Random Graphs — Week 3

Yuval Filmus

November 2, 2019

1 Moments of number of triangles

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 asymptotically non-vanishing contribution is going to come from k-tuples of disjoint triangles. To see this, we break the sum according to the isomorphism type of $S_1 \cup \cdots \cup S_k$:

$$\mathbb{E}\left[\binom{X}{k}\right] = \sum_T \sum_{\substack{S_1<\cdots<S_k \\ S_1\cup\cdots\cup S_k\approx T}} p^{e(T)} = \sum_T N_T p^{e(T)},$$

where N_T is the number of potential copies of T. If T is a union of k disjoint triangles, then

$$N_T = \frac{1}{k!}\binom{n}{3}\binom{n-3}{3}\cdots\binom{n-3(k-1)}{3} \sim \frac{1}{k!}\left(\frac{n^3}{6}\right)^k,$$

and so the corresponding summand contributes $\frac{1}{k!}(n^3/6)^k(c/n)^{3k} = \frac{1}{k!}(c^3/6)^k$ to the sum.

For the other summands, since T is a union of triangles, each vertex has degree at least 2, and so $2e(T) \geq 2v(T)$ (since $2e(T)$ is the sum of the degrees). If T is not a disjoint union, then one of the degrees is larger than 2, and so $e(T) > v(T)$. The contribution of any such T to the sum is at most $n^{v(T)}p^{e(T)} = O(n^{v(T)-e(T)}) = o(1)$. Since there are finitely such T, we conclude 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,

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

2 Poisson approximation

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 first calculate 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, known as the Bonferroni inequalities, 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 (we prove this later). 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 $\Pr[X = 0] \to e^{-\lambda}$.

3 Rest of the distribution

The Bonferroni inequalities can be extended to $\Pr[X = t]$ for arbitrary t. Here is an alternative route for analyzing this probability.

We know that with high probability, all triangles $G(n, c/n)$ are vertex disjoint. This is because the density of the following two graphs is larger than 1: two triangles sharing a vertex, and two triangles sharing an edge. Hence roughly speaking, for the graph to contain t triangles, there need to be t disjoint triangles that it contains, the other $n - 3t$ vertices supporting no triangles.

In order to formalize this idea, we will bound $\Pr[X = t]$ from both directions, starting with the upper bound. If $G \sim G(n, c/n)$ contains exactly t triangles, then either G contains two triangles which are non-vertex-disjoint (this happens with probability $o(1)$), or it contains t disjoint triangles, the other $n-3t$ vertices supporting no triangles. Therefore

$$\Pr[X = t] \leq o(1) + \frac{1}{t!}\binom{n}{3}\binom{n-3}{3}\cdots\binom{n-3(t-1)}{3}\left(\frac{c}{n}\right)^{3t} \Pr[G(n-3t, c/n) \text{ is triangle-free}].$$

Carefully repeating our calculations in the preceding section, we see that the probability that $G(n - 3t, c/n)$ is triangle-free tends to $e^{-c^3/6}$. Therefore the second summand is asymptotic to $\frac{1}{k!}(c^3/6)^k e^{-c^3/6}$. In total,

$$\Pr[X = t] \leq e^{-c^3/6} \frac{(c^3/6)^k}{k!} + o(1).$$

For the lower bound, for every k-tuple τ of vertex-disjoint triangles we construct an event E_τ in which the only triangles are τ. Since the events are disjoint, it will follow that $\Pr[X = t] \geq \sum_\tau \Pr[E_\tau]$. The event E_τ is as follows:

1. The triangles in τ belong to G, and there are no other edges among these $3k$ vertices.

2. The graph on the remaining $n - 3k$ vertices is triangle-free, and contains at most $n \log n$ edges.

3. There are no other triangles in G.

The probability that the first condition holds is

$$(c/n)^{3k}(1 - c/n)^{\binom{k}{2} - 3k} = (1 - o(1))(c/n)^{3k}.$$

The expected number of edges in $G(n - 3k, c/n)$ is at most $n^2 \cdot (c/n) = cn$, and so Markov's inequality shows that it contains more than $n \log n$ edges with probability $o(1)$. Therefore the second condition holds with probability $e^{-c^3/6} - o(1)$.

