Introduction to Random Graphs

longhuan@sjtu.edu.cn

- World Wide Web
- Internet
- Social networks
- Journal citations
- ……

Statistical properties VS Exact answer to questions

The $G(n, p)$ model

Properties of almost all graphs

Phase transition

$\boldsymbol{G}(n, p)$ Model

- $\boldsymbol{G}(n, p)$ Model [Erd$\ddot{o}$s and R$\acute{e}$nyi1960]: $|\mathrm{V}| = n$ is the number of vertices, and for and different $u, v \in V,\ \Pr(\{u, v\} \in E) = p$.

- **Example.** If $p = \frac{d}{n}$.

Then $\boldsymbol{E}(\deg(v)) = \frac{d}{n}(n-1) \approx d$

$$n \approx n - 1$$

Example: $\boldsymbol{G}(n, 1/2)$

$K = \deg(v)$

$$\Pr(K = k) = \binom{n-1}{k}\left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{n-k}$$

$$\approx \boxed{\binom{n}{k}\left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{n-k}} = \frac{1}{2^n}\binom{n}{k}$$

$E(K) = n/2$

$Var(K) = n/4$

Independence!

Binomial Distribution

Recall: Central Limit Theorem

Normal distribution (Gauss Distribution):

$X \sim N(\mu, \sigma^2)$, with density function:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad -\infty < x < +\infty$$

As long as $\{X_i\}$ is independent identically distributed with $E(X_i) = \mu$, $D(X_i) = \sigma^2$, then $\sum_{i=1}^{n} X_i$ can be approximated by normal distribution $(n\mu, n\sigma^2)$when n is large enough.

注意上面只要求了每个Xi独立同分布，且均值和方差一致，并没有
要求它们是什么分布！！

- $\boldsymbol{G}(n, 1/2)$

$$\mu = n\mu' = E(K) = \frac{n}{2},$$

$$\sigma^2 = n(\sigma')^2 = Var(K) = n/4$$

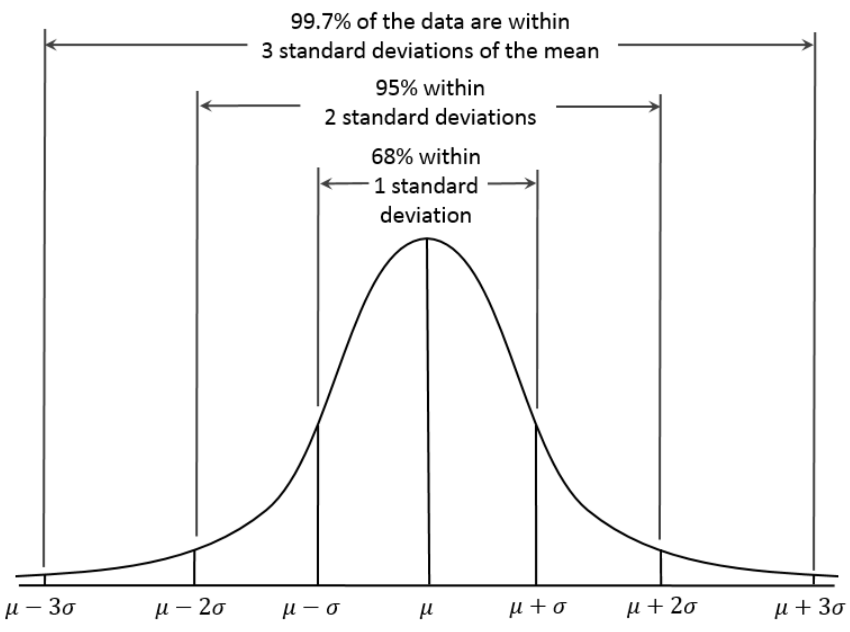
(CLT) Near the mean, the binomial distribution is well approximated by the normal distribution.

$$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(k-n\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{\pi n/2}} e^{-\frac{(k-n/2)^2}{n/2}}$$

It works well when $k = \Theta\,(n)$.

- $G(n, 1/2)$: for any $\epsilon > 0$, the degree of each vertex almost surely is within $(1 \pm \epsilon)\frac{n}{2}$.

Proof. As we can approximate the distribution by



$$\frac{1}{\sqrt{\pi n/2}} e^{-\frac{(k-n/2)^2}{n/2}}$$

$$\mu = \frac{n}{2}, \sigma = \frac{\sqrt{n}}{2}$$

$$\mu \pm c\sigma = \frac{n}{2} \pm c\frac{\sqrt{n}}{2} \approx (1 \pm \epsilon)\frac{n}{2}$$

- $G(n, p)$: for any $\epsilon > 0$, if p is $\Omega\left(\frac{\ln n}{n\epsilon^2}\right)$, then the degree of each vertex almost surely is within $(1 \pm \epsilon)np$.

