Lecture 3: Thresholds for general montone sets of graphs.

We are starting to get a picture of what the random graph $G(n,p)$ looks like for different edge probabilities. Write $p \ll p^*$ as shorthand for $p/p^* \to 0$ and $p \gg p^*$ for $p/p^* \to \infty$. In Theorem 3.6 we saw that $p \ll \frac{1}{n}$ implies whp $G(n,p)$ has no cycles and $p \gg \frac{1}{n}$ implies whp $G(n,p)$ contains a cycle (or multiple cycles) as a subgraph. We are now interested in the threshold for containing $\boxtimes$ as a subgraph. Note that $\boxtimes$ contains a cycle itself, so containing a $\boxtimes$ implies containing a cycle. Hence, the threshold function (if it exists) for containing $\boxtimes$ we would expect it to be at least as big as $\frac{1}{n}$. It turns out to be $\frac{1}{n^{2/3}}$. We prove this in Theorem 3.7.

Threshold for containing $\boxtimes$ as a subgraph

Theorem 3.7. *The function $p^*(n) = \frac{1}{n^{2/3}}$ is a threshold for $G(n,p)$ containing $\boxtimes$ as a subgraph.*

Proof. Let p be such that $p/p^* \to 0$, i.e. $n^{2/3}p \to 0$ and let $G_n \sim G(n,p)$. Let $Y_n = Y_n(G_n)$ count the number of $\boxtimes$ in G_n. For each set S of 4 vertices from $[n]$, let A_S be the event that G_n restricted to the vertices S is a $\boxtimes$. We can write Y_n in terms of these indicator random variables.

$$Y_n = \sum_{S \in \binom{[n]}{4}} 1_{A_S}$$

As expectation is linear, the expected number of $\boxtimes$ in G_n is

$$\mathbb{E}(Y_n) = \sum_{S \in \binom{[n]}{4}} \mathbb{E}(1_{A_S}) \leq n^4 p^6 = (n^{2/3}p)^6.$$

Hence $\mathbb{E}(Y_n) \to 0$ for $n^{2/3}p \to 0$. Observe $\mathbb{P}(G_n \text{ contains a } \boxtimes) = \mathbb{P}(Y_n > 0) \leq \mathbb{E}(Y_n)$ and so for $n^{2/3}p \to 0$ whp G_n does not contain $\boxtimes$ as a subgraph.

Now it remains to show that for $p/p^* \to \infty$, i.e. for $n^{2/3}p \to \infty$ that whp $G_n \sim G(n,p)$ contains a $\boxtimes$. For this part of the proof we calculate the variance of Y_n by writing $Y_n = \sum_S 1_{A_S}$ and expanding. Write $\sum_S$ for $\sum_{S \in \binom{[n]}{4}}$.

$$\mathbb{V}(Y_n) = \mathbb{E}(Y_n^2) - \mathbb{E}(Y_n)^2 \quad = \quad \mathbb{E}\Big(\Big(\sum_S 1_{A_S}\Big)^2\Big) - \Big(\sum_S \mathbb{E}(1_{A_S})\Big)^2.$$

We can rearrange a little to get an expression for the variance in terms of the probabilities of the events A_S and A_T

$$\begin{aligned} \mathbb{V}(Y_n) &= \mathbb{E}\Big(\sum_S 1_{A_S} \sum_{T \in [n]^3} 1_{A_T}\Big) - \sum_S \mathbb{E}(1_{A_S}) \sum_T \mathbb{E}(1_{A_T}) \\ &= \sum_{S,T} \Big(\mathbb{E}(1_{A_S} 1_{A_T}) - \mathbb{E}(1_{A_S})\mathbb{E}(1_{A_T})\Big) \\ &= \sum_{S,T} \Big(\mathbb{P}(A_S \,\&\, A_T) - \mathbb{P}(A_S)\mathbb{P}(A_T)\Big). \end{aligned} \tag{2}$$

If $S \cap T = \varnothing$, i.e. the vertex subsets S and T are disjoint then the events A_S and A_T are independent. Notice this is also true if S and T intersect in one vertex because they still share no edges in common. Hence if $|S \cap T| \leq 1$ then $\mathbb{P}(A_S \,\&\, A_T) = \mathbb{P}(A_S)\mathbb{P}(A_T)$ and these terms cancel in the expression for the variance (2) above.

