Lecture 6. Threshold functions of random graphs.

We have showed that surprisingly many tempting conjectures can be easily disproves by the probabilistic method and random graphs.

Question: For which $p = p_n$ is a property true for $G(n, p)$ w.h.p.?

Graph property: A graph property $\mathcal{P}$ is a subset of all graphs.

monotone (increasing): $\forall G \in \mathcal{P}$, any supergraph (adding edge) $\in \mathcal{P}$.

non-trivial: $\forall$ sufficiently large n, $\exists$ n-vertex graph $G \notin \mathcal{P}$.

Examples: containing H (monotone increasing). connected

planar (monotone, decreasing). containing vertex of deg 1. (nor).

Warmup: $\mathcal{P} = \{G : \text{triangle } K_3 \subseteq G\}$. consider $G \sim \mathcal{G}(n, p_n)$.

Let $X = \#$ of Δs. $\mathbb{E}[X] = \binom{n}{3} p^3 \approx n^3 p^3$. If $p \ll \frac{1}{n}$, $\Pr[X \geq 1] = o(1)$ by Markov's inequality. If $p \gg \frac{1}{n}$, $\mathbb{E}[X] \to \infty$. Markov fails.

Chebyshev: $\Pr[|X - \mathbb{E}[X]| \geq t] \leq \frac{Var[X]}{t^2} \Rightarrow \Pr[X = 0] = o(1)$ if $X \geq 0$, and $Var[X] = o(\mathbb{E}[X]^2)$. If $X = X_1 + \cdots + X_n$ then $Var[X] = \sum \mathbb{E}[X_i] + 2 \sum_{i<j} Cov[X_i, X_j]$. $Cov[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$.

$Cov[X_{T_1}, X_{T_2}] = \mathbb{E}[X_{T_1} X_{T_2}] - \mathbb{E}[X_{T_1}]\mathbb{E}[X_{T_2}] = p^{e(T_1 \cup T_2)} - p^{e(T_1) + e(T_2)}$

$= \begin{cases} 0 & |T_1 \cap T_2| \le 1 \\ p^5 - p^6 & |T_1 \cap T_2| = 2 \end{cases}$. $Var[X_T] = \mathbb{E}[X_T^2] - \mathbb{E}[X_T]^2 = p^3 - p^6$.

$\Rightarrow Var[X] = \sum_T Var[X_T] + \sum_{T_1 \ne T_2} Cov[X_{T_1}, X_{T_2}] = \binom{n}{3}(p^3 - p^6) + \sum_{|T_1 \cap T_2| = 2} (p^5 - p^6)$

$= \binom{n}{3}(p^3 - p^6) + \binom{n}{2}(n-2)(n-3)(p^5 - p^6) \lesssim n^3 p^3 + n^4 p^5$. We hope

$Var[X] = o(\mathbb{E}[X]^2) \Leftarrow n^3 p^3 + n^4 p^5 = o(n^6 p^6) \Leftarrow p \gg 1/n$.

Definition. (Threshold function). We say r_n is a threshold function for some graph property $\mathcal{P}$. if $Pr[\mathcal{G}(n, p_n) \in \mathcal{P}] \to \begin{cases} 0 & p_n / r_n \to 0 \\ 1 & p_n / r_n \to \infty \end{cases}$.

Theorem. A threshold function for containing a K_3 is $1/n$.

Question. What about containing a K_4? (already known $n^{-2/3}$).

General setup. Suppose $X = X_1 + \cdots + X_m$, X_i is indicator of event E_i. Denote by $i \sim j$ if $i \ne j$ but E_i, E_j are not independent. (for variance we only need pairwise independence, but will later rely on a stronger notion of dependency in Lovász local lemma). If $i \ne j$ and $i \not\sim j$ clearly $Cov[X_i, X_j] = 0$. otherwise $Cov[X_i, X_j] \le \mathbb{E}[X_i X_j] = Pr[E_i \wedge E_j]$

Thus $Var[X] = \sum_{i,j} Cov[X_i, X_j] \le \mathbb{E}[X] + \underbrace{\sum_{i \sim j} Pr[E_i \wedge E_j]}_{} \triangleq \Delta$.

