1 Threshold Phenomena in Random Graphs

We consider a model of Random Graphs by Erdős and Rényi [ER60]. To generate a random graph with \( n \) vertices, for every pair of vertices \( \{i, j\} \), we put an edge independently with probability \( p \). This model is denoted by \( G_{n,p} \).

Let \( G \) be a random \( G_{n,p} \) graph and let \( H \) be any fixed graph (on some constant number of vertices independent of \( n \)). We will be interested in understanding the probability that \( G \) contains a copy of \( H \). We start by computing this when \( H \) is \( K_4 \), the clique on 4 vertices.

**Definition 1.1** We define \( k \)-clique to be a fully connected graph with \( k \) vertices.

![4-Clique](image)

**Figure 1: 4-Clique**

As a convention, we will count a permutation of a copy of \( K_4 \) as the same copy. We define the random variable

\[
Z = \text{number of copies of } K_4 \text{ in } G = \sum_C X_C,
\]

where \( C \) ranges over all subsets of \( V \) of size 4 and the random variable \( X_C \) is defined as

\[
X_C = \begin{cases} 
1 & \text{if all pair of vertices in the set } C \text{ have an edge in between them} \\
0 & \text{otherwise}
\end{cases}.
\]
We have $\mathbb{E}[X_C] = p^6$, since the probability of connecting all 4 vertices (using 6 edges) in the 4-tuple is $p^6$. So we have the expectation of $Z$:

$$\mathbb{E}[Z] = \sum_C \mathbb{E}[X_C] = \binom{n}{4} \cdot p^6$$

We observe that

$$\mathbb{E}[Z] \to 0 \text{ when } p \ll n^{-2/3} \quad \text{and} \quad \mathbb{E}[Z] \to \infty \text{ when } p \gg n^{-2/3}.$$ 

Here, by $p \ll n^{-2/3}$, we mean that $\lim_{n \to \infty} \left(\frac{p}{n} - \frac{2}{3}\right) = 0$ and $p \gg n^{-2/3}$ is defined similarly. We will prove that there is in fact a threshold phenomenon in the probability that $G$ contains a copy of $K_4$.

### Theorem 1.2

Let $G$ be generated randomly according to the model $G_{n,p}$ graph. We have that:

- If $p \ll n^{-2/3}$, then $\mathbb{P}[G \text{ contains a copy of } K_4] \to 0$ as $n \to \infty$.
- If $p \gg n^{-2/3}$, then $\mathbb{P}[G \text{ contains a copy of } K_4] \to 1$ as $n \to \infty$.

### Proof:

As above, we define the random variable $Z$,

$$Z = \text{number of copies of } K_4 \text{ in } G = \sum_C X_C.$$

The case when $p \ll n^{-2/3}$ can be easily handled by Markov’s inequality. We get that,

$$\mathbb{P}[Z > 0] = \mathbb{P}[Z \geq 1] \leq \frac{\mathbb{E}[Z]}{1}.$$ 

Since $\mathbb{E}[Z] \to 0$ as $n \to \infty$ when $p \ll n^{-2/3}$, we get that $\mathbb{P}[G \text{ contains a copy of } K_4] \to 0$. When $p \gg n^{-2/3}$, we want to show that $\mathbb{P}[Z > 0] \to 1$, i.e., $\mathbb{P}[Z = 0] \to 0$. We use Chebyshev’s inequality to prove this. We first compute the variance of $Z$.

$$\text{Var}[Z] = \text{Var} \left( \sum_C X_C \right) = \sum_C \text{Var}[X_C] + \sum_{C \neq D} \text{Cov}[X_C, X_D]$$

Since $\mathbb{E}[X_C] = p^6$, we have $\text{Var}[X_C] = p^6 - p^{12}$. Also, for two distinct sets $C$ and $D$, we consider four different cases depending on the number of vertices they share.

- **Case 1:** $|C \cap D| = 0$. Since no vertex is shared, $X_C$ and $X_D$ are independent and hence $\text{Cov}[X_C, X_D] = 0$. 

2
- Case 2: $|C \cap D| = 1$. Since the variables $X_C$ and $X_D$ depend on pairs of vertices in the sets $C$ and $D$, and the two sets do not share any pair, we still have $\text{Cov} [X_C, X_D] = 0$.