So by this observation and (2),

$$\mathbb{V}(Y_n) \quad \leq \sum_{|S\cap T|=\{2,3,4\}} \mathbb{P}(A_S \ \&\ A_T). \tag{3}$$

We now consider the three options: $|S \cap T| = 2, 3, 4$. For each of these, for $S, T \in \binom{[n]}{4}$ with the given intersection we want to calculate $\mathbb{P}(A_S \ \&\ A_T)$. For $|S \cap T| = 2$, one edge is shared. There are 10 other edges that need to be present in order to have  on both S and on T. Hence $\mathbb{P}(A_S \ \&\ A_T) = p^{11}$ for $|S \cap T| = 2$. Similarly, for $|S \cap T| = 3$, we get $\mathbb{P}(A_S \ \&\ A_T) = p^9$ and for $|S \cap T| = 4$, we get $\mathbb{P}(A_S \ \&\ A_T) = \mathbb{P}(A_S) = p^6$.

The aim is to find an upper bound for the right hand side of (3). Hence we want to know how many $S, T \in \binom{[n]}{4}$ for each of the possible overlaps. When S and T overlap on 2 vertices, the number of ways to pick them is to first pick the set of vertices in S then pick the two vertices in S that will ovelap with T, and lastly pick the last two vertices in T (the ones that don't overlap with S). This makes $\binom{n}{4}\binom{4}{2}\binom{n}{2}$. Actually all we need is that the number of $S, T \in \binom{[n]}{4}$ which overlap on two vertices is at most n^6. Similarly the number that overlap on three vertices is at most n^5 and the number overlapping on all four vertices is at most n^4.

We can now that calculate an explicit upper bound on our variance. From (2),

$$\mathbb{V}(Y_n) \quad \leq \quad n^6p^11 + n^5p^9 + n^4p^6. \tag{4}$$

Now we have a good upper bound on the variance. What we actually want to show is that whp G_n contains a . In other words we want to show whp $Y_n > 0$.

We use the following non-obvious idea. I have some b for which I know $b > 0$ and I want to use this to show that $a > 0$. Notice it is enough to show that $|b - a| < b$.

Let's go. By some re-arraning and Chebyshev,

$$\mathbb{P}(Y_n > 0) \geq \mathbb{P}\Big(|Y_n - \mathbb{E}(Y_n)| < \mathbb{E}(Y_n)/2\Big) = 1 - \mathbb{P}\Big(|Y_n - \mathbb{E}(Y_n)| \geq \mathbb{E}(Y_n)/2\Big) \geq 1 - \frac{4\mathbb{V}(Y_n)}{\mathbb{E}(Y_n)^2}.$$

The problem is now reduced to terms we have already calculated. By (4),

$$\mathbb{P}(Y_n > 0) \geq 1 - \frac{n^6p^11 + n^5p^9 + n^4p^6}{\binom{n}{4}p^6}. \tag{5}$$

For $n^{2/3}p \to \infty$ the fraction in (5) goes to zero. Hence for $n^{2/3}p \to \infty$ whp $G(n,p)$ contains a  as a subgraph. $\square$

All 'nice' sets of graphs have a threshold function

In the following theorem, we say a graph property $\mathcal{A}$ is *non-trivial* if for each large enough n it is neither always true or always false. This is equivalent to saying that $\exists n_0$ such that the complete graphs on $n > n_0$ vertices has the property ($K_n \in \mathcal{A}$) and the empty graph on $n > n_0$ vertices does not ($\overline{K}_n, \notin \mathcal{A}$).

Theorem 3.8 (Bollobás-Thomason). *Any monotone non-trivial graph property has a threshold function p^*.*

Proof. (non-examinable)
Let $p_0 = p_0(n)$ be such that $\mathbb{P}(G(n, p_0) \in \mathcal{A}) = 1/2$. Note this must exist by the intermediate value theorem because for each integer n, $f_n(p) = \mathbb{P}(G(n, p_0) \in \mathcal{A})$ is a polynomial in p (of degree at most $\binom{n}{2}$) such that $f_n(0) = 0$ and $f_n(1) = 1$.

