

## Lecture Notes on Random Graphs

Many computational problems on graphs are NP-hard, such as Hamiltonian Cycle, Max Clique, and Max Independent Set. Are these hard on average? To study this and related questions, it helps to study random graphs.

There are two common models for random graphs. In  $G_{n,p}$ , a graph  $G$  on  $n$  nodes is chosen so that each of the possible  $\binom{n}{2}$  edges is added independently with probability  $p$ . In  $G_{n,m}$ , a graph  $G$  is chosen uniformly from all graphs on  $n$  nodes and  $m$  edges.

$G_{n,p}$  and  $G_{n,m}$  are closely related when the expected number of edges  $\binom{n}{2} \cdot p = m$ . See Section 5.6.1 of the Mitzenmacher-Upfal text for details. However,  $G_{n,p}$  is generally simpler to analyze, because we have independence. Hence we focus on this model.

### 1 Evolution of Random Graphs

We now think of growing a random graph by adding random edges one at a time. How does the random graph evolve? We study this question in the simpler  $G_{n,p}$  model.

**Theorem 1.** *For  $p = o(1/n)$  and large enough  $n$ , with high probability a graph  $G$  drawn from  $G_{n,p}$  is a forest.*

*Proof.*

$$\begin{aligned} \Pr[G \text{ is not a forest}] &= \Pr[G \text{ has a cycle}] \leq \mathbb{E}[\text{number of cycles in } G] \\ &= \sum_{k=3}^n \frac{n(n-1) \cdots (n-k+1)}{2k} p^k \\ &< \sum_{k=3}^n \frac{(pn)^k}{2k} \\ &< 2 \cdot \frac{(pn)^3}{6} = o(1). \end{aligned}$$

The last inequality uses that  $pn < 1/2$  for large enough  $n$ . □

We state the following without proof. Let  $p = c/n$ . Then with high probability the size of the largest component is:

- $\Theta(\log n)$  for  $c < 1$ ,
- $\Theta(n^{2/3})$  for  $c = 1$ , and

- $\Theta(n)$  for  $c > 1$ .

Thus, there is a sharp threshold around  $p = 1/n$ .

By the coupon collector problem, with high probability there are no isolated nodes in  $G_{n,m}$  when  $m \approx (n \ln n)/2$ . Therefore, it occurs in  $G_{n,p}$  when  $p \approx (\ln n)/n$ . When  $p$  is slightly larger, not only are there no isolated nodes, but with high probability the graph is connected.

## 2 Size of Largest Clique

With high probability, the size of a largest clique in  $G_{n,1/2}$  is  $(2 + o(1)) \log_2 n$ . To show this, we use a well-known upper bound on the size of a binomial coefficient.

**Fact:**  $\binom{n}{k} \leq \left(\frac{ne}{k}\right)^k$ .

Using this, we prove an upper bound on the clique size as follows.

$$\begin{aligned} \Pr[\exists k\text{-clique}] &\leq \mathbb{E}[\text{number of } k\text{-cliques}] = \binom{n}{k} 2^{-\binom{k}{2}} \\ &\leq \left(\frac{ne}{k}\right)^k \cdot 2^{-k(k-1)/2} \\ &= \left(\frac{ne}{k} \cdot 2^{(1-k)/2}\right)^k. \end{aligned}$$

Setting  $k = 2 \log_2 n$ , the upper bound becomes  $(\sqrt{2}e/k)^k$ , which goes to 0 as  $n \rightarrow \infty$ .

## 3 Greedy Algorithm for Finding a Large Clique

Max Clique is extremely difficult to approximate in the worst case. For any  $\epsilon > 0$ , it is NP-hard to distinguish graphs that contain a clique on  $n^{1-\epsilon}$  nodes from graphs with all cliques smaller than  $n^\epsilon$ .

We study a simple greedy algorithm to find a large clique in a random graph from  $G_{n,1/2}$ .

$S \leftarrow \emptyset$

For  $j = 1, 2, \dots, n$

    If  $S \cup \{j\}$  is a clique, then  $S \leftarrow S \cup \{j\}$

Return  $S$

To analyze this, let the random variable  $X_i$  denote the number of vertices inspected in this algorithm while  $|S| = i$ . Since the probability of a new vertex being connected to all nodes of  $S$  is  $2^{-i}$ , we conclude that  $X_i$  is a geometric random variable with success probability  $2^{-i}$ . Hence  $\mathbb{E}[X_i] = 2^i$ .

Let the random variable  $X$  denote the number of vertices inspected until the algorithm finds a clique of size  $k$ . Then

$$X = X_0 + X_1 + \dots + X_{k-1}.$$

Therefore,

$$E[X] = 1 + 2 + 2^2 + \dots + 2^{k-1} = 2^k - 1.$$

Heuristically, this suggests that the algorithm should find a clique of size  $k$  where  $2^k \approx n$ , i.e.,  $k \approx \log_2 n$ . This is about half the size of a largest clique.

Formally, we can say the following. Let  $S$  denote the output of the algorithm. Then

$$\Pr[|S| < k] = \Pr[X > n] < \frac{E[X]}{n} < \frac{2^k}{n} \leq \frac{1}{2^c},$$

if we set  $k = \log_2 n - c$ .