Proof. Omitted

$\boldsymbol{G}(n,p)$ Model: independent set and clique

Lemma. For all integers n, k with $n \geq k \geq 2$; the probability that $\mathrm{G} \in \boldsymbol{G}(n,p)$ has a set of k independent vertices is at most

$$\Pr(\alpha(G) \geq k) \leq \binom{n}{k}(1-p)^{\binom{k}{2}}$$

the probability that $\mathrm{G} \in \boldsymbol{G}(n,p)$ has a set of k clique is at most

$$\Pr(\omega(G) \geq k) \leq \binom{n}{k}(p)^{\binom{k}{2}}$$

Lemma. The expected number of k −cycles in $\mathrm{G} \in \boldsymbol{G}(n, p)$ is $E(x) = \frac{(n)_k}{2k} p^k$.

Proof. The expectation of certain n vertices $v_0, v_1, \cdots, v_{k-1}, v_0$ form a length k cycle is: p^k

The possible ways to choose k vertices to form a cycle C is $\frac{(n)_k}{2k}$.

The expectation of the number of all cycles:

$$X = \sum_C X_C = \frac{(n)_k}{2k} p^k$$

The $G(n, p)$ model

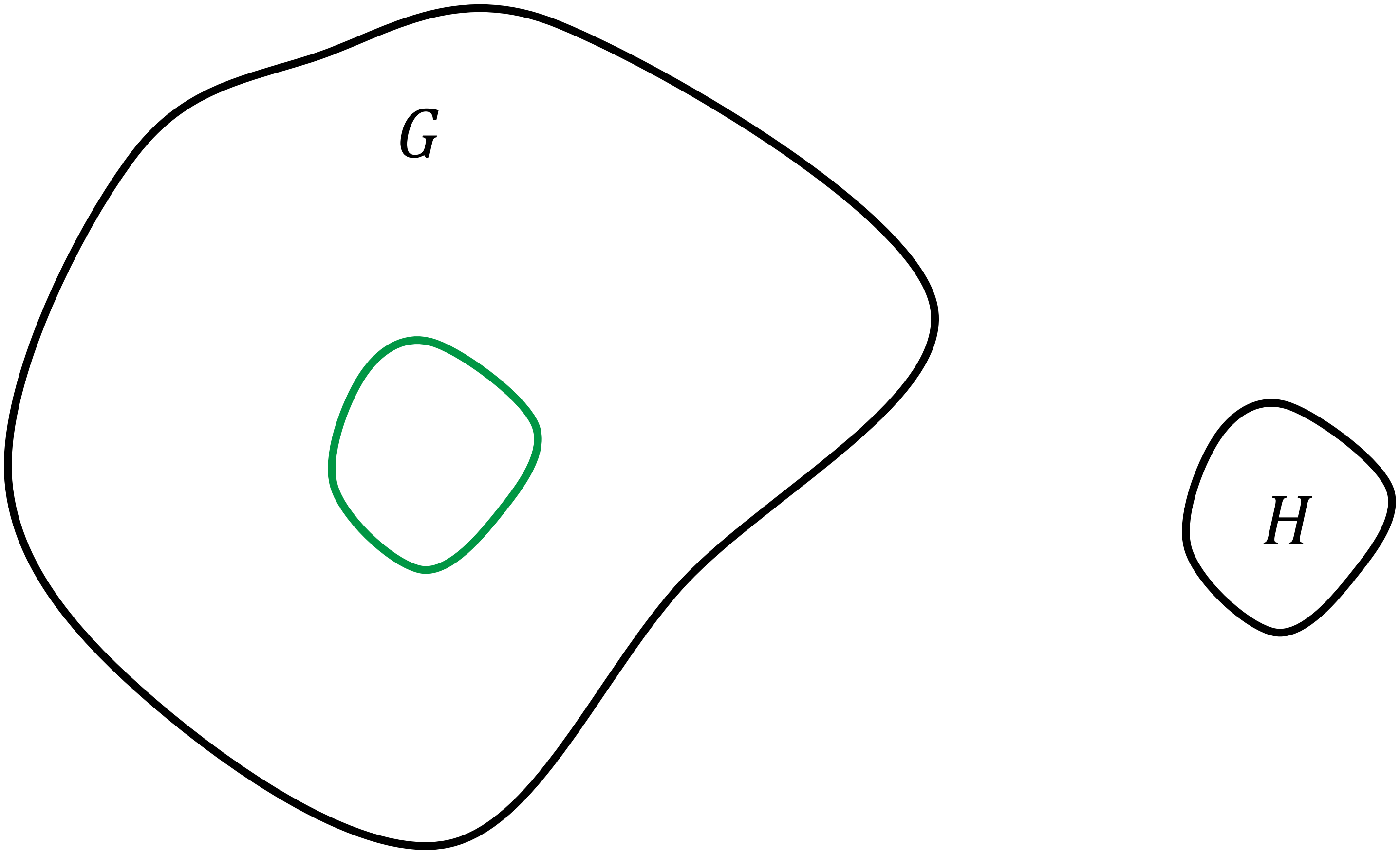
Properties of almost all graphs

Phase transition

Properties of almost all graphs