Let the threshold function be this p_0, we set $p^*(n) = p_0(n)$. It will be sufficient to prove that $\forall \varepsilon > 0$, if $p/p_0 \to 0$ then $\mathbb{P}(G(n, p) \in A) < \varepsilon$ and if $p/p_0 \to \infty$ then $\mathbb{P}(G(n, p) \in A) > 1 - \varepsilon$.

Fix $\varepsilon > 0$. We first prove that $\exists p_b$ such that $\mathbb{P}(G(n, p_b) \in \mathcal{A}) > 1 - \varepsilon$.

Consider the union of k copies of $G(n, p_0)$, for some k which we will decide later. Let $H = ([n], E(G_1) \cup \ldots \cup E(G_k))$ where each $G_i \sim G(n, p_0)$. Here the graphs G_i are all defined on the same vertex set $[n]$, and H is the random graph on this vertex set with edge set the union of the edge sets of the G_i. For any given $i \neq j \in [n]$ the probability ij is not in the edge set of H is exactly the probability that the edge ij does not appear in any of the G_i, which is $(1 - p)^k$. This means H is the random graph where each edge is present independently with probability $1 - (1 - p)^k$. Thus $H \sim G(n, 1 - (1 - p)^k)$.

The next idea is to notice that $\mathcal{A}$ monotone means that $H \in \mathcal{A}$ if $\exists i$ such that $G_i \in \mathcal{A}$. Thus,

$$\mathbb{P}(H \in \mathcal{A}) \leq 1 - \mathbb{P}(\forall i, G_i \notin \mathcal{A}) = 1 - \mathbb{P}(G_1 \notin \mathcal{A})^k = 1 - \frac{1}{2^k}. \tag{6}$$

Recall that we are still free to choose k, set $k = \lceil \log_2(1/\varepsilon) \rceil$ and let $p_1 = 1 - (1 - p)^k$. Now by our choices of p_1 and k the equation on line (6) says precisely that $\mathbb{P}(G(n, p_1) \in \mathcal{A}) > 1 - \varepsilon$.

Claim 3.9. *For all $\varepsilon > 0$ there exists function $p_\ell(n) = p_\ell(n, \varepsilon)$ such that $\mathbb{P}(G(n, p) \in \mathcal{A}) < \varepsilon$.*

Proof. (exercise) □

Claim 3.10. *For all $\varepsilon > 0$, there exists $N = N(\varepsilon)$ such and if $p/p_0 \to 0$ then (for all $n > N$) $p(n) < p_\ell(n, \varepsilon)$ and if $p/p_0 \to \infty$ then (for all $n > N$) $p(n) > p_b(n, \varepsilon)$.*

Proof. (exercise) □

With the help of these claims we can now finish the proof. Suppose $p/p_0 \to 1$ (we want to show $P(G(n, p) \in \mathcal{A}) > 1 - \varepsilon$). By Claim 3.10 $p/p_0 \to 1$ implies $p(n) > p_b(n, \varepsilon)$. Therefore by Theorem 3.4 $P(G(n, p) \in \mathcal{A}) \geq \mathbb{P}(G(n, p_b) \in \mathcal{A})$. But by our choice of $p_b(n) = p_b(n, \varepsilon)$ we know that $\mathbb{P}(G(n, p_b) \in \mathcal{A}) > 1 - \varepsilon$ and so $P(G(n, p) \in \mathcal{A}) > 1 - \varepsilon$ as required.

Now suppose $p/p_0 \to 0$ (we want to show $P(G(n, p) \in \mathcal{A}) < \varepsilon$). By Claim 3.10 $p/p_0 \to 1$ implies $p(n) < p_\ell(n, \varepsilon)$. The monotonicity result, Theorem 3.4, implies $P(G(n, p) \in \mathcal{A}) \leq \mathbb{P}(G(n, p_b) \in \mathcal{A})$. But by Claim 3.9, $P(G(n, p_b) \in \mathcal{A}) < \varepsilon$ and so we are done. □