- Case 3: $|C \cap D| = 2$. Since $C$ and $D$ share a pair of vertices, there are 11 pairs which must all have edges between them in $G$, for both $X_C$ and $X_D$ to be 1. Thus, we have $\mathbb{E} [X_C \cdot X_D] = p^{11}$ and

$$\text{Cov} [X_C, X_D] = \mathbb{E} [X_C X_D] - \mathbb{E} [X_C] \cdot \mathbb{E} [X_D] = p^{11} - p^{12}.$$  

- Case 4: $|C \cap D| = 3$. In this case $C$ and $D$ share 3 pairs and thus there are 9 distinct pairs of vertices which must all have edges between them for both $X_C$ and $X_D$ to be 1. Thus,

$$\text{Cov} [X_C, X_D] = \mathbb{E} [X_C X_D] - \mathbb{E} [X_C] \cdot \mathbb{E} [X_D] = p^9 - p^{12}.$$  

Also, there are $\binom{n}{6} \cdot \binom{5}{4}$ pairs $C$ and $D$ such that $|C \cap D| = 2$, and $\binom{n}{3} \cdot \binom{4}{4}$ pairs such that $|C \cap D| = 3$. Combining the above cases we have,

$$\text{Var} [Z] = \sum_C \text{Var} [X_C] + \sum_{C \neq D} \text{Cov} [X_C, X_D]$$

$$= \left( \frac{n}{4} \right) \cdot p^6 (1 - p^6) + \left( \frac{n}{6} \right) \cdot \left( \frac{6}{4} \right) \cdot (p^{11} - p^{12}) + \left( \frac{n}{5} \right) \cdot \left( \frac{4}{4} \right) \cdot (p^{9} - p^{12})$$

$$= O(n^4 p^6) + O(n^6 p^{11}) + O(n^5 p^9).$$

Applying Chebyshev’s inequality gives

$$\mathbb{P} [Z = 0] \leq \mathbb{P} \left[ |Z - \mathbb{E} [Z]| \geq \mathbb{E} [Z] \right] \leq \frac{\text{Var} [Z]}{\left( \mathbb{E} [Z] \right)^2}$$

$$= \frac{1}{\left( \frac{n}{4} \right)^2} \cdot p^{12} \cdot \left( O(n^4 p^6) + O(n^6 p^{11}) + O(n^5 p^9) \right)$$

$$= O \left( \frac{1}{n^4 p^6} \right) + O \left( \frac{1}{n^2 p} \right) + O \left( \frac{1}{n^3 p^3} \right).$$

For $p \gg n^{-2/3}$, all the terms on the right tend to 0 as $n \to \infty$. Hence, $\mathbb{P} [Z = 0] \to 0$ as $n \to \infty$.  

The above analysis can be extended to any graph $H$ of a fixed size. Let $Z_H$ be the number of copies of $H$ in a random graph $G$ generated according to $G_n,p$. Using the previous analysis, we have $\mathbb{E} [Z_H] = \Theta \left( n^{\left| V(H) \right|} \cdot p^{\left| E(H) \right|} \right)$. Hence, $\mathbb{E} [Z] \to 0$ when $p \ll n^{-\left| V(H) \right|/\left| E(H) \right|}$ and $\mathbb{E} [Z] \to \infty$ when $p \gg n^{-\left| V(H) \right|/\left| E(H) \right|}$. Thus, it might be tempting to conclude that $p = n^{-\left| V(H) \right|/\left| E(H) \right|}$ is the threshold probability for finding a copy of $H$. However, con-
Consider the graph in Figure 2. For this graph, we have $|V(H)|/|E(H)| = 5/7$. But for $p$ such that $p \gg n^{-5/7}$ and $p \ll n^{-2/3}$, a random $G$ is extremely unlikely to contain a copy of $K_4$ and thus also extremely unlikely to contain a copy of $H$. For an arbitrary graph $H$, it was shown by Bollobás [Bol81] that the threshold probability is $n^{-\lambda}$, where

\[ \lambda = \min_{H' \subseteq H} \frac{|V(H')|}{|E(H')|}. \]

2 Chernoff/Hoeffding Bounds

We now derive sharper concentration bounds for sums of independent random variables. We start by considering $n$ independent Boolean random variables $X_1, \ldots, X_n$, where $X_i$ takes value 1 with probability $p_i$ and 0 otherwise. Let $Z = \sum_{i=1}^{n} X_i$. We set $\mu = \mathbb{E}[Z] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \sum_{i=1}^{n} \mu_i$. We will try to derive a bound on the probability $\mathbb{P}[Z \geq t]$ for $t = (1 + \delta)\mu$. Using the fact that the function $e^x$ is strictly increasing, we get that for $\lambda > 0$

\[ \mathbb{P}[Z \geq (1 + \delta)\mu] = \mathbb{P}[e^{\lambda Z} \geq e^{\lambda(1+\delta)\mu}] \leq e^{\lambda(1+\delta)\mu}. \]

