6.207/14.15: Networks
Lecture 11: Giant Component, Generalized Random Graphs
Outline

- Emergence and size of a giant component in Erdös-Renyi graphs
- An application: contagion and diffusion
- Generalized random graph models
- Graphs with prescribed degrees — configuration model
- Emergence of a giant component in the configuration model

Reading:
Giant Component

- We have shown that when $p(n) \ll \frac{\log(n)}{n}$, the Erdös-Renyi graph is disconnected with high probability.
- In cases for which the network is not connected, the component structure is of interest.
- We have argued that in this regime the expected number of isolated nodes goes to infinity. This suggests that the Erdös-Renyi graph should have an arbitrarily large number of components.
- We will next argue that the threshold $p(n) = \frac{\lambda}{n}$ plays an important role in the component structure of the graph.
  - For $\lambda < 1$, all components of the graph are “small”.
  - For $\lambda > 1$, the graph has a (unique) giant component, i.e., a component that contains a constant fraction of the nodes.
Emergence of the Giant Component—1

- We will analyze the component structure in the vicinity of \( p(n) = \frac{\lambda}{n} \) using a branching process approximation.
- We assume \( p(n) = \frac{\lambda}{n} \).
- \( B(n, \frac{\lambda}{n}) \): binomial random variable with parameters \( n, \frac{\lambda}{n} \).
- Consider starting from node 1 and exploring the graph.

(a) Erdos-Renyi graph process.  
(b) Branching Process Approx.
Emergence of the Giant Component—2

- We first consider the case when $\lambda < 1$.
- Let $Z^G_k$ and $Z^B_k$ denote the number of individuals at stage $k$ for the graph process and the branching process approximation, respectively.
- In view of the “overcounting” feature of the branching process, we have $Z^G_k \leq Z^B_k$ for all $k$.
- From branching process analysis (see Lecture 3 notes), we have $\mathbb{E}[Z^B_k] = \lambda^k$,
  
  (since the expected number of children is given by $n \times \frac{\lambda}{n} = \lambda$).
- Let $S_1$ denote the number of nodes in the Erdös-Renyi graph connected to node 1, i.e., the size of the component which contains node 1.
- Then, we have
  
  $$\mathbb{E}[S_1] = \sum_k \mathbb{E}[Z^G_k] \leq \sum_k \mathbb{E}[Z^B_k] = \sum_k \lambda^k = \frac{1}{1 - \lambda}.$$
Emergence of the Giant Component—3

- The preceding result suggests that for $\lambda < 1$, the sizes of the components are "small".

**Theorem**

Let $p(n) = \frac{\lambda}{n}$ and assume that $\lambda < 1$. For all (sufficiently large) $a > 0$, we have

$$\mathbb{P}\left( \max_{1 \leq i \leq n} |S_i| \geq a \log(n) \right) \to 0 \quad \text{as} \quad n \to \infty.$$

*Here $|S_i|$ is the size of the component that contains node $i$."

- This result states that for $\lambda < 1$, all components are small [in particular they are of size $O(\log(n))]$.
- Proof is beyond the scope of this course.
○ We next consider the case when $\lambda > 1$.
○ We claim that $Z_k^G \approx Z_k^B$ when $\lambda^k \leq O(\sqrt{n})$.
○ The expected number of conflicts at stage $k + 1$ satisfies
  \[
  \mathbb{E}[\text{number of conflicts at stage } k + 1] \approx np^2 \mathbb{E}[Z_k^2] = n \frac{\lambda^2}{n^2} \mathbb{E}[Z_k^2].
  \]
  \[
  \mathbb{E}[Z_k^2] = \text{var}(Z_k) + \mathbb{E}[Z_k]^2 = \lambda^k + \lambda^{2k} \approx \lambda^{2k}.
  \]
○ Combining the preceding two relations, we see that the conflicts become non-negligible only after $\lambda^k \approx \sqrt{n}$.
Emergence of the Giant Component—5

- Hence, there exists some $c > 0$ such that $\Pr(\text{there exists a component with size } \geq c\sqrt{n} \text{ nodes}) \to 1$ as $n \to \infty$.