- For a graph property P , when $n \to \infty$, If the *limit* of the probability of $G \in \boldsymbol{G}(n, p)$ having the property tends to
  - **1**: we say than the property holds for almost all (almost every / almost surely) $G \in \boldsymbol{G}(n, p)$.
  - **0**: we say than the property holds for almost no $G \in \boldsymbol{G}(n, p)$.

Proposition. For every constant $p \in (0,1)$ and every graph H, almost every $G \in \boldsymbol{G}(n,p)$ contains an induced copy of H.



Proposition. For every constant $p \in (0,1)$ and every graph H, almost every $G \in \boldsymbol{G}(n,p)$ contains an induced copy of H.

Proof. $V(G) = \{v_0, v_1, \dots, v_{n-1}\}, k = |H|$

Fix some $U \in \binom{V(G)}{k}$, then $\Pr(U \cong H) = r > 0$

r depends on p, k not on n.

There are $\lfloor n/k \rfloor$ disjoint such U.

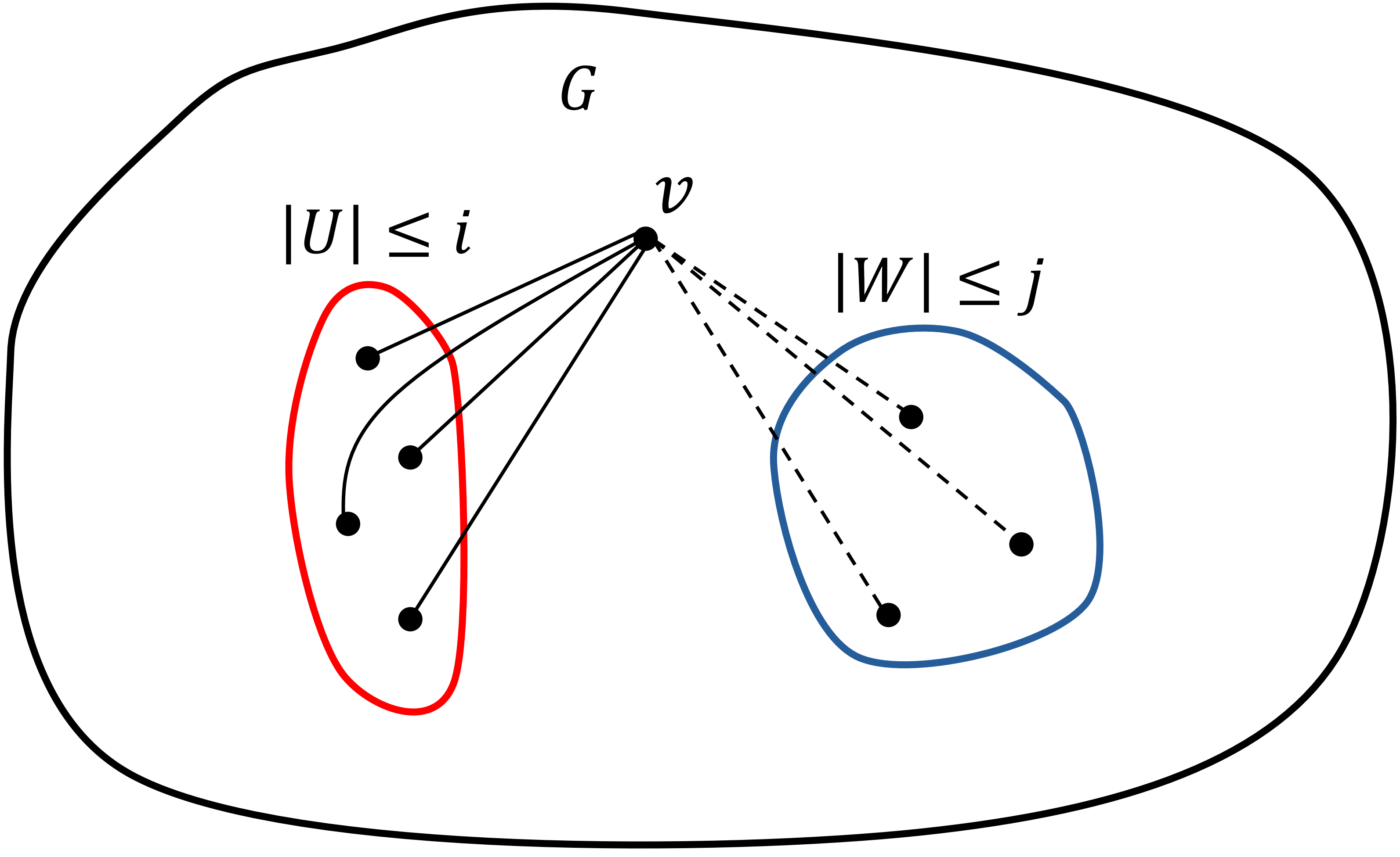
The probability that none of the $G[U]$ is isomorphic to H is: $= (1-r)^{\lfloor n/k \rfloor}$