We now have:

\[ \mathbb{E}[e^{\lambda Z}] = \mathbb{E}[e^{\lambda(X_1+\ldots+X_n)}] = \mathbb{E}\left[\prod_{i=1}^{n} e^{\lambda X_i}\right] \overset{\text{independence}}{=} \prod_{i=1}^{n} \mathbb{E}[e^{\lambda X_i}] = \prod_{i=1}^{n} [\mu_i e^\lambda + (1 - \mu_i)] = \prod_{i=1}^{n} [1 + \mu_i(e^\lambda - 1)]. \]

At this point, we utilize the simple but very useful inequality:

\[ \forall x \in \mathbb{R}, \quad 1 + x \leq e^x. \]
Since all the quantities in the previous calculation are non-negative, we can plug the above inequality in the previous calculation and we get:

$$E[e^{\lambda Z}] \leq \prod_{i=1}^{n} \exp\left( (e^\lambda - 1)\mu_i \right) = \exp\left( (e^\lambda - 1)\mu \right)$$

Thus, we get

$$P[Z \geq (1 + \delta)\mu] \leq \exp\left( (e^\lambda - 1)\mu - \lambda(1 + \delta)\mu \right).$$

We now want to minimize the right hand-side of the above inequality, with respect to $\lambda$. Setting the derivative of the exponent to zero, we get

$$e^\lambda \mu - (1 + \delta)\mu = 0 \Rightarrow \lambda = \ln(1 + \delta).$$

Using this value for $\lambda$, we get

$$P[Z \geq (1 + \delta)\mu] \leq \frac{\exp\left( (e^\lambda - 1)\mu \right)}{\exp\left( \lambda(1 + \delta)\mu \right)} = \frac{e^{\delta \mu}}{(1 + \delta)^{1+\delta}} = \left( \frac{e^\delta}{(1 + \delta)^{1+\delta}} \right)^\mu.$$ 

**Exercise 2.1** Prove similarly that

$$P[Z \leq (1 - \delta)\mu] \leq \left( \frac{e^{-\delta}}{(1 - \delta)^{1-\delta}} \right)^\mu.$$ 

(Note that $P[Z \leq (1 - \delta)\mu] = P[e^{-\lambda Z} \geq e^{-\lambda(1-\delta)\mu}]$.) When $\delta \in (0, 1)$, the bounds above expressions can be simplified further. It is easy to check that

$$\left( \frac{e^\delta}{(1 + \delta)^{1+\delta}} \right)^\mu \leq e^{-\delta^2 \mu/3}, \quad 0 < \delta < 1.$$ 

So we get:

$$P[Z \geq (1 + \delta)\mu] \leq e^{-\delta^2 \mu/3}, \quad \text{for } 0 < \delta < 1.$$ 

Similarly:

$$P[Z \leq (1 - \delta)\mu] \leq e^{-\delta^2 \mu/3}, \quad \text{for } 0 < \delta < 1.$$ 

Combining the two we get

$$P[|Z - \mu| \geq \delta \mu] \leq 2 \cdot e^{-\delta^2 \mu/3}, \quad \text{for } 0 < \delta < 1.$$ 

The above is only one of the proofs of the Chernoff-Hoeffding bound. A delightful paper by Mulzer [Mull18] gives several other proofs with different applications.
2.1 Coin tosses once more

We will now compare the above bound with what we can get from Chebyshev’s inequality. Let’s assume that $X_1, \ldots, X_n$ are independent coin tosses, with $\mathbb{P}[X_i = 1] = \frac{1}{2}$. We want to get a bound on the value of $Z = \sum_{i=1}^{n} X_i$. Using Chebyshev’s inequality, we get that

$$
\mathbb{P}[|Z - \mu| \geq \delta \mu] \leq \frac{\text{Var}[Z]}{\delta^2 \mu^2}.
$$

And since in this particular case we have that $\text{Var}[Z] = n/4$ and $\mu = n/2$, we get that

$$
\mathbb{P}[|Z - \mu| \geq \delta \mu] \leq \frac{1}{\delta^2 n}.
$$

The above bound is only inversely polynomial in $n$, while the Chernoff-Hoeffding bound gives

$$
\mathbb{P}[|Z - \mu| \geq \delta \mu] \leq 2 \cdot \exp\left(-\frac{\delta^2 n}{6}\right),
$$

which is exponentially small in $n$. This fact will prove very useful when taking a union bound over a large collection of events, each of which may be bounded using a Chernoff-Hoeffding bound.

References

