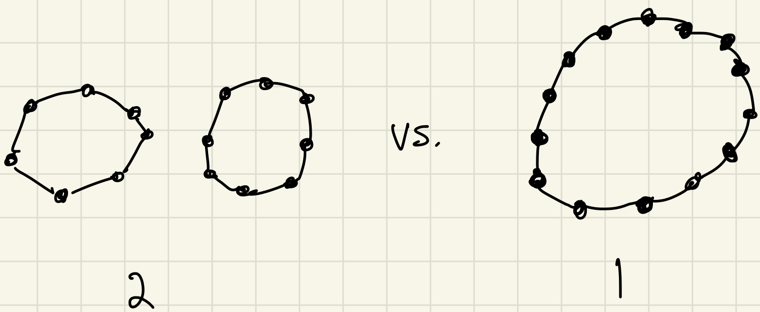
Lecture 3

Topics:

- 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

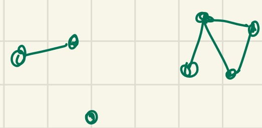
Given: Graph G, max degree Δ ⟸ Adjacency list representation
parameter ε

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

Output: let C = # conn comp in G

output y s.t.

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



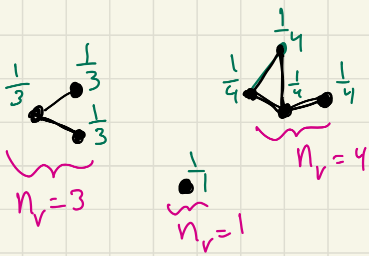
$C = 3$

Model: Adjacency list data structure

Main Insight:

A different characterization of # conn comp:

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



<u>observation</u>: $\forall$ conn comp $A \subseteq V$

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

So $\boxed{\text{\# conn comp } c = \sum_{u \in V} \frac{1}{n_u}}$

Why better?

maybe not?

Compute n_u $O(n)$ time?

sum n terms?

estimate
estimate
estimate

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

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

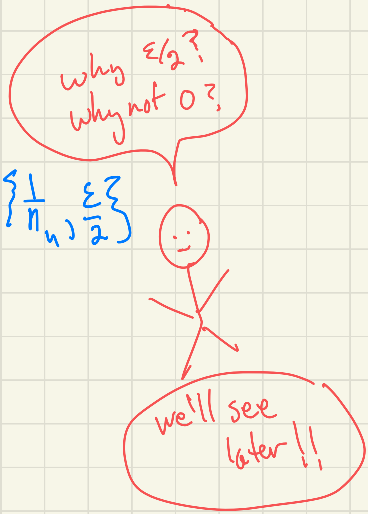
2) estimate sum via sampling bnds $\leftarrow$ additive error but implies multiplicative error

First:

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

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

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



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

Corr $|c - \hat{c}| \leq \frac{\varepsilon n}{2}$

idea: if n_u is really big ($> 2/\varepsilon$), can be hard to compute

but in this case $\frac{1}{n_u}$ is small $\leftarrow \leq \varepsilon/2$

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

$n_v \equiv$ # nodes in v's conn comp

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

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

conn comp $c = \sum_{u \in V} \frac{1}{n_u}$

def $\hat{n}_u \equiv \min\{n_u, 2/\varepsilon\}$

$$\hat{c} \equiv \sum_{u \in V} \frac{1}{\hat{n}_u}$$

Lemma $\forall u$ $\left|\frac{1}{\hat{n}_u} - \frac{1}{n_u}\right| \le \varepsilon/2$

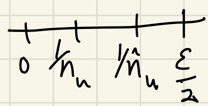
Corr $|c - \hat{c}| \le \frac{\varepsilon n}{2}$

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

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



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

$\blacksquare$

$n_v \equiv$ # nodes in v's conn comp

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

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

conn comp $c = \sum_{u \in V} \frac{1}{n_u}$

def $\hat{n}_u \equiv \min\{n_u, 2/\varepsilon\}$

$$\hat{c} \equiv \sum_{u \in V} \frac{1}{\hat{n}_u}$$

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

Corr $|c - \hat{c}| \leq \frac{\varepsilon n}{2}$

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

Algorithm:

Do BFS from u until

- visit whole component of u — $\hat{n}_u \leftarrow n_u$

OR

- visit $2/\varepsilon$ distinct nodes — $\hat{n}_u \leftarrow 2/\varepsilon$

Output # visited nodes

runtime: $O(\underset{\substack{\uparrow \\ \text{time per step of BFS}}}{\Delta} \cdot 1/\varepsilon) \leftarrow$ max # of nodes visited

actually, $O(\Delta + 1/\varepsilon)$ suffices ... why?