- Moreover, between any two components of size $\sqrt{n}$, the probability of having a link is given by

\[
\Pr(\text{there exists at least one link}) = 1 - (1 - \frac{\lambda}{n})^n \approx 1 - \exp(-\lambda),
\]

i.e., it is a positive constant independent of $n$.

- This argument can be used to see that components of size $\leq \sqrt{n}$ connect to each other, forming a connected component of size $qn$ for some $q > 0$, a giant component.
Size of the Giant Component

- Form an Erdös-Renyi graph with \( n - 1 \) nodes with link formation probability \( p(n) = \frac{\lambda}{n} \), \( \lambda > 1 \).
- Now add a last node, and connect this node to the rest of the graph with probability \( p(n) \).
- Let \( q \) be the fraction of nodes in the giant component of the \( n - 1 \) node network. We can assume that for large \( n \), \( q \) is also the fraction of nodes in the giant component of the \( n \)-node network.
- The probability that node \( n \) is not in the giant component is given by
  \[
  P(\text{node } n \text{ not in the giant component}) = 1 - q \equiv \rho.
  \]
- The probability that node \( n \) is not in the giant component is equal to the probability that none of its neighbors is in the giant component, yielding
  \[
  \rho = \sum_{k=0}^{n-1} p_k \rho^k \equiv \Phi(\rho).
  \]
- Like before, this equation has a fixed point \( \rho^* \in (0, 1) \).
An Application: Contagion and Diffusion

- Consider a society of $n$ individuals.
- A randomly chosen individual is infected with a contagious virus.
- Assume that the network of interactions in the society is described by an Erdös-Renyi graph with link probability $p$.
- Assume that any individual is immune with a probability $\pi$.
- We would like to find the expected size of the epidemic as a fraction of the whole society.
- The spread of disease can be modeled as:
  - Generate an Erdös-Renyi graph with $n$ nodes and link probability $p$.
  - Delete $\pi n$ of the nodes uniformly at random.
  - Identify the component that the initially infected individual lies in.
- We can equivalently examine a graph with $(1 - \pi)n$ nodes with link probability $p$. 
An Application: Contagion and Diffusion

- We consider 3 cases:
  - $p(1 - \pi)n < 1$:
    
    $$\mathbb{E}[\text{size of epidemic as a fraction of the society}] \leq \frac{\log((1 - \pi)n)}{n} \approx 0.$$
  
  - $1 < p(1 - \pi)n < \log((1 - \pi)n)$:
    
    $$\mathbb{E}[\text{size of epidemic as a fraction of the society}] = \frac{qq(1 - \pi)n + (1 - q)\log((1 - \pi)n))}{n} \approx q^2(1 - \pi),$$
    
    where $q$ denotes the fraction of nodes in the giant component of the graph with $(1 - \pi)n$ nodes, i.e., $q = 1 - e^{-q(1-\pi)np}$.
  
  - $p > \frac{\log((1-\pi)n)}{(1-\pi)n}$:
    
    $$\mathbb{E}[\text{size of epidemic as a fraction of the society}] = (1 - \pi).$$
We have seen that the Erdös-Renyi model has a Poisson degree distribution, which falls off very fast.

Our next goal is to generate random networks with a “given degree distribution”.

One of the most widely method used for this purpose is the configuration model developed by Bender and Canfield in 1978.

The configuration model is specified in terms of a degree sequence, i.e., for a network of \( n \) nodes, we have a desired degree sequence \((k_1, \ldots, k_n)\), which specifies the degree \( k_i \) of node \( i \), for \( i = 1, \ldots, n \).

- Given a degree distribution \( p_k \), we can generate the degree sequence for \( n \) nodes by sampling the degrees independently from the distribution \( p_k \), i.e., \( k_i \sim p_k \).
- A law of large numbers argument establishes that the frequency of degrees \( p_k^{(n)} \) converges to the degree distribution \( p_k \) as \( n \) goes to infinity.
Configuration Model—2

- Given the degree \( k_i \) for node \( i \) for all \( i = 1, \ldots, n \), we create a random network with these degrees as follows:

- We give each node \( i, k_i \) “stubs” sticking out of it, which are ends of edges-to-be (there are a total of \( \sum_i k_i = 2m \) stubs, where \( m \) is the number of edges).