$\Pr[\neg(H \subseteq G \text{ induced})]$: $\leq (1-r)^{\lfloor n/k \rfloor}$

$$\Big\downarrow n \to \infty$$

$$0$$

Proposition. For every constant $p \in (0,1)$ and $i, j \in N$, almost every graph $G \in \boldsymbol{G}(n, p)$ has the property $P_{i,j}$.



Proposition. For every constant $p \in (0,1)$ and $i, j \in N$, almost every graph $G \in \boldsymbol{G}(n, p)$ has the property $P_{i,j}$.

Proof. Fix U, W and $v \in G - (U \cup W)$, $q = 1 - p$,

The probability that $P_{i,j}$ holds for v: $p^{|U|} q^{|W|} \geq p^i q^j$

The probability there's no such v for chosen U, W:

$$= \left(1 - p^{|U|} q^{|W|}\right)^{n-|U|-|W|} \leq \left(1 - p^i q^j\right)^{n-i-j}$$

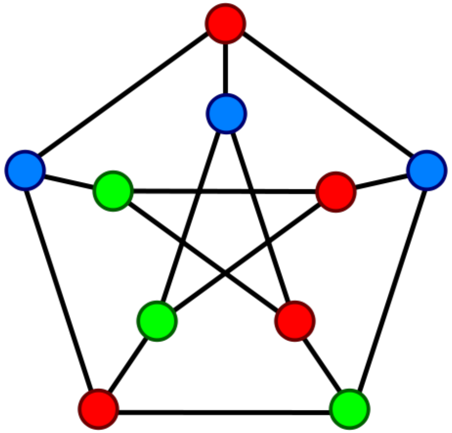
The upper bound for the number of different choice of U, W: n^{i+j}

The probability there exists some U, W without suitable v:

$$\leq n^{i+j} \left(1 - p^i q^j\right)^{n-i-j} \xrightarrow{n \to \infty} 0$$

Colouring

- **Vertex coloring:** to $G = (V, E)$, a vertex coloring is a map $c: V \to S$ such that $c(v) \neq c(w)$ whenever v and w are adjacent.
- **Chromatic number $\boldsymbol{\chi(G)}$:** the smallest size of S.



$\chi(G) = 3$

Colouring

- **Vertex coloring:** to $G = (V, E)$, a vertex coloring is a map $c: V \to S$ such that $c(v) \neq c(w)$ whenever v and w are adjacent.
- **Chromatic number $\chi(G)$:** the smallest size of S.
- **Some famous results：**
  - Whether $\chi(G) = k$ is NP-complete.
  - Every Planar graph is 4-colourable.
  - [Grt$\ddot{o}$zsch 1959]Every Planar graph not containing a triangle is 3-colourable.

Proposition. For every constant $p \in (0,1)$ and every $\epsilon > 0$, almost every graph $G \in \boldsymbol{G}(n,p)$ has chromatic number $\chi(G) > \frac{\log(1/q)}{2+\epsilon} \cdot \frac{n}{logn}$

Proof. The size of the maximum independent set in G: $\alpha(G)$

$$\Pr(\alpha(G) \geq k) \leq \binom{n}{k} q^{\binom{k}{2}} \leq n^k q^{\binom{k}{2}}$$

(*)

$$= q^{k\frac{\log n}{\log q}+\frac{1}{2}k(k-1)} = q^{\frac{k}{2}\left(-\frac{2\log n}{\log(1/q)}+k-1\right)}$$

Take $k = (2+\epsilon)\frac{\log n}{\log(1/q)}$ then (*) tends to ∞ with n.

