Lecture 2: Threshold for a random graph containing a cycle

For a series of events $E_1, E_2, \dots$ we say that E_n occurs *with high probability*, abbreviated *whp*, if $\mathbb{P}(E_n) \to 1$ as $n \to \infty$.

Theorem 1.6. *Let $\mathcal{A}_\circ$ be the set of graphs which contain a cycle as a subgraph then the function $p^* = \frac{1}{n}$ is a threshold for $\mathcal{A}_\circ$.*

Proof. Let $p = p(n)$ be any function such that $p/p^* \to 0$, i.e. such that $np \to 0$. Now sample the random graph $G_n \sim G(n,p)$. We want to show that whp G_n does not contain a cycle as a subgraph.

Let $X_n = X_n(G_n)$ be the random variable which counts the number of cycles in G_n. For example the number of cycles in the following graphs is #() = 1, #() = 0 and lastly #() = 3 as the graph  contains two s and the 4-cycle .

The probability that G_n has a cycle is at most the expectation of X_n:

$$\mathbb{P}(G_n \text{ has a cycle}) = \mathbb{P}(X_n > 0) = \sum_{k=1} \mathbb{P}(X_n = k) \le \sum_{k=0} k\mathbb{P}(X_n = k) = \mathbb{E}(X_n),$$

and so it will be enough to show that $\mathbb{E}(X_n) \to 0$ as $n \to \infty$.

Let $\mathcal{S}$ be the set of all places in the graph where a cycle could occur. Explicitly, $\mathcal{S}_k$ is the set of all subsets of k vertices ordered up to rotation and orientation of the cycle and $\mathcal{S} = \cup_{k\ge3}\mathcal{S}_k$. For $S \in \mathcal{S}$ define A_S to be the event that a cycle occurs on S in the random graph G_n. As expectation is linear,

$$\mathbb{E}(X_n) = \sum_{S\in\mathcal{S}} \mathbb{E}(1_{A_S}) = \sum_{k\ge3}\sum_{S\in\mathcal{S}_k} \mathbb{P}(A_S) \tag{1}$$

For $S \in \mathcal{S}_k$ the probability that a cycle occurs on S is p^k as we need each of the k independent edges which form the cycle to be present in our random graph. We want to know $|\mathcal{S}_k|$. The number of ordered sets of size k is $\binom{n}{k}k!$ - which overcounts each $S \in \mathcal{S}_k$ by $2k$ times. Why $2k$? Once for each starting position on the cycle ($\times k$), and once for each direction of the cycle ($\times 2$). Hence[2] $\mathcal{S}_k = \binom{n}{k}k!/(2k) = \binom{n}{k}(k-1)!/2$. Thus by (1),

$$\mathbb{E}(X_n) = \sum_{i\ge3}\binom{n}{k}\frac{(k-1)!}{2}p^k.$$

Now note that $\binom{n}{i}i! = n(n-1)\dots(n-i+1) \le n^i$ and we get

$$\mathbb{E}(G_n) \le \sum_{k\ge3} n^k p^k = \frac{n^3p^3}{1-np},$$

which so $E(X_n)$ goes to zero for $np \to 0$. Hence as $\mathbb{P}(G_n \text{ has a cycle}) \le \mathbb{E}(X_n)$ we have proven that whp G_n has no cycle, i.e. whp $G_n \notin \mathcal{A}_\circ$ for $p/p^* \to 0$.

For the second part of the proof we need to show that whp $G_n \in \mathcal{A}_\circ$ for $np \to \infty$. Recall (Q 2a) that any graph on n vertices with at least n edges must contain a cycle. We show that for $p = 3/n$ whp the number of edges in $G(n,p)$ is at least n. By Theorem 1.4 this implies for $p \ge 3/n$ that whp $G(n,p)$ contains a cycle as required.

[2]For the purpose of the proof it would be enough to establish that $\mathcal{S}_k \le \binom{n}{k}k!$.

Let $G_n \sim G(n, 3/n)$ and write Y_n for the number of edges in G_n. Notice $Y_n = \sum_{1\le i<j\le n} 1_{ij\in E(G_n)}$ is the sum of $\binom{n}{2}$ independent random variables each of which is 1 with probability $p = 3/n$ and 0 with probability $1-p$. Thus Y has the binomial distribution[3] of $\text{bin}(\binom{n}{2}, p)$ with expectation $\mathbb{E}(Y_n) = \binom{n}{2}p$ and variance $\mathbb{V}(Y_n) = \binom{n}{2}p(1-p)$.