We hope $Var[X] = o(\mathbb{E}[X])^2$. So $\Delta = o(\mathbb{E}[X])^2$ suffices. Furthermore.

$\sum_{i \sim j} Pr[E_i \wedge E_j] = \sum_i Pr[E_i] \sum_{j \sim i} Pr[E_j \mid E_i]$. In many symmetric cases.

$\sum_{j\sim i} \Pr[E_j \mid E_i]$ does not depend on i. Denote it by Δ^*. (or we may set $\Delta^* = \max_i \sum_{j\sim i} \Pr[E_j \mid E_i]$). $\Rightarrow \Delta = \sum_i \Pr[E_i]\,\Delta^* = \mathbb{E}[X]\,\Delta^*$.

Lemma. If $\mathbb{E}[X] \to \infty$ and $\Delta^* = o(\mathbb{E}[X])$. then $X > 0$ w.h.p.

Now consider the property of containing K_4. $\forall S$. $E_S = S$ forms a K_4.

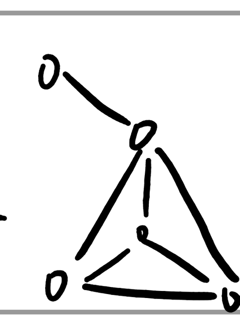
$\forall S, T$ of size 4. $S \sim T$ iff $|S \cap T| \geq 2$. There are two cases:

① $|S\cap T| = 2$. $\sum_T \Pr[E_T \mid E_S] \leq 6\binom{n}{2}\Pr[E_T \mid E_S] = 6\binom{n}{2}p^5 \approx n^2 p^5$.

② $|S\cap T| = 3$. $\sum_T \Pr[E_T \mid E_S] = 4(n-4)\Pr[E_T \mid E_S] \leq 4np^3 \approx np^3$.

$\Rightarrow \Delta^* \approx n^2p^5 + np^3 = o(n^4p^6) = o(\mathbb{E}[X])$ if $n^2p \gg 1$, $np \gg 1$.

Conjecture: $X = \#$ of K_4 (or other H). $\mathbb{E}[X] \to \infty \Rightarrow X > 0$ whp.

Counterexample. Consider $H =$  $\mathbb{E}[X] \approx n^5p^7 \to \infty$ if $p \gg n^{-5/7}$.

However, H contains K_4. If $p \ll n^{-2/3}$ there is no K_4 in $G(n,p)$.

In fact the threshold function is $n^{-2/3}$ since we can easily extend a K_4 to H as long as K_4 exists. $n^{-2/3} \gg n^{-5/7}$. why? $\mathrm{Var}[X]$ is large. no concentration. Δ^* has a term np if two copies of H overlap with a K_4 and $np \ll n^5p^7$ if $p \ll n^{-2/3}$. Chebyshev fails.

Definition: The edge-vertex ratio of $G=(V,E)$: $\rho(G) = |E|/|V|$.

The maximum subgraph ratio is given by $m(G) = \max_{H \subseteq G} \rho(H)$.

Theorem (Bollobás, 1981). Fix a graph $H = (V, E)$. Then $p = n^{-1/m(H)}$ is a threshold function for containing H as a subgraph. Furthermore, if $p \gg n^{-1/m(H)}$, then $X_H = \#$ of copies of H in $G(n,p)$ is w.h.p.

$$X_H \approx \mathbb{E}[X] = \binom{n}{|V|} \frac{|V|!}{|Aut(H)|} p^{|E|} \approx \frac{n^{|V|} p^{|E|}}{|Aut(H)|}.$$

Proof. Let H' be the subgraph of H achieving the maximum e-v ratio, i.e. $m(H) = \rho(H')$. If $p \ll n^{-1/m(H)}$, $\mathbb{E}[X_{H'}] = o(1) \Rightarrow X_{H'} = 0$ whp.

Now assume $p \gg n^{-1/m(H)}$. Count the labelled copies of H in $G(n,p)$.

Let L be a labelled copy of H in K_n. A_L be the event of $L \subseteq G(n,p)$.

