Today: continue estimating the number of edges

Recall:

Vertices partitioned into buckets $B_0, B_1, B_2, \ldots$

$$B_i = \left\{ v \in V : \left(1+\frac{\varepsilon}{10}\right)^i \leq \deg(v) < \left(1+\frac{\varepsilon}{10}\right)^{i+1} \right\}$$

Number of buckets $= k = O\left(\frac{1}{\varepsilon}\log n\right)$

Want to categorize buckets as

- heavy if $|B_i| > T$
- light if $|B_i| \leq T$

T = some threshold

Three types of edges:

1. vertex in heavy bucket to vertex in heavy bucket
2. heavy light
3. light light

$H = \{ v \in V : v \text{ in heavy bucket} \}$

$L = \{ v \in V : v \text{ in light bucket} \}$

$$\deg(v) = \underbrace{d_H(v)}_{\text{\# of } v\text{'s neighbors in } H} + \underbrace{d_L(v)}_{\text{\# of } v\text{'s neighbors in } L}$$

$$\#\text{edges} = \frac{1}{2}\left(\underbrace{\sum_{v \in H} d_H(v)}_{①} + \underbrace{\underbrace{\sum_{v \in H} d_L(v)}_{} + \underbrace{\sum_{v \in L} d_H(v)}_{}}_{②} + \underbrace{\sum_{v \in L} d_L(v)}_{③} \right)$$

equal

Idea:

- learn which buckets are heavy and which are light
- estimate the first two sums:
  - edges ③ ignored ← negligible number
  - edges ② counted only once ← factor 2 approximation

$\boxed{21-2}$

Algorithm:

- set $T = \frac{\varepsilon^{3/2} n^{1/2}}{C_1 \log n}$

$\boxed{C_1, C_2 \text{ - sufficiently large constants}}$

- sample $s = C_2 \cdot \left(\frac{n}{T}\right) \cdot \frac{1}{\varepsilon^2} \cdot \log n$ vertices with replacement
- For each i, let s_i = # samples in bucket i
- For each bucket B_i:

  if $\frac{s_i}{s} \cdot n \geq T$, mark it heavy

  otherwise, mark it light
- Output

$$\sum_{\substack{i \\ B_i \text{ heavy}}} \frac{s_i}{s} \cdot n \cdot \left(1+\frac{\varepsilon}{10}\right)^i$$

Total queries: $O\left(\frac{\sqrt{n}}{\varepsilon^{O(1)}} \log^{O(1)} n\right)$

Why this works with probability $\frac{99}{100}$:

- Buckets B_i with $|B_i| < T/2$: classified as light

$\boxed{21-3}$

– Buckets B_i with $|B_i| \geq T/2$:

$\frac{s_i}{s} \cdot n$ is a $\left(1+\frac{\varepsilon}{10}\right)$- approximation of $|B_i|$

This follows via the Chernoff bound + the union bound

If this holds:

$$\#\text{edges } ③ \leq \Big(2 \cdot T \cdot \overbrace{O(\varepsilon^{-1} \log n)}^{\#\text{buckets}}\Big)^2 \leq \underbrace{\frac{\varepsilon}{100} \cdot n}$$

(Light-to-light)

No isolated vertices $m \geq n/2$

$$m\left(1-\frac{\varepsilon}{10}\right) \leq \frac{m}{1+\frac{\varepsilon}{10}} \leq \frac{1}{2} \sum_i |B_i| \left(1+\frac{\varepsilon}{10}\right)^i \leq m$$

$$\frac{1}{2} m\left(1-\frac{\varepsilon}{10}\right) - \frac{\varepsilon m}{50} \leq \frac{1}{2} \sum_{\substack{i \\ B_i \text{ heavy}}} |B_i| \left(1+\frac{\varepsilon}{10}\right)^i \leq m$$

$$\vee\!/$$

$$\frac{1}{2} m\left(1-\frac{\varepsilon}{5}\right)$$

$$\frac{1}{2} m\left(1-\frac{\varepsilon}{5}\right)\left(1-\frac{\varepsilon}{10}\right) \leq \underbrace{\frac{1}{2} \sum_{\substack{i \\ B_i \text{ heavy}}} \left(\frac{s_i}{s} \cdot n\right)\left(1+\frac{\varepsilon}{10}\right)^i}_{\text{the algorithm's output}} \leq m\left(1+\frac{\varepsilon}{10}\right)$$

$$\vee\!/$$

$$\frac{1}{2} m (1-\varepsilon)$$

$(2+\varepsilon)$- approximation

$\boxed{21-4}$

Getting $1+\varepsilon$ approximation:

- So far limited to checking degrees of random vertices
- Now also neighbor queries allowed
- Counter the loss of half the weight of edges of type 2
- For each sampled vertex in heavy bucket, select random neighbor and see if it's in heavy or light bucket
- For bucket i, let δ_i = fraction of observed neighbors in light buckets
- Output

$$\sum_{\substack{i \\ B_i \text{ heavy}}} \frac{s_i}{s} \cdot n \cdot (1+\delta_i) \cdot \left(1-\frac{\varepsilon}{10}\right)^i$$

Good estimate via Hoeffding's inequality