Random Graphs — Week 2

Yuval Filmus

October 30, 2019

1 Subgraph thresholds

Last week we proved the following theorem:

Theorem 1. *If $p = o(1/n)$ then with high probability, $G(n,p)$ contains no triangles. If $p = \omega(1/n)$ then with high probability, $G(n,p)$ contains triangles.*

Now we want to generalize this to arbitrary graphs H. The starting point is calculating the expected number of copies of H in $G(n,p)$. If H has $v(H)$ vertices and $e(H)$ edges, then the expected number of copies of H in $G(n,p)$ is proportional to

$$n^{\underline{v(H)}} p^{e(H)},$$

where $n^{\underline{k}} = n(n-1)\cdots(n-k+1)$ is the number of ways to choose k vertices from $[n]$, with order.

This formula could double count copies of H. For example, if H is a triangle, then we count each triangle 6 times. In order to get the actual expectation, we need to divide by the number of *automorphisms* of H. At any rate, the order of growth of the number of copies is

$$\Theta(n^{v(H)} p^{e(H)}).$$

This suggests the following candidate for the threshold for appearance of H:

$$p \approx n^{-v(H)/e(H)}.$$

Indeed, a first moment argument shows that if $p = o(n^{-v(H)/e(H)})$ then with high probability, $G(n,p)$ contains no copies of H. But is this the right threshold?

Consider the graphs H_ℓ, formed from K_4 by adjoining a path of length ℓ edges to one of the vertices. If $p = o(n^{-2/3})$ then with high probability $G(n,p)$ doesn't contain any copies of K_4. However,

$$\frac{v(H_\ell)}{e(H_\ell)} = \frac{4+\ell}{6+\ell} \approx 1,$$

which predicts a threshold of $n^{-(1-\epsilon)}$ for H_ℓ when ℓ is large. Something is wrong here, since $n^{-(1-\epsilon)} \ll n^{-2/3}$.

The solution is to go over all subgraphs of H, and take the one with the largest *density*. Let $d(K) = e(K)/v(K)$, and define $m(H) = \max_{K \subseteq H} d(K)$. The first moment argument shows that if $p = o(n^{-1/m(H)})$ then with high probability, $G(n,p)$ contains no copies of H.

For the opposite direction (if $p = \omega(n^{-1/m(H)})$ then with high probability $G(n,p)$ contains a copy of H), recall that we need to show that $\mathbb{V}[X] = o(\mathbb{E}[X]^2)$, where X is the expected number of copies of H in $G(n,p)$. Our starting point is the following formula for $\mathbb{V}[X]$:

$$\mathbb{V}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \sum_{H_i,H_j \in \mathcal{H}} \Pr[H_i, H_j \in G(n,p)] - \sum_{H_i,H_j \in \mathcal{H}} \Pr[H_i \in G(n,p)] \Pr[H_j \in G(n,p)] = \sum_{H_i,H_j \in \mathcal{H}} (p^{|H_i \cup H_j|} - p^{2e(H)}),$$

where $\mathcal{H}$ is the set of all potential copies of H in $G(n,p)$, and $|H_i \cup H_j|$ is the total number of edges in H_i, H_j.

What we need to do now is to estimate how many pairs (H_i, H_j) have a given number of edges in common. It makes sense to split the sum according to the isomorphism type of the intersection:

$$\mathbb{V}[X] = \sum_{K \subseteq H} \sum_{\substack{H_i,H_j \in \mathcal{H} \\ H_i \cap H_j \approx K}} (p^{2e(H)-|K|} - p^{2e(H)}) \leq \sum_{\substack{K \subseteq H \\ e(K)>0}} \sum_{\substack{H_i,H_j \in \mathcal{H} \\ H_i \cap H_j \approx K}} p^{2e(H)-e(K)}.$$

If K contains $v(K)$ edges, then H_i, H_j together cover $2v(H) - v(K)$ vertices, and so the total number of pairs H_i, H_j whose intersection is K is at most $n^{2v(H)-v(K)}$. Therefore we can estimate

$$\mathbb{V}[X] \leq \sum_{\substack{K \subseteq H \\ e(K)>0}} n^{2v(H)-v(K)} p^{2e(H)-e(K)} = (n^{v(H)} p^{e(H)})^2 \sum_{\substack{K \subseteq H \\ e(K)>0}} n^{-v(K)} p^{-e(K)}.$$

