CS 4540/CS 8803: Algorithmic Theory of Intelligence	Fall 2023

Lecture 7: Random Graph

Instructor: Santosh Vempala *Lecture date: 9/27*

Scribed by: Mirabel Reid, Xinyuan Cao

1 Erdos-Renyi Random Graph

We consider an undirected random graph $G_{n,p}$ of n vertices. For vertices i, j, (i, j) is an edge in G with probability p. That is, $\forall i, j, \mathbb{P}((i, j) \in E) = p$. We will discover the properties of the random graph.

Lemma 1. *The probability that there is no edge in G is $(1-p)^{\binom{n}{2}}$.*

If $p = 1/n$, by Lemma 1,

$$\mathbb{P}(\text{graph has no edge}) = (1 - 1/n)^{\binom{n}{2}} \simeq e^{-(n-1)/2}$$

This is exponentially small.

Lemma 2. *The probability that there exists isolated vertices is upper bounded by $n(1-p)^{n-1}$.*

Corollary 3. *For $p > \frac{\ln ne}{n}$, the probability that there exists isolated vertex is smaller than $1/e$.*

Proof. We check as follows.

$$\left(1 - \frac{\ln ne}{n}\right)^{n-1} \leq e^{-\ln ne} = \frac{1}{ne}$$

□

Theorem 4 (Sharp Threshold)**.** *Let $p = \frac{\ln n + c}{n}$. Then the probability that $G_{n,p}$ is connected satisfies*

$$\mathbb{P}(G_{n,p} \textit{ is connected}) = \begin{cases} 1 & \textit{for } c \xrightarrow[n\to\infty]{} \infty \\ e^{-e^{-c}} & \textit{for } c \textit{ is a constant} \\ 0 & \textit{for } c \xrightarrow[n\to\infty]{} -\infty \end{cases}$$

Before we prove this theorem, let's examine what it means to have a sharp threshold. We will define a *graph property* as a measurable property which is either true or false for any fixed graph. Therefore, if we sample a graph G from a random graph model (such as $G_{n,p}$), we can talk about the *probability* that G has property P. An interesting family of graph properties are monotone properties.

Definition 5. *A graph property P is monotone if, once the property is true for G, it cannot be made false by adding edges.*

Connectivity is an example of a monotone increasing property. Once G is connected, it cannot suddenly become unconnected if you add more edges to the graph. Another example is the existence of a k-clique, or the existence of a cycle of length k.

We have the following theorem about monotone properties in G (not proven in class):

Theorem 6. *Any monotone graph property has a sharp threshold in $G_{n,p}$.*

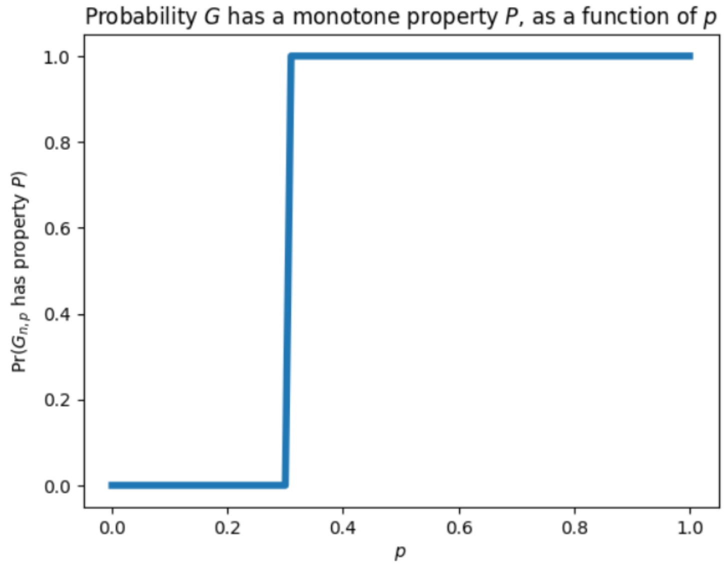

Figure 1: Example of a sharp threshold for a monotone graph property.