$\therefore \Pr(\alpha(G) \geq k) \xrightarrow{n\to\infty} 0 \Rightarrow$ No k vertices can have the same colour.

$$\therefore \chi(G) > \frac{n}{k} = \frac{\log(1/q)}{2+\epsilon} \cdot \frac{n}{\log n}$$

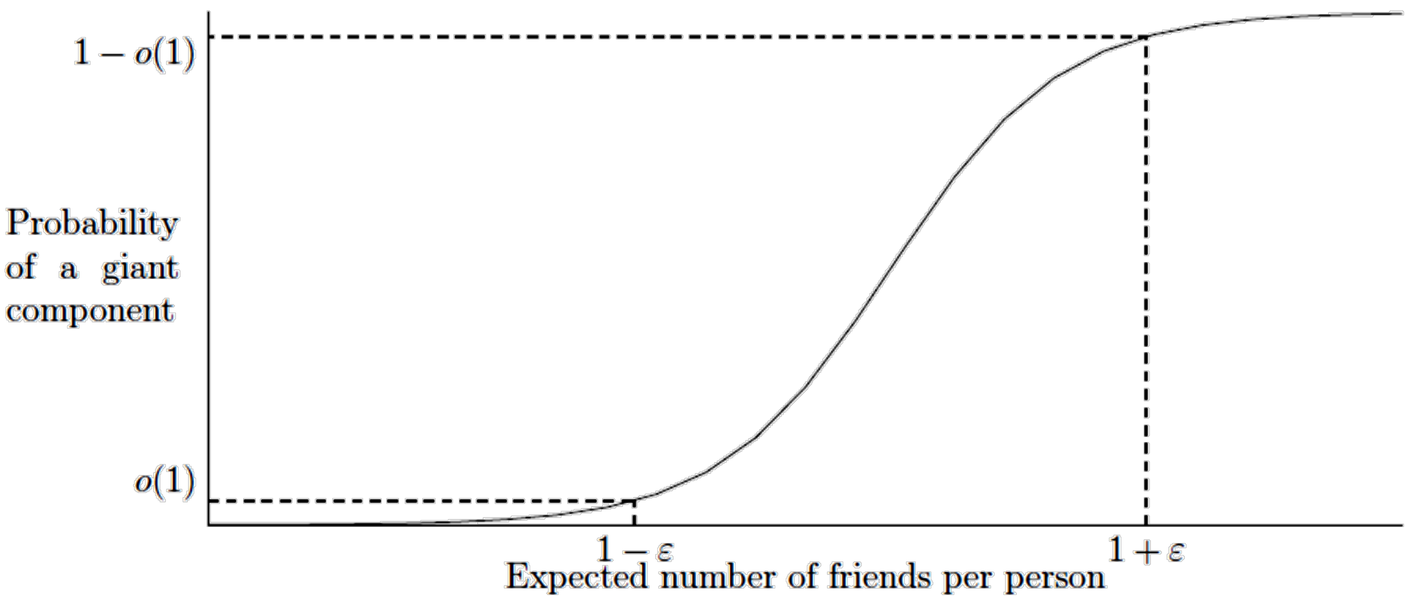
The $G(n, p)$ model

Properties of almost all graphs

Phase transition

Phase transition

The interesting thing about the $\boldsymbol{G}(n, p)$ model is that even though edges are chosen independently, certain global properties of the graph emerge from the independent choice.



Phase transition

Definition. If there exists a function $p(n)$ such that

- when $\lim_{n\to\infty}\left(\frac{p_1(n)}{p(n)}\right)=0$, $\boldsymbol{G}(n, p_1(n))$ almost surely does not have the property.
- when $\lim_{n\to\infty}\left(\frac{p_2(n)}{p(n)}\right)=\infty$, $\boldsymbol{G}(n, p_2(n))$ almost surely has the property.

Then we say phase transition occurs and $p(n)$ is the threshold.

Every increasing property has a threshold.

