Random Graphs — Week 1

Yuval Filmus

October 30, 2019

1 Erdős–Rényi random graphs

For most of the class, we will be concerned with the $G(n,p)$ model of random graphs, also known as Erdős–Rényi random graphs. In this model, there is a fixed vertex set $[n] = \{1,\ldots,n\}$, and each of the potential $\binom{n}{2}$ edges is included with probability p independently. When $p = 1/2$, this is the same as picking a uniformly random graph on the vertex set $[n]$.

Erdős and Rényi actually considered the related $G(n,m)$ model, in which we choose a random graph on the vertex set $[n]$ with exactly m edges. The models $G(n,p)$ and $G(n,m)$ have similar behavior, where $m = p\binom{n}{2}$ is the expected number of edges in $G(n,p)$.

2 Monotone properties

Let $c(p)$ be the probability that a random $G(n,p)$ graph contains a triangle. This is an example of a *monotone property*: if a graph contains a triangle, then it will still contain a triangle if we add more edges to it.

We claim that $c(p)$ is a continuous, strictly increasing function of p. To prove that $c(p)$ is increasing, we use the technique of *coupling*.

Lemma 1. *If $p \leq q$ then $c(p) \leq c(q)$.*

Proof 1. Sample G_1 from the distribution $G(n,p)$. For each edge absent from G_1, add it with probability $r := \frac{q-p}{1-p}$ (independently), and call the resulting graph G_2. The probability that a specific edge belongs to G_2 is $p + (1-p)r = q$, and so G_2 has the distribution $G(n,q)$.

The resulting distribution (G_1, G_2) on pairs of graphs is a *coupling* of the distributions $G(n,p)$ and $G(n,q)$. This means that its marginals have these distributions. Furthermore, the construction guarantees that $G_1 \subseteq G_2$ always.

We can now prove the lemma:

$$c(p) = \Pr[G_1 \text{ contains a triangle}] \leq \Pr[G_2 \text{ contains a triangle}] = c(q). \qquad \square$$

Proof 2. Here is another way to construct the same coupling. Consider the complete graph K_n. For each edge, choose a weight according to the uniform distribution on

$[0,1]$. Let G_p be the graph consisting of all edges whose weight is at most p, and define G_q analogously. Then (G_p, G_q) is a coupling of $G(n,p), G(n,q)$ such that $G_p \subseteq G_q$ always. $\square$

The fact that $c(p)$ is continuous and strictly increasing follows from the fact that $c(p)$ is a polynomial.

Lemma 2. *The function $c(p)$ is a polynomial of degree at most $\binom{n}{2}$, and so is continuous and strictly increasing.*

Proof. Let $\mathcal{G}$ be the set of all graphs on $[n]$ containing a triangle. For each $G \in \mathcal{G}$, the probability that G is sampled from $G(n,p)$ is $p^{|G|}(1-p)^{\binom{n}{2}-|G|}$, where $|G|$ is the number of edges in G; note that this is a polynomial in p. Therefore $c(p) = \sum_{G \in \mathcal{G}} p^{|G|}(1-p)^{\binom{n}{2}-|G|}$ is a polynomial.

Since $c(p)$ is a polynomial, it is continuous. Since it is increasing and nonconstant (as $c(0) = 0$ and $c(1) = 1$), it must be strictly increasing. $\square$

We can define the *critical probability* of containing a triangle to be the probability p^* such that $c(p^*) = 1/2$ (the value $1/2$ here is arbitrary; any constant will do). How can we determine p^*?

3 Appearance of triangles

Let us denote by X the number of triangles in a $G(n,p)$ graph. Using linearity of expectation, it is not difficult to calculate the expected number of triangles:

$$\mathbb{E}[X] = \binom{n}{3} p^3 \sim \frac{(np)^3}{6}.$$

Here $\binom{n}{3}$ is the number of potential triangles, and p^3 is the probability that $G(n,p)$ contains a specific triangle.

This formula suggests that p^* should be of order of magnitude $1/n$, since when $p \ll 1/n$, the expected number of triangles is very small, and when $p \gg 1/n$, it is very large. In order to show that this is indeed the case, we will need to use the *first moment method* as well as the *second moment method*.

3.1 First moment method

Markov's inequality implies that if $\mathbb{E}[X]$ is small, then it is unlikely that $X \neq 0$, since

$$\Pr[X > 0] = \Pr[X \geq 1] \leq \mathbb{E}[X].$$

(In this particular case, the same bound also follows from the union bound.)

