

Graph Mining Class Notes on Feb.6 2007

Xiang Yang

Department of Computer Science, Kent State University, Kent, Ohio, U.S.A.,

General background:

Assume $X \sim B(n, 1/2)$ then $X = \sum_{i=1}^n Y_i$. Y_i is Bernoulli trial.

$$E(Y_i) = 1/2, \text{var}(Y_i) = 1/4$$

$$E(X) = E[\sum_{i=1}^n Y_i] = n \cdot E[Y_i] = n/2$$

1 Markov Inequality

$$P(X \geq a) \leq \frac{E[X]}{a}. \quad X \text{ is a random variable and } X \geq 0$$

$$\begin{aligned} \text{Proof. } E[X] &= \sum_{i=1}^n i \cdot P(X = i) \\ &= \sum_{i < a} i \cdot P(X = i) + \sum_{i \geq a} i \cdot P(X = i) \\ &\geq \sum_{i < a} i \cdot P(X = i) + \sum_{i \geq a} a \cdot P(X = i) \\ &\geq 0 + \sum_{i \geq a} i \cdot P(X = i) \\ &= a \cdot P(X \geq a) \end{aligned}$$

If $X \geq 0$ and X is integer and $E[X] < 1$ then we also have $P(X > 0) \leq E[X]$

2 Chebyshev Inequality

$$P(|X - E[X]| \geq a) \leq \frac{\text{var}[X]}{a^2}$$

$$\begin{aligned} \text{Proof. } P(|X - E[X]| \geq a) &= P((X - E[X])^2 \geq a^2) \\ &\leq \frac{E[(X - E[X])^2]}{a^2} \\ &\leq \frac{\text{var}[X]}{a^2} \end{aligned}$$

Corollary 1. $P(|X - E[X]| \geq a \cdot \delta(X)) \leq 1/a^2$
($\delta(X) = \sqrt{\text{var}(X)}$)

Corollary 2. $P(|X - E[X]| \geq aE[X]) \leq \frac{\text{var}(X)}{a^2[E(X)]^2}$

3 Three models of random graph

1. $G(n, p)$ [probabilistic model]
2. $G(n, m)$ [static model]
3. Dynamic model $|V(G)| = n$ Initially no edge. At each new time point, randomly add a new edge to graph.

4 Threshold function

$G| = A \implies$ If adding any new edge into graph G then $G \Rightarrow G'$
 $G'| = A$

Properties:

1. Connectivity
2. Giant Component $\theta(n)$
3. k_n

$P(n)$ is called a threshold function if

1. $r(n) \ll p(n), \lim_{n \rightarrow +\infty} P[G(n, r(n))| = A] = 0$
2. $r(n) \gg p(n), \lim_{n \rightarrow +\infty} P[G(n, r(n))| = A] = 1$

$f(r) = P[G(n, r)| = A]$ when $P(n)$ is a threshold function $f(r)$ jumps from zero to one around $r = P(n)$

In the dynamic model

1. The graph does not have property A when number of edges $\ll \binom{n}{2} \cdot P(n)$
2. Almost always have property A when number of edges $\gg \binom{n}{2} \cdot P(n)$

$P(G_{n,p}| = k_4)$

$r(n) \ll P(n) P(G_{n,r}| = A) = 0$ $r(n) \gg P(n) P(G_{n,r}| = A) = 0$

Let Y = number of K_4 in a random graph $G_{n,p}$

$$\begin{aligned} E(Y) &= \binom{n}{4} P^6 \\ &= \frac{n(n-1)(n-2)(n-3)}{4 \times 3 \times 2 \times 1} P^6 \\ &\approx n^4 \cdot p^6 \end{aligned}$$

Let $E(Y) = 1$ we have $P = n^{-2/3}$, a threshold.

Next class we will talk about:

When $n \rightarrow \infty, \frac{r(n)}{p(n)} \rightarrow 0$ and $P(Y \geq 1) \rightarrow 0$