Phase transition

Probability	Transition
$p = o(\frac{1}{n})$	Forest of trees, no component of size greater than $O(\log n)$
$p = \frac{d}{n}, d < 1$	All components of size $O(\log n)$
$p = \frac{d}{n}, d = 1$	Components of size $O(n^{\frac{2}{3}})$
$\boxed{p = \frac{d}{n}, d > 1}$	Giant component plus $O(\log n)$ components
$p = \sqrt{\frac{2 \ln n}{n}}$	Diameter two
$p = \frac{1}{2} \frac{\ln n}{n}$	Giant component plus isolated vertices
$p = \frac{\ln n}{n}$	Disappearance of isolated vertices Appearance of Hamilton circuit Diameter $O(\log n)$
$p = \frac{1}{2}$	Clique of size $(2 - \epsilon) \ln n$

First moment method

Markov's Inequality: Let x be a random variable that assumes only nonnegative values. Then for all $a > 0$

$$\Pr(x \geq a) \leq \frac{\boldsymbol{E}[x]}{a}$$

First moment method : for non-negative, integer valued variable x

$$\Pr(x > 0) = \Pr(x \geq 1) \leq \boldsymbol{E}(x)$$

$$\therefore \ \Pr(x = 0) = 1 - \Pr(x > 0) \geq 1 - \boldsymbol{E}(x)$$

First moment method : for non-negative , integer valued variable x

$$\Pr(x > 0) = \Pr(x \geq 1) \leq \boldsymbol{E}(x)$$

$$\therefore \Pr(x = 0) = 1 - \Pr(x > 0) \geq 1 - \boldsymbol{E}(x)$$

- If the expectation goes to 0: the property almost surely does not happen.
- If the expectation does not goes to 0:

e.g. Expectation = $\frac{1}{n} \times n^2 + \frac{n-1}{n} \times 0 = n$

i.e., a *vanishingly small* fraction of the sample contribute a lot to the expectation.

Chebyshev's Inequality

- For any $a > 0$,

$$\Pr(|X - E(X)| \geq a) \leq \frac{Var[X]}{a^2}$$

Second moment method

Theorem. Let $x(n)$ be a random variable with $\boldsymbol{E}(x) > 0$. If

$$Var(x) = o\left(\boldsymbol{E}^2(x)\right)$$

Then x is almost surely greater than zero.

Proof. If $\boldsymbol{E}(x) > 0$, then for $x \le 0$,

$$\Pr(x \le 0) \le \Pr\left(|x - \boldsymbol{E}(x)| \ge \boldsymbol{E}(x)\right)$$

$$\le \frac{Var(x)}{E^2(x)} \to 0$$

Example : Threshold for graph diameter two (two degrees of separation)

Probability	Transition
$p = o(\frac{1}{n})$	Forest of trees, no component of size greater than $O(\log n)$
$p = \frac{d}{n}, d < 1$	All components of size $O(\log n)$
$p = \frac{d}{n}, d = 1$	Components of size $O(n^{\frac{2}{3}})$
$p = \frac{d}{n}, d > 1$	Giant component plus $O(\log n)$ components
$p = \sqrt{\frac{2 \ln n}{n}}$	Diameter two
$p = \frac{1}{2} \frac{\ln n}{n}$	Giant component plus isolated vertices
$p = \frac{\ln n}{n}$	Disappearance of isolated vertices Appearance of Hamilton circuit Diameter $O(\log n)$
$p = \frac{1}{2}$	Clique of size $(2 - \epsilon) \ln n$

Example : Threshold for graph diameter two (two degrees of separation)

- **Diameter:** the maximum length of the shortest path between a pair of nodes.
- **Theorem:** The property that $\boldsymbol{G}(n, p)$ has diameter two has a sharp threshold at $p = \sqrt{2}\sqrt{\frac{\ln n}{n}}$.

sharp threshold是指p1(n), p2(n)都是cp(n)的形式

Example : Threshold for graph diameter two (two degrees of separation)

Theorem. The property that $\boldsymbol{G}(n,p)$ has diameter two has a sharp threshold at $p = \sqrt{2}\sqrt{\frac{\ln n}{n}}$

Proof. For any two different vertices $i < j$,

$$I_{ij} = \begin{cases} 1 & \{\mathrm{i,j}\} \notin E, \text{no other vertex is adjacent to both } i \text{ and } j \\ 0 & otherwise \end{cases}$$

$$x = \sum_{i<j} I_{ij}$$

If $\boldsymbol{E}(x) \xrightarrow{n\to\infty} 0$, then for large n, almost surely the diameter is at most two.