The expected number of edges is $\mathbb{E}(Y_n) = \binom{n}{2}\frac{3}{n} = \frac{3n}{2}(1-1/n)$. Hence to show that whp the number of edges is at least n it is sufficient to show that whp $|\mathbb{E}(Y_n) - Y_n| < n/3$. But we can do this using Chebyshev's inequality

$$\mathbb{P}(|\mathbb{E}(Y_n) - Y_n| \ge n/3) \le \frac{\mathbb{V}(Y_n)}{(n/3)^2} = \frac{3^3}{2n}\left(1-\frac{1}{n}\right)\left(1-\frac{3}{n}\right) \to 0.$$

Hence whp $e(G_n) \ge n$ and (consequently) whp G_n contains a cycle for $np \to \infty$. □

At the end of lecture 2 we began a proof that the threshold function for having  as a subgraph is $p^* = \frac{1}{n^{2/3}}$. This will be continued in lecture 3.

[3]There are many ways to calculate the expectation and variance of the binomial random variable $\text{bin}(t,p)$ and this is not part of the course but for completeness we write out one method below.

For any random variable taking values in $\{0,1,\dots,t\}$, one can construct the polynomial (known as the probability generating function of X), $f(x) = \sum_{i=1}^{t} \mathbb{P}(X = k)x^k$ and note that $f'(x)\big|_{x=1} = \mathbb{E}(X)$ and $f''(x)\big|_{x=1} = \mathbb{E}(X(X-1))$.

Hence for the random variable Y with distribution $\text{bin}(t,p)$ the probability generating function is $f(x) = \sum_{i=1}^{t}\binom{t}{k}p^k(1-p)^{t-k}x^k = (px-(1-p))^t$ which means

$$f'(x))\big|_{x=1} = tp(x+(1-p))^{t-1})\big|_{x=1} = np,$$

and

$$f''(x))\big|_{x=1} = t(t-1)p^2(x+(1-p))^{t-2})\big|_{x=1} = t(t-1)p.$$

Thus $\mathbb{E}(Y) = tp$ and the variance is

$$\mathbb{V}(Y) = \mathbb{E}(Y(Y-1)) + \mathbb{E}(Y) - \mathbb{E}(Y^2) = t(t-1)p^2 + np - t^2p^2 = tp(1-p).$$

Exercise 5. Show whp $np \to \infty$ implies whp G_n contains  i.e. a 3-cycle[4].

Let Y_n count the number of  in G_n and for any 3-subset of vertices $S \subset V(G)$ let A_S be the event that G_n restricted to the vertices S is a .

(a) Show by linearity of expectation that:

$$\mathbb{V}(Y_n) = \sum_{S,T \in \binom{[n]}{3}} \Big(\mathbb{P}(A_S \,\&\, 1_{A_T}) - \mathbb{P}(A_S)\mathbb{P}(A_T) \Big).$$

(b) Notice that when the sets of vertices S and T don't intersect that the events A_S and A_T are independent. What about when they intersect on one vertex? Using (a) show that:

$$\mathbb{V}(Y_n) \leq \sum_{|S \cap T| = \{2,3\}} \mathbb{P}(A_S \,\&\, A_T).$$

(c) After some case analysis and from (b) show: $\mathbb{V}(Y_n) \leq n^4p^5 + n^3p^3$.

(d) From (c) conclude that whp $Y_n > 0$. *Hint: use Chebyshev's inequality.*

Exercise 6. Given $k \in \mathbb{N}$, let $\mathcal{P}_k$ be the set of graphs which have a path on k vertices as a subgraph. Find the threshold function for $\mathcal{P}_3$ (containing the path  as a subgraph) and for $\mathcal{P}_4$. Can you find the threshold for $\mathcal{P}_k$ in terms of k and n?

[4] This exercise demonstrates a different way to prove the second part of Theorem 1.6. In the proof we showed that whp $e(G_n) \geq n$ for $np \to \infty$ and from this and Q 2a we concluded that $np \to \infty$ implies whp G_n has a cycle.