

Lecture 2, part 1: Average degrees (final words),

Ronitt Rubinfeld

MIT

Estimating degrees

- Given: graph G , $m = \Omega(n)$ edges, with
 - Degree queries: on v , return $\deg(v)$
 - Neighbor queries: on (v, j) return j th neighbor of v
- Output: estimate of average degree $\bar{d} = \frac{2m}{n}$
 - i.e., output $\tilde{d}$ s.t. $\Pr[|\tilde{d} - \bar{d}| \leq \epsilon \bar{d}] \geq 1 - \delta$

Naïve sampling

- Pick s vertices $v_1, \dots, v_s$
- Output average degree of sample $\frac{1}{s} \sum_i \deg(v_i)$

Needs $\Omega(n)$ samples

Lower bound

$\sqrt{n}$ queries

(Total) ordering on vertices

- $u < v$ if:
 - $\deg(u) < \deg(v)$
 - OR $\deg(u) = \deg(v)$ and $ID(u) < ID(v)$. (assuming unique ID's)
- Orient edges from small to large
- $\deg^+(u)$ = number of nbrs of u such that $u < v$. (outdegree)
- $\sum_{u \in V} \deg^+(u) = m = \frac{n}{2} \cdot \bar{d}$

Idea: Lower variance by counting degree from smaller side of each edge

algorithm

Algorithm naïve:

- $k \leftarrow ??$
- For $i = 1$ to k do:
 - Pick $v_i \in_U V$
 - Let $X_i \leftarrow \deg(v_i)$
- Output $\tilde{d} \leftarrow \frac{1}{k} \sum_{i=1}^k X_i$

New algorithm:

- $k \leftarrow \frac{16}{\epsilon^2} \cdot \sqrt{n}$
- For $i = 1$ to k do:
 - Pick $v_i \in_U V$
 - Pick $u_i \in_U N(v_i)$
 - If $v_i < u_i$ then $X_i \leftarrow 2 \deg(v_i)$
 - Else $X_i \leftarrow 0$
- Output $\tilde{d} \leftarrow \frac{1}{k} \sum_{i=1}^k X_i$

Sample deg+ by
"correcting"
sampling of deg

Lemmas from last time

- $\forall v \in V \deg^+(v) \leq \sqrt{2m}$
- $E[X_i] = \bar{d}$
- $\text{Var}[X_i] \leq 4 \sqrt{2m} \cdot \bar{d}$

Two useful facts about variance:

- Chebysev's Inequality: $\Pr[|X - E[X]| \geq b] \leq \frac{\text{Var}[X]}{b^2}$
- Lemma: Let $Y = \frac{1}{k} \sum_{i=1}^k X_i$ where X_i 's iid. Then $\text{Var}[Y] = \frac{1}{k} \text{Var}[X_i]$

Lemmas from last time

- $\forall v \in V \deg^+(v) \leq \sqrt{2m}$
- $E[X_i] = \bar{d}$
- $Var[X_i] \leq 4\sqrt{2m} \cdot \bar{d}$ so by lemma $Var[\tilde{d}] \leq 4\frac{\sqrt{2m}}{k} \cdot \bar{d}$

- By Chebyshev: $\Pr[|\tilde{d} - \bar{d}| \geq \epsilon \bar{d}] \leq \frac{var[\tilde{d}]}{(\epsilon \bar{d})^2} \leq \dots \leq \frac{1}{4}$

Good estimate
with prob $\geq 3/4$

Since $\bar{d} = E[d]$

Since $\bar{d} = \frac{2m}{n}$ and by
choice of $k = 16\sqrt{n}/\epsilon^2$