Example : Threshold for graph diameter two (two degrees of separation)

Theorem. The property that $\boldsymbol{G}(n,p)$ has diameter two has a sharp threshold at $p = \sqrt{2}\sqrt{\frac{\ln n}{n}}$

Proof. For any two different vertices $i < j$,

$$I_{ij} = \begin{cases} 1 & \{\text{i, j}\} \notin E, \text{no other vertex is adjacent to both } i \text{ and } j \\ 0 & otherwise \end{cases}$$

$$x = \sum_{i<j} I_{ij}$$

$$\boxed{\boldsymbol{E}(x) = \binom{n}{2}(1-p)(1-p^2)^{n-2}}$$

Take $p = c\sqrt{\frac{\ln n}{n}}$, $E(x) \cong \frac{n^2}{2}\left(1 - c\sqrt{\frac{\ln n}{n}}\right)\left(1 - c^2\frac{\ln n}{n}\right)^n$

$$\cong \frac{n^2}{2}e^{-c^2 \ln n} = \frac{1}{2}n^{2-c^2}$$

Example : Threshold for graph diameter two (two degrees of separation)

Theorem. The property that $\boldsymbol{G}(n,p)$ has diameter two has a sharp threshold at $p = \sqrt{2}\sqrt{\frac{\ln n}{n}}$

Proof. For any two different vertices $i < j$,

$$I_{ij} = \begin{cases} 1 & \{\mathrm{i,j}\} \notin E, \text{no other vertex is adjacent to both } i \text{ } and \text{ } j \\ 0 & otherwise \end{cases}$$

$$x = \sum_{i<j} I_{ij} \qquad\qquad \boldsymbol{E}(x) = \binom{n}{2}(1-p)(1-p^2)^{n-2}$$

Take $p = c\sqrt{\frac{\ln n}{n}}$, $c > \sqrt{2}$, $\lim_{n\to\infty} \boldsymbol{E}(x) = 0$

For large n, almost surely the diameter is at most two.

Theorem. The property that $\boldsymbol{G}(n, p)$ has diameter two has a sharp threshold at $p = \sqrt{2}\sqrt{\frac{\ln n}{n}}$

- Take $p = c\sqrt{\frac{\ln n}{n}}$, $c > \sqrt{2}$, $\lim_{n\to\infty} \boldsymbol{E}(x) = 0$

Theorem. The property that $\boldsymbol{G}(n, p)$ has diameter two has a sharp threshold at $p = \sqrt{2}\sqrt{\frac{\ln n}{n}}$

- Take $p = c\sqrt{\frac{\ln n}{n}}$, $c < \sqrt{2}$,

$$\boldsymbol{E}(x^2) = \boldsymbol{E}\left(\sum_{i<j} I_{ij}\right)^2$$

If $Var(x) = o\left(\boldsymbol{E}^2(x)\right)$, then for large n, almost surely the diameter will be larger than two.

Theorem. The property that $\boldsymbol{G}(n,p)$ has diameter two has a sharp threshold at $p = \sqrt{2}\sqrt{\frac{\ln n}{n}}$

- Take $p = c\sqrt{\frac{\ln n}{n}}$, $c < \sqrt{2}$

$$\boldsymbol{E}(x^2) = \boldsymbol{E}\left(\sum_{i<j} I_{ij}\right)^2 = \boldsymbol{E}\left(\sum_{i<j} I_{ij} \sum_{k<l} I_{kl}\right) = \boldsymbol{E}\left(\sum_{\substack{i<j\\k<l}} I_{ij}\, I_{kl}\right) = \sum_{\substack{i<j\\k<l}} \boldsymbol{E}\left(I_{ij} I_{kl}\right)$$

$$a = |\{i,j,k,l\}|$$

$$\boldsymbol{E}(x^2) = \sum_{\substack{i<j\\k<l\\a=4}} \boldsymbol{E}\left(I_{ij} I_{kl}\right) + \sum_{\substack{i<j\\i<k\\a=3}} \boldsymbol{E}\left(I_{ij} I_{ik}\right) + \sum_{\substack{i<j\\ \\a=2}} \boldsymbol{E}\left(I_{ij}^2\right)$$

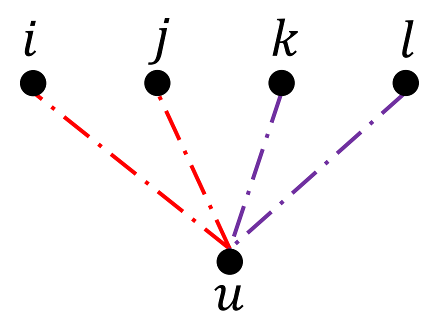
Theorem. The property that $G(n,p)$ has diameter two has a sharp threshold at $p = \sqrt{2}\sqrt{\frac{\ln n}{n}}$

- Take $p = c\sqrt{\frac{\ln n}{n}}$, $c < \sqrt{2}$

$$E(x^2) = \boxed{\sum_{\substack{i<j\\k<l\\a=4}} E(I_{ij}I_{kl})} + \sum_{\substack{i<j\\i<k\\a=3}} E(I_{ij}I_{ik}) + \sum_{\substack{i<j\\ \\a=2}} E(I_{ij}^2)$$

$$E(I_{ij}I_{kl}) \le (1-p^2)^{2(n-4)} \le \left(1 - c^2\frac{\ln n}{n}\right)^{2n}(1+o(1)) \le n^{-2c^2}(1+o(1))$$