The first factor, $(n^{v(H)} p^{e(H)})^2$, is $\Theta(\mathbb{E}[X]^2)$. Therefore, to complete the proof, we need to show that the sum is $o(1)$. Since there are finitely many summands, it suffices to show that each of them is $o(1)$. Indeed,

$$n^{-v(K)} p^{-e(K)} = o(n^{-v(K)} n^{e(K)/m(H)}) = o(n^{e(K)/d(K)-v(K)}) = o(1).$$

We have proved the following result:

Theorem 2. *If $p = o(n^{-1/m(H)})$ then with high probability, $G(n,p)$ contains no copy of H. If $p = \omega(n^{-1/m(H)})$ then with high probability, $G(n,p)$ contains a copy of H.*

2 Distribution around threshold

The threshold for appearance of a triangle is around $p = 1/n$. How many triangles does $G(n,p)$ contains when $p = c/n$? We have already seen that the expected number of triangles is $\mathbb{E}[X] \sim (np)^3/6 = c^3/6$. In order to understand more about the distribution, we need to compute higher moments of X.

Let us start with the second moment, which we have essentially already computed. It will be a bit nicer to compute

$$\mathbb{E}\left[\binom{X}{2}\right] = \sum_{S<T\in\binom{[n]}{3}} \Pr[\triangle_S, \triangle_T \in G(n,p)] = \sum_{S<T\in\binom{[n]}{3}} p^{|\triangle_S \cup \triangle_T|}.$$

(Here $S < T$ is with respect to some arbitrary ordering of $\binom{[n]}{3}$.) Just as in the preceding section, we need to look at the possible ways in which $\triangle_S, \triangle_T$ can intersect:

- S, T are disjoint. There are $\frac{1}{2}\binom{n}{3}\binom{n-3}{3} \sim \frac{1}{2}n^6/6^2$ such pairs, and each one contributes p^6 to the sum, for a total of $\sim \frac{1}{2}c^6/6^2$.
- S, T intersect at a vertex. There are $\frac{1}{2}\binom{n}{3} \cdot 3 \cdot \binom{n-3}{2} = O(n^5)$ such pairs, and each one contributes p^6 to the sum, for a total of $o(1)$.
- S, T intersect at an edge. There are $\frac{1}{2}\binom{n}{3} \cdot 3 \cdot \binom{n-3}{1} = O(n^4)$ such pairs, and each one contributes p^5 to the sum, for a total of $o(1)$.

Note that S cannot equal T. In total,

$$\mathbb{E}\left[\binom{X}{2}\right] \sim \frac{1}{2}\frac{c^6}{6^2} \sim \frac{\mathbb{E}[X]^2}{2}.$$

We can estimate higher moments in a similar way:

$$\mathbb{E}\left[\binom{X}{k}\right] = \sum_{S_1<\cdots<S_k} p^{|S_1\cup\cdots\cup S_k|}.$$

The only non-negligible contribution is going to be from disjoint $S_1, \ldots, S_k$, of which there are $\sim (n^3/6)^k/k!$ tuples, implying that

$$\mathbb{E}\left[\binom{X}{k}\right] \sim \frac{1}{k!}\left(\frac{c^3}{6}\right)^k \sim \frac{\mathbb{E}[X]^k}{k!}.$$

Essentially, what this means is that X behaves as the sum of *independent* random variables. Indeed, consider a random variable Y which is the sum of m Bernoulli random variables Y_i with $\Pr[Y_1 = 1] = \lambda/m$. Then $\mathbb{E}[Y] = \lambda$, and more generally,

$$\mathbb{E}\left[\binom{Y}{k}\right] = \sum_{1\le i_1<\cdots<i_k\le m} \lambda^k = \binom{m}{k}\left(\frac{\lambda}{m}\right)^k \sim \frac{\lambda^k}{k!}.$$

What is the distribution of Y? We can estimate it explicitly:

$$\Pr[Y = t] = \binom{m}{t}\left(\frac{\lambda}{m}\right)^t\left(1-\frac{\lambda}{m}\right)^{m-t} \sim e^{-\lambda}\frac{\lambda^t}{t!}.$$

This distribution is known as the *Poisson distribution*.

