4.2 Lecture 21 (19/Nov): $G(n,p)$ thresholds

4.2.1 Threshold for H as a subgraph in $G(n,p)$

Working with low moments of random variables can be incredibly effective, even when we are not specifically looking for limited-independence sample spaces. Here is a prototypical example. "When" does a specific, constant-sized graph H, show up as a subgraph of a random graph selected from the distribution $G(n,p)$? We have in mind that we are "turning the knob" on p. If H has any edges then when $p = 0$, with probability 1 there is no subgraph isomorphic to H. When $p = 1$, with probability 1 such subgraphs are everywhere [2]. In between, for any finite n, the probability is some increasing function of p. But we won't take n finite, we will take it tending to ∞.

So the question is,[3] can we identify a function $\pi(n)$ such that in the model $G(n, p(n))$, with $[\![H]\!]$ denoting the event that there is an H in the random graph G,

(a) If $p(n) \in o(\pi(n))$, then $\lim_n \Pr([\![H]\!]) = 0$.

(b) If $p(n) \in \omega(\pi(n))$, then $\lim_n \Pr([\![H]\!]) = 1$.

Such a function $\pi(n)$ is known as the *threshold* for appearance of H. It follows from work of Bollobas and Thomason [19] that monotone events—events that must hold in G' if they hold in some $G \subseteq G'$—always have a threshold function.

(A related but incomparable statement: for a monotone graph property, i.e., a monotone property invariant under vertex permutations, for any $\varepsilon > 0$ there is a $p(n)$ such that $\Pr_{p(n)}(\text{property}) \leq \varepsilon$ and $\Pr_{p(n)+O(1/\log n)}(\text{property}) \geq 1 - \varepsilon$. See [42].)

4.2.2 Most pairs independent: threshold for K_4 in $G(n,p)$

Let $S \subset \{1, \ldots, n\}$, $|S| = 4$. Let X_S be the event that K_4 occurs as a subgraph of G at S—that is, when you look at those four vertices, all the edges between them are present. Conflating X_S with its indicator function and letting X be the number of K_4's in G, we have

$$X = \sum_S X_S$$

and

$$E(X) = \binom{n}{4} p^6.$$

We are interested in $\Pr(X > 0)$. Let $\pi(n) = n^{-2/3}$.

(a) For $p(n) \in o(\pi(n))$, $E(X) \in o(1)$, so $\Pr([\![K_4]\!]) \in o(1)$ and therefore $\lim_n \Pr([\![K_4]\!]) = 0$.

(b) For $1 > p(n) \in \omega(\pi(n))$, $E(X) \in \omega(1)$. We'd like to conclude that likely $X > 0$ but we do not have enough information to justify this, as it could be that X is usually 0 and occasionally very large.[4] We will exclude that possibility for K_4 by studying the next moment of the distribution.

Before carrying out this calculation, though, we have to make one important note. Since the event $[\![K_4]\!]$ is monotone, $[p \leq p'] \Rightarrow [\Pr_{G(n,p)}[\![K_4]\!] \leq \Pr_{G(n,p')}[\![K_4]\!]]$. (An easy way to see this is by choosing reals iid uniformly in $[0,1]$ at each edge, and placing the edge in the graph if the rv is above the p

[2] Today we focus on $H = K_4$, the 4-clique, but more generally this method will establish the probability of any fixed graph H occurring as a subgraph in G, that is, there is an injection of vertices of H into the vertices of G such that every edge of H is present in G. This is different from asking that H occur as an *induced* subgraph of G, which requires also that non-edges of H be non-edges in G. That is an interesting question too but different in an essential way: the event is not monotone in G.

[3] Recall $p(n) \in o(\pi(n))$ means that $\limsup p(n)/\pi(n) = 0$, and $p(n) \in \omega(\pi(n))$ means that $\limsup \pi(n)/p(n) = 0$.

