LECTURE 9

THM (Irregularity of sparse random graphs)

> $\exists$ abs. const $c>0$ such that if $d < c \log N$,
> $G(n,p)$ has a vertex with too large degree with high prob.: ← "hub"
> $$\mathbb{P}\{\exists i : \deg(i) \geq 10d\} \geq 0.9$$

Proof $\deg(i) \sim$ Binomial with mean d

$\Rightarrow$ By the reverse Chernoff's inequality (Hw 3 Problem 3),

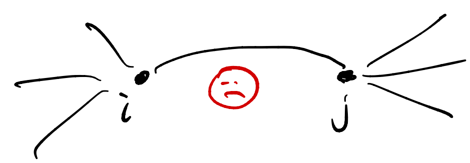
- $\mathbb{P}\{\underbrace{\deg(i) \geq 10d}_{E_i}\} \geq e^{-\mu}\left(\frac{\mu}{t}\right)^t$ (with $\mu = d$, $t = 10d$)

$$= e^{-d}\left(\frac{\cancel{d}}{10\cancel{d}}\right)^{10d} \geq e^{-20d} \geq e^{-20c\log N} \geq \frac{100}{N}$$ (choose $c>0$ a small const.)

- $$\mathbb{P}\left(\bigcup_{i=1}^{N} E_i\right) = 1 - \mathbb{P}\left(\bigcap_{i=1}^{N} E_i^c\right) = 1 - \prod_{i=1}^{N} \mathbb{P}(E_i^c)$$ (independence)

$$\geq 1 - \prod_{i=1}^{N}(1-\mathbb{P}(E_i)) \geq 1 - \left(1 - \frac{100}{N}\right)^N \geq 0.9$$ (DIY)

ERROR! $\deg(i)$ are NOT independent

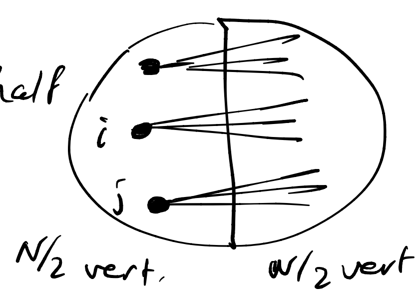


FIX: Pass to a bipartite subgraph of G:

split the vertices into two halves:

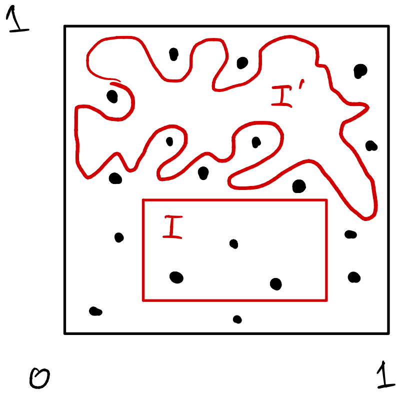
HW Count only the edges from left half

These degrees are independent



QED

DISCREPANCY

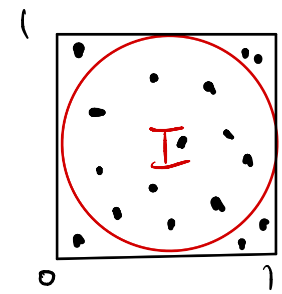


- Throw N random points into the square $[0,1)^2$ independently and uniformly.
- $\forall$ subset $I \subset [0,1)^2$, the expected # pts in I is proportional to the <u>area of I</u>

- Why? # pts in I: $N_I \sim \text{Binom}(N, \lambda)$

$$\Rightarrow \mathbb{E} N_I = N\lambda.$$

- <u>Ex</u>: probabilistic computation of π:



$$\mathbb{E}\left[\frac{N_I}{N}\right] = \text{area}(I) = \pi r^2 = \frac{\pi}{4}.$$

- <u>Error</u>: $\quad Var(N_I) = N\lambda(1-\lambda) \leq N\lambda$

$\Rightarrow$ by Chebyshev's inequality, with prob. ≥ 0.99:

$$\boxed{N_I = N\lambda \pm C\sqrt{N\lambda}} \qquad (*)$$

& Chernoff improves the confidence.

- Is this true for all I simultaneously, i.e. with prob. ≥ 0.99, $(*)$ holds $\forall I \subset [0,1)^2$?

- **NO**: see I' on the pic above.

- $\Rightarrow$ difference between "<u>instance</u>" and "<u>uniform</u>" results:

(a) $\forall I$, w.h.p. $(*)$ holds — TRUE

(b) w.h.p, $\forall I$ $(*)$ holds — FALSE.

- Does the result hold uniformly over rectangles I?

YES (almost)

THM (Discrepancy) A set of N independent uniform random pts in $[0,1]^2$ satisfies the following with probability ≥ 0.99.

$\forall$ axis-aligned rectangle $I \subset [0,1]$ we have:

$$\lambda_I N - C\sqrt{\lambda_I N \log N} \leq N_I \leq \lambda_I N + C\sqrt{\lambda_I N \log N},$$

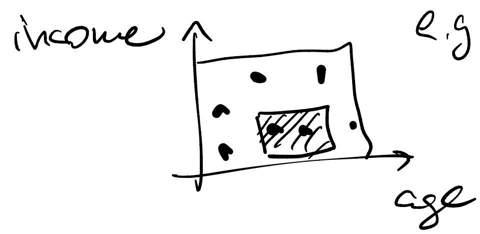
as long as LHS ≥ 0 and $N \geq C_1$ (LHS stands for "left-hand side")

Here: N_I = # points in I,

λ_I = area of I,

C, C_1 = absolute constants (large enough)

- Relevance for statistic: representative sampling.



e.g. We want a sample in which all age brackets and income brackets are fairly represented

- Difficulty: need to control all rectangles I at once, ∞ many of them (hard to take union bound)

- More on discrepancy theory (ergodic etc.)