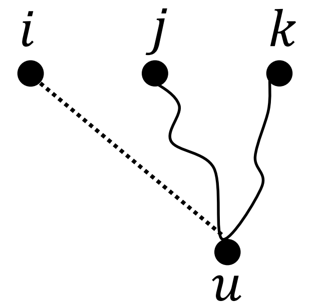
$$\sum_{\substack{i<j\\k<l\\a=4}} E(I_{ij}I_{kl}) \le n^{4-2c^2}(1+o(1))$$



Theorem. The property that $\boldsymbol{G}(n, p)$ has diameter two has a sharp threshold at $p = \sqrt{2}\sqrt{\frac{\ln n}{n}}$

- Take $p = c\sqrt{\frac{\ln n}{n}}$, $c < \sqrt{2}$

$$\boldsymbol{E}(x^2) = \sum_{\substack{i<j \\ k<l \\ a=4}} \boldsymbol{E}\left(I_{ij} I_{kl}\right) + \boxed{\sum_{\substack{i<j \\ i<k \\ a=3}} \boldsymbol{E}\left(I_{ij} I_{ik}\right)} + \sum_{\substack{i<j \\ a=2}} \boldsymbol{E}\left(I_{ij}^2\right)$$



Theorem. The property that $\boldsymbol{G}(n,p)$ has diameter two has a sharp threshold at $p=\sqrt{2}\sqrt{\frac{\ln n}{n}}$

- Take $p=c\sqrt{\frac{\ln n}{n}}$, $c<\sqrt{2}$

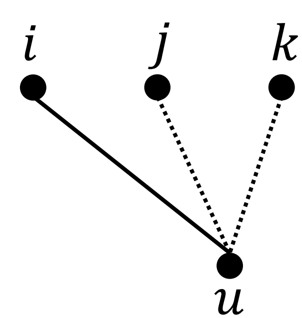
$$\boldsymbol{E}(x^2)=\sum_{\substack{i<j\\k<l\\a=4}}\boldsymbol{E}(I_{ij}I_{kl})+\boxed{\sum_{\substack{i<j\\i<k\\a=3}}\boldsymbol{E}(I_{ij}I_{ik})}+\sum_{\substack{i<j\\a=2}}\boldsymbol{E}(I_{ij}^2)$$

$$\Pr(I_{ij}I_{ik}=1)\le 1-p+p(1-p)^2=1-2p^2+p^3\approx 1-2p^2$$

$$\boldsymbol{E}(I_{ij}I_{ik})\le(1-2p^2)^{n-3}=\left(1-\frac{2c^2\ln n}{n}\right)^{n-3}$$

$$\cong e^{-2c^2\ln n}=n^{-2c^2}$$

$$\sum_{\substack{i<j\\i<k}}\boldsymbol{E}(I_{ij}I_{ik})\le n^{3-2c^2}$$



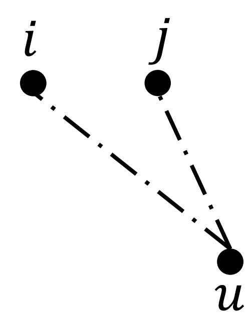
Theorem. The property that $G(n,p)$ has diameter two has a sharp threshold at $p = \sqrt{2}\sqrt{\frac{\ln n}{n}}$

- Take $p = c\sqrt{\frac{\ln n}{n}}$, $c < \sqrt{2}$

$$E(x^2) = \sum_{\substack{i<j\\ k<l\\ a=4}} E(I_{ij}I_{kl}) + \sum_{\substack{i<j\\ i<k\\ a=3}} E(I_{ij}I_{ik}) + \boxed{\sum_{\substack{i<j\\ \\ a=2}} E(I_{ij}^2)}$$

$$E(I_{ij}^2) = E(I_{ij})$$

$$\sum_{ij} E(I_{ij}^2) = E(x) \cong \frac{1}{2} n^{2-c^2}$$



Theorem. The property that $\boldsymbol{G}(n, p)$ has diameter two has a sharp threshold at $p = \sqrt{2}\sqrt{\frac{\ln n}{n}}$

- Take $p = c\sqrt{\frac{\ln n}{n}}$, $c < \sqrt{2}$

$$\boldsymbol{E}(x^2) \leq \boldsymbol{E}^2(x)(1 + o(1))$$

A graph almost surely has at least one bad pair of vertices and thus diameter greater than two.