Using this, we immediately get that when $p \ll 1/n$, a $G(n,p)$ graph typically doesn't contain any triangles.

Theorem 1. *If $p = o(1/n)$ then with high probability, $G(n,p)$ contains no triangles. (This means that $\Pr_{G \sim G(n,p)}[G$ contains a triangle$] = o(1)$.)*

3.2 Second moment method

Markov's inequality only allows us to bound the probability that X is much larger than its expectation. To bound the probability that X is much smaller than its expectation, we use Chebyshev's inequality:

$$\Pr[X = 0] \leq \Pr[|X - \mathbb{E}[X]| \geq \mathbb{E}[X]] \leq \frac{\mathbb{V}[X]}{\mathbb{E}[X]^2}.$$

Calculating $\mathbb{V}[X]$ is straightforward, if a bit tiring. Since $\mathbb{V}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$ and we already know $\mathbb{E}[X]$, it remains to compute $\mathbb{E}[X^2]$. We can write X as the sum of indicator variables I_S, where S goes over all subsets of $[n]$ of size 3, and I_S indicates that the graph contains the triangle formed by S. Expanding $\mathbb{E}[X^2]$, we get

$$\mathbb{E}[X^2] = \sum_{S,T \in \binom{[n]}{3}} \mathbb{E}[I_S I_T] = \sum_{S,T \in \binom{[n]}{3}} \Pr[G \text{ contains } \triangle_S, \triangle_T].$$

The probability that the graph contains both triangles supported on S and on T depends on the total number of edges in both triangles:

- If $S = T$ then the two triangles share all edges, and so the probability is p^3. There are $\binom{n}{3}$ such pairs.
- If $|S \cap T| = 2$ then the two triangles share one edge, and so the probability is p^5. There are $12\binom{n}{4}$ such pairs (together, the two triangles form a K_4 with one missing edge; we need to choose which edge is missing, and which of the two degree 2 vertices belongs to S).
- Otherwise, the two triangles might share a vertex, but they share no edge, and so the probability is p^6.

In total, we get

$$\mathbb{E}[X^2] = \binom{n}{3}p^3 + 12\binom{n}{4}p^5 + \left(\binom{n}{3}^2 - \binom{n}{3} - 12\binom{n}{4}\right)p^6.$$

Therefore

$$\mathbb{V}[X] = \binom{n}{3}(p^3 - p^6) + 12\binom{n}{4}(p^5 - p^6) = O(n^3p^3 + n^4p^5).$$

Since $\mathbb{E}[X] = \Theta(n^3p^3)$, the quantity in Chebyshev's inequality is

$$\frac{\mathbb{V}[X]}{\mathbb{E}[X]^2} = \frac{O(n^3p^3 + n^4p^5)}{\Theta(n^6p^6)} = O\left(\frac{1}{n^3p^3} + \frac{1}{n^2p}\right).$$

Theorem 2. *If $p = \omega(1/n)$, then with high probability $G(n,p)$ contains a triangle.*

Proof. If $p = \omega(1/n)$ then $pn = \omega(1)$, and so

$$\frac{\mathbb{V}[X]}{\mathbb{E}[X]^2} = O\left(\frac{1}{(np)^3} + \frac{1}{n(np)}\right) = \frac{1}{\omega(1)} + \frac{1}{n\omega(1)} = o(1).$$

Therefore, by Chebyshev's inequality we get $\Pr[X = 0] = o(1)$. □

4 Where do we go from here?

Our investigation has so far led to interesting results, but leaves many questions open, such as the following:

1. The theorems state what happens when $p \ll 1/n$ and when $p \gg 1/n$, but leave a gap in the middle. What happens when $p = c/n$, for constant c? In particular, what is the probability that the graph contains a triangle? More generally, what is the distribution of the number of triangles?
2. What is the approximate distribution of the number of triangles when $p = \omega(1/n)$? What is the probability that the number of triangles deviates significantly from the average? What does the graph look like in the latter case?
3. What is the probability that the graph does contain a triangle when $p = o(1/n)$?
4. What if we want a guarantee which is better than just "with high probability"?
5. What happens if we replace triangle with a different graph?
6. Do the triangles "bunch up" in specific parts of the graph?

We will see answers to some of these questions in future weeks.