(we will come back to this)

$n_v \equiv$ # nodes in v's conn comp

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

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

conn comp $c = \sum_{u \in V} \frac{1}{n_u}$

$= n \cdot \left(\text{average value of } \frac{1}{n_u}\right)$

def $\hat{n}_u \equiv \min \{ n_u, 2/\varepsilon \}$

$$\hat{c} \equiv \sum_{u \in V} \frac{1}{\hat{n}_u}$$

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

Corr $|c - \hat{c}| \leq \frac{\varepsilon n}{2}$

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

Output $$\tilde{c} = \underbrace{\frac{n}{r} \sum_{u \in U} \frac{1}{\hat{n}_u}}_{\text{average value of sample } U \text{ where } |U| = r}$$

runtime: $$O\Big(\underbrace{\frac{\Delta}{\varepsilon}}_{\text{algorithm to compute } \hat{n}_a} \cdot \frac{1}{\varepsilon^3}\Big) = O\left(\frac{\Delta}{\varepsilon^4}\right)$$

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

$n_v \equiv$ # nodes in v's conn comp

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

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

conn comp $c = \sum_{u \in V} \frac{1}{n_u}$

Theorem:

$$\Pr\left[|\tilde{c} - \hat{c}| \le \frac{\varepsilon n}{2}\right] \ge 3/4$$

Proof

Chernoff Bound:

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

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

Then: $\Pr\left[\left|\frac{S}{r} - p\right| \ge \delta p\right] \le e^{-\Omega(rp\delta^2)}$

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

$$\text{so } \frac{S}{r} = \frac{\sum_{i=1}^{r} X_i}{r} = \underbrace{\frac{1}{r} \sum_{i=1}^{r} \frac{1}{\hat{n}_{u_i}}}_{\tilde{c}/n}$$

recall — Output $= \tilde{c}$

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

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

recall $r = b/\varepsilon^3$

$$1 \le \hat{n}_u \le 2/\varepsilon$$ (from definition)

so $$1 \ge \frac{1}{\hat{n}_u} \ge \varepsilon/2$$

$$\Longrightarrow \quad n \ge \underbrace{\sum_{u \in V} \frac{1}{\hat{n}_u}}_{\equiv \hat{C}} \ge \frac{\varepsilon n}{2}$$

so $$\frac{\varepsilon}{2} \le \frac{\hat{C}}{n} \le 1$$

$$\Pr\left[|\tilde{C} - \hat{C}| \ge \frac{\varepsilon n}{2}\right] \le \Pr\left[|\tilde{C} - \hat{C}| \ge \frac{\varepsilon}{2} \cdot \hat{C}\right]$$

$$= \Pr\left[\left|\frac{\tilde{C}}{n} - \frac{\hat{C}}{n}\right| \ge \frac{\varepsilon}{2} \cdot \frac{\hat{C}}{n}\right]$$

$$\le e^{-\Omega(r p \delta^2)}$$

$$= e^{-\Omega\left(b/\varepsilon^{\cancel{3}} \cdot \frac{\hat{C}}{n} \cdot \frac{\varepsilon^{\cancel{2}}}{4}\right)} \le e^{-\Omega(b)}$$

$\ge \frac{\varepsilon}{2}$

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

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

$\blacksquare$

Finishing up:

We had

$$|c - \hat{c}| \le \frac{\varepsilon}{2} \cdot n \qquad \text{by earlier corollary}$$

now we have $|\hat{c} - \tilde{c}| \le \frac{\varepsilon}{2} n$

$$\Downarrow$$

$$\Big|\underset{\substack{\uparrow \\ \#\text{ conn comp}}}{c} - \underset{\substack{\uparrow \\ \text{output}}}{\tilde{c}}\Big| \le \varepsilon n \qquad \text{by } \Delta \neq$$