Fix a monotone graph property P. This theorem implies that there exists a $p^* \in [0, 1]$ such that for any small constant c,

$$Pr(G \sim G_{n,p^*+c} \text{ has property } P) = 1 - o(1)$$

$$Pr(G \sim G_{n,p^*-c} \text{ has property } P) = o(1)$$

That is, the probability either goes to 0 or to 1 as $n \to \infty$. If we were to graph the probability that $G \sim G_{n,p}$ has property P, the graph would look like Figure 1

Theorem 4 states that the threshold for connectivity is $p^* = \frac{\ln n}{n}$. Let's prove it now.

Proof of Theorem 4. Since proving connectivity directly is complicated, let's bound the probability by another property: the existence of an isolated vertex. We will show the following inequality:

$$Pr(G \text{ is connected}) \leq Pr(G \text{ has an isolated vertex}) \leq (1 + o(1))Pr(G \text{ is connected})$$

The first inequality is easy. If G is connected, then by definition it does not have an isolated vertex. Therefore, being connected is less likely.

The second inequality is more involved. Let's bound the probability that the graph is connected by summing over all possible connected components of size k.

$$\begin{aligned} Pr(G \text{ is connected}) &\leq \sum_{k=1}^{n} Pr(\exists \text{ a } k\text{-component in } G) \\ &= Pr(G \text{ has an isolated vertex}) + \sum_{k=2}^{n} Pr(\exists \text{ a } k\text{-component in } G) \\ &\text{Since a 1-component is an isolated vertex.} \end{aligned}$$

For a fixed k, there are $\binom{n}{k}$ ways to choose the component. For it to be connected, there must be at least one tree within the edges. To get an upper bound on the probability, we can sum over all

k^{k-2} trees to get the following bound:

$$
\begin{aligned}
Pr(\exists \text{ a } k\text{-component in } G) &\leq \binom{n}{k} k^{k-2} p^{k-1} (1-p)^{k(n-k)} \\
&\leq \left(\frac{ne}{k}\right)^k k^{k-2} p^{k-1} e^{-pk(n-k)} \\
&\leq ne^k k^{-2} (\ln n + c)^{k-1} e^{-\frac{\ln n + c}{n} k(n-k)} \quad \text{substituting } p = \frac{\ln n + c}{n} \text{ and simplifying.}
\end{aligned}
$$

Next, we will split into two cases.
Case 1: $k \geq 10$

$$
\begin{aligned}
Pr(\exists \text{ a } k\text{-component in } G) &\leq ne^k k^{-2} (\ln n + c)^{k-1} e^{-\frac{1}{2}(\ln n + c)k} \\
&= ne^k k^{-2} (\ln n + c)^{k-1} n^{-k/2} e^{-ck/2} \\
&\leq \frac{1}{n^2} \text{ for all } k \geq 10
\end{aligned}
$$

Case 2: $k < 10$, so $k(n-k) \geq 2(n-2)$

$$
\begin{aligned}
Pr(\exists \text{ a } k\text{-component in } G) &\leq ne^k k^{-2} (\ln n + c)^{k-1} e^{-\frac{2(n-2)}{n}(\ln n + c)} \\
&\leq ne^{10} 2^{-2} (\ln n + c)^9 n^{-2+\frac{4}{n}} e^{-2c} \\
&= \frac{1}{n^{1-o(1)}}
\end{aligned}
$$

Summing over all k, we get:

$$
\sum_{k=2}^{n} Pr(\exists \text{ a } k\text{-component in } G) = o(1)
$$

Thus,

$$
Pr(G \text{ is connected}) = Pr(G \text{ has an isolated vertex}) + o(1)
$$

Finally, using the result from Corollary 3, we get the theorem statement. $\square$