- We choose two stubs uniformly at random and create an edge between the corresponding nodes.

- We choose another pair from the remaining \( 2m - 2 \) stubs, connect those and continue until all the stubs are used up.

- **Remarks:**
  - This process generates each possible matching of stubs with equal probability.
  - The sum of degrees needs to be even (or else an entry will be left out at the end).
  - It is possible to have self-edges and multiedges.
Distribution of the Degree of a Neighboring Node—1

- We will use a branching process approximation to study the giant component in the configuration model.
- For this we need to understand the distribution of the degree of a neighboring node, i.e., given some node $i$ with degree $d_i$, consider a neighbor $j$. What is the degree distribution of node $j$?

\[ k_1 \sim p_k \]
\[ k_2 \sim \tilde{p}_k \]
\[ \text{# of children} = k_2 - 1 \]

- **Naive intuition**: Same distribution as node $i$.
- **Example**: Consider a graph with 4 nodes and links \{1,2\}, \{2,3\}, \{3,4\}.
  - We have $p_1 = p_2 = 1/2$. Pick a link at random, then randomly pick an end of it, there is a 2/3 chance of finding a node with degree 2 and 1/3 chance of finding a node with degree 1.
  - Higher degree nodes are involved in a higher percentage of the links.
The degree of a node we reach by following a randomly chosen edge is not given by $p_k$.

In the configuration model, an edge emerging from a node has equal chance of terminating at any of the stubs.

Since there are $2m$ stubs in total, the probability of this edge ending at any particular node of degree $k$ is $k/2m$.

Since the total number of nodes with degree $k$ is given by $np_k$, the probability of the edge attaching to a node with degree $k$ is given by

$$\frac{k}{2m} np_k = \frac{kp_k}{\langle k \rangle},$$

where $\langle k \rangle$ is the expected degree in the network and the equality follows from the relation $2m = n\langle k \rangle$. 


Distribution of the Degree of a Neighboring Node—3

- Intuitively, there are $k$ edges that arrive at a node of degree $k$, we are $k$ times as likely to arrive at that node than another node that has degree 1.
- Thus, the degree distribution of the neighboring node $\tilde{p}_k$ is proportional to $kp_k$.

$$\tilde{p}_k = \frac{kp_k}{\sum_j jp_j} = \frac{kp_k}{\langle k \rangle}.$$
Emergence of a Giant Component in the Configuration Model—1

- We will use a branching process approximation to analyze the emergence of the giant component.
  - We ignore self loops (can be shown to have small probability) and conflicts (do not matter until the graph grows to a substantial size).

- Note that we have

\[
\mu = \mathbb{E}[\text{number of children}] = \mathbb{E}[k - 1] \\
= \sum_k k\tilde{p}_k - 1 \\
= \sum_k \frac{k^2p_k}{\langle k \rangle} - 1 \\
= \frac{\langle k^2 \rangle}{\langle k \rangle} - 1.
\]
Emergence of a Giant Component in the Configuration Model—2

- Using the branching process analysis, this yields the following threshold for the emergence of the giant component:
  
  **Subcritical:** $\mu < 1$, or equivalently
  
  \[
  \frac{\langle k^2 \rangle}{\langle k \rangle} < 2 \iff \langle k(k - 2) \rangle < 0.
  \]

  **Supercritical:** $\mu > 1$, or equivalently
  
  \[
  \langle k(k - 2) \rangle > 0.
  \]

- In the case of an Erdös-Renyi graph, we have $\langle k^2 \rangle = \langle k \rangle + \langle k \rangle^2$, and so the giant component emerges when
  
  \[
  \langle k \rangle^2 > \langle k \rangle \iff \langle k \rangle > 1.
  \]

- Since $\langle k \rangle = (n - 1)p$ in the Erdös-Renyi graph, this indeed yields the threshold function $t(n) = \frac{1}{n}$ for the emergence of the giant component.