∎

Approximate Min Spanning Tree (MST)

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

$n = |V|$

max degree Δ

each edge has weigt

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

ie. $w_{uv} \notin E$

Connected

assumption that edges integral not necessary

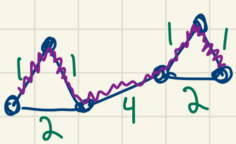
(2) ε

Output

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

tree

output $\hat{M}$ s.t. $(1-\varepsilon)\cdot M \le \hat{M} \le (1+\varepsilon)\cdot M$



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

A different characterization of MST:

<u>def</u> $E^{(i)} = \{ (u,v) \mid w_{u,v} \in \{1..i\} \}$ edges of wt $\le i$

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

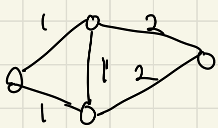
$C^{(i)} = \#$ conn comp of $G^{(i)}$

Some examples:

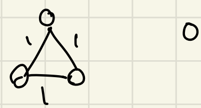
1) $w = 1$ only size 1 wts + connected (by assumption)

here $M = n-1$

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



G



$G^{(1)}$

$C^{(1)} = 2$

$$E^{(i)} = \{(u,v) \mid w_{u,v} \in \{1..i\}\}$$

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

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

Kruskal's idea for MST:
(paraphrased)

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

$\Longrightarrow$ need $C^{(1)} - 1$ wt 2 edges

Total MST wt:

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

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

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

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

Kruskal's $\Rightarrow$
this is <u>same</u>
$\forall$ MSTs why?

<u>Pf</u>. let α_i = # edges of wt i in <u>any</u> MST of G

$$\sum_{i>\ell} \alpha_i = \#\text{ conn comp of } G^{(\ell)} - 1$$
$$= C^{(\ell)} - 1$$

$$E^{(i)} = \{ (u,v) \mid w_{u,v} \in \{1 .. i\} \}$$
$$G^{(i)} = (V, E^{(i)})$$
$$C^{(i)} = \#\text{ conn comp of } G^{(i)}$$

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

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

$$M = \sum_{i=1}^{w} i \cdot \alpha_i$$
$$= \sum_{i=1}^{w} \alpha_i + \sum_{i=2}^{w} \alpha_i + \dots + \underbrace{\sum_{i=w}^{w} \alpha_i}_{\alpha_w}$$
$$= \underbrace{C^{(0)} - 1}_{n-1} + C^{(1)} - 1 + C^{(2)} - 1 + \dots + C^{(w-1)} - 1$$
$$= n - w + \sum_{i=1}^{w-1} C^{(i)}$$

∎

$E^{(i)} = \{(u,v) \mid w_{u,v} \in \{1..i\}\}$

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

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

Approximation algorithm:

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

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

let $\varepsilon' = \frac{\varepsilon}{2w}$

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

Runtime:

$\tilde{O}\left(\frac{\Delta}{(\varepsilon')^4}\right) = \tilde{O}\left(\frac{\Delta w^4}{\varepsilon^4}\right)$ per call to $\underbrace{\text{approx-CC}}_{\text{how do you work with } G^{(i)}?}$

Total: $\tilde{O}\left(\frac{\Delta w^5}{\varepsilon^4}\right)$

since calling approx-CC on subgraphs of G, can't save 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!!!

$E^{(i)} = \{(u,v) \mid w_{u,v} \in \{1..i\}\}$

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

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

Approximation algorithm:

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

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

let $\varepsilon' = \frac{\varepsilon}{2w}$

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

Approximation guarantee:

how? HW 0

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

union bnd

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

If $\uparrow$ happens, $|M - \hat{M}| \le w \cdot \frac{\varepsilon}{2w} \cdot n = \frac{\varepsilon n}{2}$ ← additive error bnd

since $M \ge n - 1 \ge \frac{n}{2}$ $\quad |M - \hat{M}| \le \varepsilon M$ ← mult error bnd

Which edges are on MST?

What if w is big?