Note that the first two conditions are independent, since the first one depends only on the edges among the $3k$ triangle vertices, and the second one depends only on the edges among the remaining $n - 3k$ vertices. The third condition will only depend on the remaining edges.

Now suppose that the first two conditions hold. There are two ways in which the graph can contain a triangle other than the ones in τ. First, there could be a triangle sharing an edge with one of the triangles in τ. There are $3k(n - 3k)$ such potential triangles, and each of them belongs to G with probability $(c/n)^2$, and so this case happens with probability $O(1/n)$.

Second, there could be a triangle sharing a vertex with one of the triangles in τ. Such a triangle should contain one of the $n \log n$ edges involving the remaining $n - 3k$ vertices. There are $3kn \log n$ such potential triangles, and each of them belongs to G with probability $(c/n)^2$, so this case happens with probability $O(\log n/n)$. In total, assuming the first two conditions, the third one is met with probability $1 - o(1)$.

Summarizing,

$$\Pr[E_\tau] = (1 - o(1))(c/n)^{3k} \cdot (e^{-c^3/6} - o(1)) \cdot (1 - o(1)) \sim e^{-c^3/6}(c/n)^{3k}.$$

There are $\frac{1}{k!}\binom{n}{3}\binom{n-3}{3} \cdots \binom{n-3(k-1)}{3} \sim \frac{1}{k!}(n^3/6)^k$ possible τ, and so, since the events E_τ are disjoint,

$$\Pr[X = k] \geq \sum_\tau \Pr[E_\tau] \sim \frac{(n^3/6)^k}{k!} \cdot e^{-c^3/6}(c/n)^{3k} = e^{-c^3/6}\frac{(c^3/6)^k}{k!}.$$

Combining this with the upper bound, we conclude

$$\Pr[X = k] \longrightarrow e^{-c^3/6}\frac{(c^3/6)^k}{k!}.$$

4 Bonferroni inequalities

Finally, let us prove the Bonferroni inequalities. Suppose that $E_1, \ldots, E_m$ are events. We want to bound the probability that none of the events happen by an expression of the form

$$p_k := 1 - \sum_{i=1}^{m} \Pr[E_i] + \sum_{1 \leq i < j \leq m} \Pr[E_i \wedge E_j] - \cdots \pm \sum_{1 \leq i_1 < \cdots < i_k \leq m} \Pr[E_1 \wedge \cdots \wedge E_k].$$

Consider any point x in the sample space, and suppose that it belongs to t of the events. Its contribution to the sum above is $\Pr[x]$ times the polynomial

$$P_k(t) = 1 - t + \binom{t}{2} - \cdots \pm \binom{t}{k}.$$

Note that $P_k(0) = 1$ and for $1 \leq t \leq k$,

$$P_k(t) = 1 - t + \binom{t}{2} - \cdots \pm \binom{t}{t} = 0.$$

Therefore the roots of P_k are $t = 1, \dots, t = k$. The sign of P_k is constant for all $t > k$. Since $P_k(t) \sim (-1)^k t^k / k!$, it follows that when $t > k$, $P_k(t)$ is positive if k is even and negative if k is odd.

Consequently, when k is even, the contribution of x to $q := \Pr[\neg E_1 \wedge \cdots \wedge \neg E_k]$ is always bounded above by its contribution to p_k, and when k is odd, it is bounded below. Therefore $q \leq p_k$ for even k and $q \geq p_k$ for odd k.

5 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.

A different large deviation question asks for the probability that $G(n,p)$ contains a triangle when $p = o(1/n)$. I'm not sure what is the answer to this question!