For fixed L. $\Delta^* = \sum_{L' \sim L} \Pr[A_{L'} \mid A_L] = \sum_{L' \sim L} p^{|E(L') \setminus E(L)|}$. Note that $\#$ of L' that $L' \sim L \approx n^{|V(L') \setminus V(L)|}$, and $p \gg n^{-1/m(H)} \geq n^{-1/\rho(L' \cap L)} = n^{-|V(L') \cap V(L)| / |E(L') \cap E(L)|}$. So $\Delta^* \approx \sum n^{|V(L') \setminus V(L)|} p^{|E(L') \setminus E(L)|} \ll n^{|V(L)|} p^{|E(L)|}$

$\Rightarrow \Delta^* \ll \mathbb{E}[X_H] \Rightarrow Var[X] = \mathbb{E}[X_H] + o(\mathbb{E}[X_H])^2$. $\square$

Existence of threshold: for which property $\mathcal{P}$, a threshold exist?

Simpler question: $\Pr[G(n,p) \in \mathcal{P}]$ increasing of p if $\mathcal{P}$ monotone?

Consider a simple case: $\mathcal{F}$ is a family of subset of $[n]$. $\mathcal{F}$ is called upward closed set (or up-set) if $\forall S \subseteq T$ and $S \in \mathcal{F}$, then $T \in \mathcal{F}$.

Theorem. Let $\mathcal{F}$ be a nontrivial ($\mathcal{F} \neq \phi, 2^{[n]}$) up-set of $[n]$. Then $f(p) = \Pr[\mathrm{Bin}([n], p) \in \mathcal{F}]$ is a strictly increasing function.

Proof. (by coupling). coupling: joint distribution of (X, Y) where marginal distributions are given. $\forall\, 0 \le p < q < 1$, construct a coupling. Pick a uniform random vector $(x_1, \dots, x_n) \in [0,1]^n$. Let $A = \{i : x_i \le p\}$ and $B = \{j : x_j \le q\}$. Clearly A has the same distribution as $\mathrm{Bin}([n], p)$ and B has the same distribution as $\mathrm{Bin}([n], q)$. But $A \subseteq B$. Thus $f(p) = \Pr[A \in \mathcal{F}] < \Pr[B \in \mathcal{F}] = f(q)$. $\square$

Another proof, also by two-round exposure coupling: Let $0 \le p < q \le 1$. Construct A, B as follows. $\forall i \in [n]$, add i into A with probability p. If $i \in A$, add i into B. Otherwise add it into B with probability $1 - \frac{1-q}{1-p}$. Then $\Pr[i \in B] = p + (1-p)(1 - \frac{1-q}{1-p}) = q$ and independent. So B has the same distribution as $\mathrm{Bin}([n], q)$, and $A \subseteq B$. $\square$

Theorem (Bollobás & Thomason, 1987). Every nontrivial monotone (increasing) graph property has a threshold function.

Proof. Consider k independent copies $G_1, G_2, \dots, G_k$ of $G(n,p)$. Their union $G_1 \cup G_2 \cup \cdots \cup G_k$ has the same distribution of $G(n, 1-(1-p)^k)$. By the monotonicity of $\mathcal{P}$, if $G_1 \cup G_2 \cup \cdots \cup G_k \notin \mathcal{P}$, then $G_1, \dots, G_k \notin \mathcal{P}$. By independence, $\Pr[G(n, 1-(1-p)^k) \notin \mathcal{P}] \le \Pr[G(n,p) \notin \mathcal{P}]^k$.

Let $f(p) = f_n(p) = \Pr[G(n,p) \in \mathcal{P}]$. Note that $(1-p)^k \ge 1-kp$ by convexity. $\forall$ monotone $\mathcal{P}$, and any positive integer $k \le 1/p$, we have $1-f(kp) \le 1-f(1-(1-p)^k) \le (1-f(p))^k$. Fix a sufficiently large n. Since $f(p)$ is a continuous strictly increasing function from 0 to 1 as p goes from 0 to 1, there is some critical $p_c = p_c(n)$ s.t. $f(p_c) = \frac{1}{2}$.

