Threshold for appearance of any constant H in $G(n,p)$

These notes are following the beginning of Chapter 5 in Frieze & Karoński's Introduction to Random Graphs, which is freely available online.

For any complete subgraph K_t, let X be the number of copies of K_t in $G(n,p)$. Then we have

$$\mathbb{E}(X) = \binom{n}{t} p^{\binom{t}{2}} \sim n^t p^{\binom{t}{2}}.$$

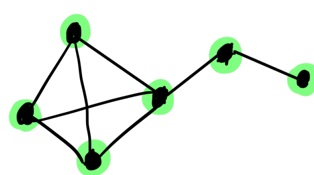
This implies that when $p \ll n^{-t/\binom{t}{2}}$ we have $\mathbb{E}(X) \to 0$, which shows (by the first moment method) that $\mathbb{P}(X = 0) \to 1$. When $p \gg n^{-t/\binom{t}{2}}$, we have $\mathbb{E}(X) \to \infty$, and we can use the second moment method to show that $\mathbb{P}(X = 0) \to 0$.

In the case of complete graphs K_t, everything works as expected. For example, $p^* = n^{-4/6}$ is a threshold for the appearance of K_4.

Things get a little more complicated for the appearance of general subgraphs H. Let X_H be the number of copies of H in $G(n,p)$. First of all, H can appear in different ways on the same set of vertices (unlike K_t). Let a symmetry or automorphism of H be a permutation of the vertices that preserves edge relations. Let $\mathrm{Aut}(H)$ be the set of all such automorphisms. (In fact, this forms a group, but that is not important here.) If H has n_H vertices, then H can appear on a set of n_H vertices in $n_H!/|Aut(H)|$ ways. Please verify this for yourself if needed. This exact number will not end up mattering for the derivation, since it is a constant, but it gives us an accurate expression for the expected value of X_H to start with. We have

$$\mathbb{E}(X_H) = \binom{n}{n_H} \frac{n_H!}{|Aut(H)|} p^{m_H} \sim n^{n_H} p^{m_H},$$

where m_H is the number of edges of H. Perhaps, just as with the complete graphs, we can obtain that $p^* = n^{-n_H/m_H}$ is a threshold for the appearance of H? This turns out not to be the case. For example, consider the following graph:



This graph has 6 vertices and 8 edges, so we would expect that a threshold of $p^* = n^{-6/8}$ will work. It is true that when $p \ll p^*$ we have $\mathbb{E}(X_H) \to 0$ and $\mathbb{P}(X = 0) \to 1$ as usual. However, note that, for example $4/6 < 5/7 < 6/8$. If we set $p = 5/7$, we have $p \gg p^*$, but also $p \ll n^{-4/6}$. This implies that with high probability, $G(n,p)$ has no copies of K_4, so it cannot have any copies of H either (since $K_4 \subseteq H$). The reason that this does not graph does not work as expected is that its edges are clustered: it is unbalanced. To help gain intuition, consider the following two graphs, both on 4 vertices and 2 edges. Do you expect that they have the same threshold of appearance?

In the following theorem, we'll show that the appearance threshold of subgraphs depends only on their densest parts, since this part is the last to appear in $G(n,p)$ (as we increase p). First, let

$$d(H) = \frac{m_H}{n_H}.$$

This looks a lot like edge density, but note that edge density is usually defined as $m_H \binom{n_H}{2}^{-1}$. Furthermore, we let

$$d^*(H) = \max_{H' \subseteq H} d(H').$$

We say that a graph H is balanced if $d^*(H) = d(H)$.

Theorem 1. *For a fixed graph H with at least one edge, let X_H be the number of copies in $G(n,p)$. We have*

$$\mathbb{P}(X = 0) \to \begin{cases} 1, & \text{if } p \ll n^{-1/d^*(H)}, \\ 0, & \text{if } p \gg n^{-1/d^*(H)}. \end{cases}$$

Exercise 1. *Prove the first part of Theorem 1, for the case $p \ll n^{-1/d^*(H)}$.* 2

For the second part, we will need to show that when $p \gg n^{-1/d^*(H)}$, we have $\mathrm{var}(X_H) \ll \mathbb{E}(X_H)^2$. Let

$$X_H = Y_1 + \cdots + Y_s,$$

where $s = \binom{n}{n_H} \frac{n_H}{|Aut(H)|}$ is the total number of possible copies of H, and each Y_i is the indicator random variable of its appearance.

Exercise 2. *Show that we can write* 1

$$\mathrm{var}(X_H) = \sum_{i=1}^{s} \sum_{j=1}^{s} \mathbb{P}(Y_i = 1, Y_j = 1) - \mathbb{P}(Y_i = 1)\mathbb{P}(Y_j = 1).$$

Exercise 3. *Next, show that we can write* 2

$$\mathrm{var}(X_H) = \sum_{H' \subseteq H, m_{H'} > 0} O(n^{2n_H - n_{H'}}(p^{2m_H - m_{H'}} - p^{2m_H})).$$

Exercise 4. *Finish the proof of Theorem 1.* 3

Exercise 5. *According to Theorem 1, any cycle C_k of constant length k behaves "the same" in terms of the threshold of appearance. Note that this only describes asymptotic behavior and ignores constants. Perform simulations on a set of values of n, p, k, and describe any observed differences between cycles of different lengths.* 3

Threshold for appearance of any cycle in $G(n,p)$

According to Theorem 1, the threshold function for a cycle C_k of constant length k is $p^* = \frac{1}{n}$. Now, we ask a slightly different question: what is the threshold for the appearance of **any** cycle, including long ones, in $G(n,p)$? It turns out that this also $p^* = \frac{1}{n}$.