[4] When we study not K_4-subgraphs, but other subgraphs, this can really happen. We'll discuss this below.

or p' threshold.) This means that it is enough to show that K_4 "shows up" slightly above π. This is useful because some of our calculations break down far above π, not because there is anything wrong with the underlying statement but because the inequalities we use are not strong enough to be useful there and a direct calculation would need to take account of further moments.

To simplify our remaining calculations, then, let

$$p = n^{-2/3}g(n), \quad \text{so } n^4p^6 = g^6$$

for any sufficiently small $g(n) \in \omega(1)$; we'll see how this is helpful in the calculations.

By an earlier exercise,

$$\mathrm{Var}(X) = \sum_S \mathrm{Var}(X_S) + \sum_{S \neq T} \mathrm{Cov}(X_S, X_T)$$

X_S is a coin (or Bernoulli rv) with $\Pr(X_S = 1) = p^6$. The variance of such an rv is $p^6(1-p^6)$.

The covariance terms are more interesting.

1. If $|S \cap T| \leq 1$, no edges are shared, so the events are independent and $\mathrm{Cov}(X_S, X_T) = 0$.
2. If $|S \cap T| = 2$, one edge is shared, and a total of 11 specific edges must be present for both cliques to be present. A simple way to bound the covariance is (since $E(X_S), E(X_T) \geq 0$) that $\mathrm{Cov}(X_S, X_T) = E(X_S X_T) - E(X_S)E(X_T) \leq E(X_S X_T) = p^{11}$.
3. If $|S \cap T| = 3$, three edges are shared, and a specific 9 edges must be present for both cliques to be present. Similarly to the previous case, $\mathrm{Cov}(X_S, X_T) \leq p^9$.

$$\begin{aligned}
\mathrm{Var}(X) &\leq \binom{n}{4}p^6(1-p^6) + \binom{n}{2,2,2}p^{11} + \binom{n}{3,1,1}p^9 \\
&\in O(n^4p^6 + n^6p^{11} + n^5p^9) \\
&= O(g^6n^{4-4} + g^{11}n^{6-22/3} + g^9n^{5-6}) \quad \text{from } p = n^{-2/3}g(n) \\
&= O(g^6 + g^{11}n^{-4/3} + g^9n^{-1}) \qquad (4.4) \\
&= O(g^6) \quad \text{provided } g^5 \in O(n^{4/3}) \text{ and } g^3 \in O(n)
\end{aligned}$$

This gives us the key piece of information. For $g \in \omega(1)$ but not *too* large, we have

$$\frac{\mathrm{Var}(X)}{(E(X))^2} \in \frac{O(g^6)}{\Theta((n^4p^6)^2)} = \frac{O(g^6)}{\Theta(g^{12})} = O(g^{-6}) \subseteq o(1)$$

and we have only to apply the Chebyshev inequality (Cor. 14) or better yet Paley-Zygmund 74 to conclude that $\Pr(X = 0) \in o(1)$ and so

$$\lim_n \Pr([\![K_4]\!]) = 1. \qquad (4.5)$$

Since $[\![K_4]\!]$ is a monotone event, (4.5) holds even for g above the range we needed for the calculation to hold. (Note, though, since there is so much "room" in the calculation, we could even have used the upper bound $O(g^11)$ on 4.4, and not resorted to this monotonicity argument.)

Exercise: Show that the threshold for appearance, as a subgraph, of the graph with 5 edges and 4 vertices is $n^{-4/5}$.

Comment: For a general H the threshold for appearance of H in $G(n,p)$ as a subgraph is determined not by the ratio ρ_H of edges to vertices, but by the maximum of this ratio over induced subgraphs of H, call it $\rho_{\max H}$. We'll see this on a problem set (and see [8]). If these numbers are different then above n^{-1/ρ_H} the expected number of H's starts tending to ∞ but almost certainly we have none; once we cross the higher threshold $n^{-1/\rho_{\max H}}$, there is an "explosion" of many of these subgraphs appearing. (They show up highly intersecting in the fewer copies of the critical induced subgraph.)