We claim that p_c is a threshold function. If $p = p(n) \gg p_c$, then letting $k = k(n) = \lfloor p/p_c \rfloor \to \infty$, $1-f(p) \le (1-f(p_c))^k = 2^{-k} \to 0$, so $f(p) \to 1$. Analogously, if $p \ll p_c$, then letting $k = \lfloor p_c/p \rfloor \to \infty$, we have $\frac{1}{2} = 1-f(p_c) \le (1-f(p))^k$. Thus $f(p) \to 0$ as $n \to \infty$. $\square$

Remark. In fact, using the method of moments, # of Δs converges to a Poisson distribution. Thus $\Pr[\Delta \subseteq G(n, c_n/n)] = \begin{cases} 1-e^{-c^3/6} & c_n \to c \\ 1 & c_n \to \infty \end{cases}$

However, consider some other properties, such as "no isolated v".

We have $\Pr[G(n,p) \text{ has no isolated vertex}] = e^{-e^{-c}}$ if $c_n \to c$ where $p = \frac{\log n + c_n}{n}$ and $c \in \mathbb{R} \cup \{-\infty, \infty\}$. (exercise!). This property shows a stronger notion of threshold: sharp threshold.

Definition: We say r_n is a sharp threshold for some graph property $\mathcal{P}$. if $\forall \delta > 0$, $\Pr[G(n, p_n) \in \mathcal{P}] \to \begin{cases} 0 & \text{if } p_n \le (1-\delta) r_n \\ 1 & \text{if } p_n \ge (1+\delta) r_n \end{cases}$

Sharp threshold: $[f^{-1}(\varepsilon), f^{-1}(1-\varepsilon)] = \Theta(r_n) \to [f^{-1}(\varepsilon), f^{-1}(1-\varepsilon)] = o(r_n)$.

Remark: Roughly speaking, all monotone graph property with a cocurse threshold may be approximated by a local property (having some H as a subgraph). (Friedgut's sharp threshold theorem, 1999).

Conjecture: $\forall k \ge 3$. $\exists d_k$ s.t. $\Pr[G(n, d/n) \text{ is } k\text{-colorable}] \to \begin{cases} 1 & d < d_k \\ 0 & d > d_k \end{cases}$.

Theorem (Achlioptas & Friedgut, 2000). $\forall k \ge 3$. $\exists$ function $d_k(n)$ s.t. $\forall \varepsilon > 0$, $\Pr[G(n, d(n)/n) \text{ is } k\text{-colorable}] \to \begin{cases} 1 & \text{if } d(n) < d_k(n) - \varepsilon \\ 0 & \text{if } d(n) > d_k(n) + \varepsilon \end{cases}$.

Open question: whether $d_k(n)$ has a limit d_k?

Consider an easy case: chromatic number of $G(n, 1/2)$. We do know it has concentration in the course Advanced Algorithms.

Start from a question: what is the clique number of $G(n, 1/2)$?

Let $X = \#$ of cliques of $G(n, 1/2)$. We have $f_k = \mathbb{E}[X] = \binom{n}{k} 2^{-\binom{k}{2}}$.

Fix S of size k. $\forall T$ of size k. $T \sim S \Rightarrow |S \cap T| \geqslant 2$. So we have

$$\Delta^* = \sum_{T \sim S} \Pr[A_T \mid A_S] = \sum_{i=2}^{k-1} \binom{k}{i} \binom{n-k}{k-i} 2^{\binom{i}{2} - \binom{k}{2}} \leq 2^{-\binom{k}{2}} \sum k^i n^{k-i} 2^{\binom{i}{2}}$$

$$\leq 2^{-\binom{k}{2}} n^k \sum \left(\frac{k 2^i}{n} \right)^i = o(f_k) \text{ if } f_k \to \infty. \Rightarrow \Pr[X > 0] = 1 - o(1).$$

Theorem. (Bollobás - Erdős, 1976. Matula 1976). $\omega(G(n, 1/2)) \approx 2 \log n$.

Corollary: $\alpha(G(n, 1/2)) \approx 2 \log_2 n \Rightarrow \chi(G(n, 1/2)) \geqslant (1 + o(1)) \frac{n}{2 \log_2 n}$.