The calculations above suggest that the number of triangles in $G(n, c/n)$ should have roughly Poisson distribution, with $\lambda = c^3/6$. The proof is a small exercise in calculus.

We will concentrate on calculating the probability of having no triangles. The idea is to use the inclusion-exclusion formula:

$$\Pr[X = 0] = 1 - \sum_{S_1 \in \binom{[n]}{3}} \Pr[\triangle_{S_1} \in G(n,p)] + \sum_{S_1 < S_2 \in \binom{[n]}{3}} \Pr[\triangle_{S_1}, \triangle_{S_2} \in G(n,p)] + \cdots .$$

In fact, we will need to use a stronger form of the formula, in which we cut the sum in the middle. The stronger form states that the direction of the error matches that of the following term. So for example

$$\begin{aligned} \Pr[X = 0] &\leq 1 \\ &\geq 1 - \sum_{S_1 \in \binom{[n]}{3}} \Pr[\triangle_{S_1} \in G(n,p)] \\ &\leq 1 - \sum_{S_1 \in \binom{[n]}{3}} \Pr[\triangle_{S_1} \in G(n,p)] + \sum_{S_1 < S_2 \in \binom{[n]}{3}} \Pr[\triangle_{S_1}, \triangle_{S_2} \in G(n,p)] \end{aligned}$$

and so on. Observe that the ℓth summand is exactly $\mathbb{E}[\binom{X}{\ell}]$. Now comes the calculus part.

Denote the ℓth partial summand by Σ_ℓ ($\Sigma_0 = 1$, and so on), so that $\Sigma_{2\ell+1} \leq \Pr[X = 0] \leq \Sigma_{2\ell}$. The calculations above show that

$$s_\ell := \lim_{n\to\infty} \Sigma_\ell = 1 - \lambda + \frac{\lambda^2}{2} - \cdots \pm \frac{\lambda^k}{k!}.$$

We also know that

$$\lim_{\ell\to\infty} s_\ell = \sum_{k\geq 0} (-1)^k \frac{\lambda^k}{k!} = e^{-\lambda}.$$

Therefore, for every $\epsilon > 0$ we can find ℓ so that $|s_{2\ell} - e^{-\lambda}|, |s_{2\ell+1} - e^{-\lambda}| \leq \epsilon/2$. We can also find N such that for $n \geq N$, $|\Sigma_{2\ell} - s_{2\ell}|, |\Sigma_{2\ell+1} - s_{2\ell+1}| \leq \epsilon/2$. In total, we deduce that for $n \geq N$,

$$\Pr[X = 0] \leq \Sigma_{2\ell} \leq s_{2\ell} + \frac{\epsilon}{2} \leq e^{-\lambda} + \epsilon,$$

and similarly $\Pr[X = 0] \geq e^{-\lambda} - \epsilon$. In total, $|\Pr[X = 0] - e^{-\lambda}| \leq \epsilon$. Since this holds for every $\epsilon > 0$, it follows that $\Pr[X = 0] \to e^{-\lambda}$.

Using similar ideas, one can show that $\Pr[X = t] \to e^{-\lambda}\lambda^t/t!$ for every constant t, matching a Poisson distribution. Using some more calculus, it follows that X converges (in any reasonable metric) to a Poisson distribution with mean λ.

3 Where do we go from here?

One obvious question is generalizing the Poisson limit law to other graphs. Such a law doesn't hold for all graphs, but it does hold for all *strictly balanced* graphs, which are graphs for which $m(H) = d(H)$ (*balanced*), and moreover the maximum is achieved uniquely.

A different question is what happens when p is above the threshold. When $np \to \infty$ and $n^2(1-p) \to \infty$, the distribution of the number of triangles is roughly normal, in the sense that for every fixed t,

$$\Pr\left[\frac{X - \mathbb{E}[X]}{\sqrt{\mathbb{V}[X]}} < t\right] \to \Pr[N(0,1) < t].$$

One is also interested in large deviation properties. For each $\epsilon > 0$, the probability that $X < (1-\epsilon)\,\mathbb{E}[X]$ or that $X > (1+\epsilon)\,\mathbb{E}[X]$ behaves asymptotically as e^{-Cn^2}, where C depends on ϵ and on the direction. The lower tail ($X < (1-\epsilon)\,\mathbb{E}[X]$) is much easier to analyze than the upper tail.