Theorem 2. *Let $X_\circ$ be the number of cycles in $G(n,p)$. Then,*

$$\mathbb{P}(X_\circ = 0) \to \begin{cases} 1, & \text{if } p \ll \frac{1}{n}, \\ 0, & \text{if } p \gg \frac{1}{n}. \end{cases}$$

Exercise 6. *Show that* 1

$$\mathbb{E}(X_\circ) = \sum_{k=3}^{n} \binom{n}{k} \frac{(k-1)!}{2} p^k.$$

Exercise 7. *Show that* 1

$$\mathbb{E}(X_\circ) \leq \sum_{k=3}^{n} n^k p^k.$$

Exercise 8. *Prove the first part of Theorem 2, for the case when* $p \ll \frac{1}{n}$. 1

Estimating the variance of $X_\circ$ is trickier, but we will use a different trick for the second part. When $p = \frac{1}{n}$, the expected number of edges in $G(n,p)$ is around $n/2$. Note that when a graph has $\geq n$ edges, it **must** have a cycle. If we let $p = \frac{2+\epsilon}{n}$ for $\epsilon > 0$, then the expected number of edges is around $(1 + \epsilon/2)n$.

Exercise 9. *Prove the second part of Theorem 2, for the case when* $p \gg \frac{1}{n}$. *Note that you* 1
can actually prove a stronger statement here.

Solutions to Selected Exercises

Exercise 1. *Prove the first part of Theorem 1, for the case* $p \ll n^{-1/d^*(H)}$.

Let $H^* \subseteq H$ be a subgraph of H such that $d(H^*) = d^*(H)$. Let

$$p \ll n^{-1/d^*(H)} = n^{-n_{H^*}/m_{H^*}}.$$

We have

$$\mathbb{E}(X_{H^*}) \sim n^{n_{H^*}} p^{m_{H^*}} \to 0.$$

By the first moment method, this implies that $\mathbb{P}(X_{H^*} = 0) \to 1$. When a graph has no copies of H^* it also cannot have any copies of H (since $H^* \subseteq H$), and therefore $\mathbb{P}(X_H = 0) \to 1$.

Exercise 4. *Finish the proof of Theorem 1.*

We have

$$\begin{aligned}
\operatorname{var}(X_H) &= \sum_{H' \subseteq H, m_{H'} > 0} O(n^{2n_H - n_{H'}}(p^{2m_H - m_{H'}} - p^{2m_H})) \\
&= \sum_{H' \subseteq H, m_{H'} > 0} O(n^{2n_H - n_{H'}} p^{2m_H - m_{H'}}) \\
&= n^{2n_H} p^{2m_H} \sum_{H' \subseteq H, m_{H'} > 0} O(n^{-n_{H'}} p^{-m_{H'}}) \\
&= \mathbb{E}(X_H)^2 \sum_{H' \subseteq H, m_{H'} > 0} O(n^{-n_{H'}} p^{-m_{H'}})
\end{aligned}$$

For our result to work, we therefore need that

$$\sum_{H' \subseteq H, m_{H'} > 0} O(n^{-n_{H'}} p^{-m_{H'}}) \to 0. \tag{1}$$

Let $p = \omega p^* = \omega n^{-1/d^*(H)}$, where $\omega \to \infty$. We can write

$$n^{-n_{H'}} p^{-m_{H'}} = \frac{n^{\frac{m_{H'}}{d^*(H)} - n_{H'}}}{\omega^{m_{H'}}}.$$

By the definition of $d^*(H)$, we have that for any $H' \subseteq H$

$$\frac{1}{d^*(H)} \leq \frac{n_{H'}}{m_{H'}},$$

and therefore

$$\frac{1}{d^*(H)} - \frac{n_{H'}}{m_{H'}} \leq 0 \quad \Leftrightarrow \quad \frac{m_{H'}}{d^*(H)} - n_{H'} \leq 0.$$

Therefore,

$$\frac{n^{\frac{m_{H'}}{d^*(H)} - n_{H'}}}{\omega^{m_{H'}}} = O\left(\frac{1}{\omega^{m_{H'}}}\right) \to 0.$$

We are not quite done yet. We still need to sum the above over all subgraphs $H' \subseteq H$. Since there are only a constant number of such subgraphs, we obtain the result in 1, and concludde that $\operatorname{var}(X_H) \ll \mathbb{E}(X_H)^2$ and therefore $\mathbb{P}(X_H = 0) \to 0$.

Exercise 8. *Prove the first part of Theorem 2, for the case when* $p \ll \frac{1}{n}$.

Exercise 9. *Prove the second part of Theorem 2, for the case when* $p \gg \frac{1}{n}$. *Note that you can actually prove a stronger statement here.*