6.207/14.15: Networks
Lecture 4: Erdös-Renyi Graphs and Phase Transitions

Daron Acemoglu and Asu Ozdaglar
MIT

September 21, 2009

Outline

- Phase transitions
- Connectivity threshold
- Emergence and size of a giant component
- An application: contagion and diffusion

Reading:

- Jackson, Sections 4.2.2-4.2.5, and 4.3.

Phase Transitions for Erdös-Renyi Model

- Erdös-Renyi model is completely specified by the link formation probability $p(n)$.
- For a given property A (e.g. connectivity), we define a threshold function $t(n)$ as a function that satisfies:

$$\mathbb{P}(\text{property } A) \to 0 \qquad \text{if} \quad \frac{p(n)}{t(n)} \to 0, \text{and}$$

$$\mathbb{P}(\text{property } A) \to 1 \qquad \text{if} \quad \frac{p(n)}{t(n)} \to \infty.$$

  - This definition makes sense for "monotone or increasing properties," i.e., properties such that if a given network satisfies it, any supernetwork (in the sense of set inclusion) satisfies it.
- When such a threshold function exists, we say that a phase transition occurs at that threshold.
- Exhibiting such phase transitions was one of the main contributions of the seminal work of Erdös and Renyi 1959.

Threshold Function for Connectivity

Theorem

(Erdös and Renyi 1961) A threshold function for the connectivity of the Erdös and Renyi model is $t(n) = \frac{\log(n)}{n}$.

- To prove this, it is sufficient to show that when $p(n) = \lambda(n)\frac{\log(n)}{n}$ with $\lambda(n) \to 0$, we have $\mathbb{P}(\text{connectivity}) \to 0$ (and the converse).
- However, we will show a stronger result: Let $p(n) = \lambda\frac{\log(n)}{n}$.

$$\text{If } \lambda < 1, \qquad \mathbb{P}(\text{connectivity}) \to 0, \tag{1}$$

$$\text{If } \lambda > 1, \qquad \mathbb{P}(\text{connectivity}) \to 1. \tag{2}$$

Proof:

- We first prove claim (1). To show disconnectedness, it is sufficient to show that the probability that there exists at least one isolated node goes to 1.

Proof (Continued)

- Let I_i be a Bernoulli random variable defined as

$$I_i = \begin{cases} 1 & \text{if node } i \text{ is isolated,} \\ 0 & \text{otherwise.} \end{cases}$$

- We can write the probability that an individual node is isolated as

$$q = \mathbb{P}(I_i = 1) = (1-p)^{n-1} \approx e^{-pn} = e^{-\lambda \log(n)} = n^{-\lambda}, \qquad (3)$$

  where we use $\lim_{n\to\infty} \left(1 - \frac{a}{n}\right)^n = e^{-a}$ to get the approximation.

- Let $X = \sum_{i=1}^{n} I_i$ denote the total number of isolated nodes. Then, we have

$$\mathbb{E}[X] = n \cdot n^{-\lambda}. \qquad (4)$$

- For $\lambda < 1$, we have $\mathbb{E}[X] \to \infty$. We want to show that this implies $\mathbb{P}(X = 0) \to 0$.
  - In general, this is not true.
  - Can we use a Poisson approximation (as in the example from last lecture)? No, since the random variables I_i here are dependent.
  - We show that the variance of X is of the same order as its mean.

Proof (Continued)

- We compute the variance of X, $\text{var}(X)$:

$$\begin{aligned}\text{var}(X) &= \sum_i \text{var}(I_i) + \sum_i \sum_{j \neq i} \text{cov}(I_i, I_j) \\ &= n\text{var}(I_1) + n(n-1)\text{cov}(I_1, I_2) \\ &= nq(1-q) + n(n-1)\Big(\mathbb{E}[I_1 I_2] - \mathbb{E}[I_1]\mathbb{E}[I_2]\Big),\end{aligned}$$

  where the second and third equalities follow since the I_i are identically distributed Bernoulli random variables with parameter q (dependent).

- We have

$$\begin{aligned}\mathbb{E}[I_1 I_2] &= \mathbb{P}(I_1 = 1, I_2 = 1) = \mathbb{P}(\text{both 1 and 2 are isolated}) \\ &= (1-p)^{2n-3} = \frac{q^2}{(1-p)}.\end{aligned}$$

- Combining the preceding two relations, we obtain

$$\begin{aligned}\text{var}(X) &= nq(1-q) + n(n-1)\left[\frac{q^2}{(1-p)} - q^2\right] \\ &= nq(1-q) + n(n-1)\frac{q^2 p}{1-p}.\end{aligned}$$

Proof (Continued)

- For large n, we have $q \to 0$ [cf. Eq. (3)], or $1 - q \to 1$. Also $p \to 0$. Hence,

$$\begin{aligned} \text{var}(X) &\sim nq + n^2 q^2 \frac{p}{1-p} \sim nq + n^2 q^2 p \\ &= nn^{-\lambda} + \lambda n \log(n) n^{-2\lambda} \\ &\sim nn^{-\lambda} = \mathbb{E}[X], \end{aligned}$$

  where $a(n) \sim b(n)$ denotes $\frac{a(n)}{b(n)} \to 1$ as $n \to \infty$.

- This implies that

$$\mathbb{E}[X] \sim \text{var}(X) \geq (0 - \mathbb{E}[X])^2 \mathbb{P}(X = 0),$$

  and therefore,

$$\mathbb{P}(X = 0) \leq \frac{\mathbb{E}[X]}{\mathbb{E}[X]^2} = \frac{1}{\mathbb{E}[X]} \to 0.$$

- It follows that $\mathbb{P}(\text{at least one isolated node}) \to 1$ and therefore, $\mathbb{P}(\text{disconnected}) \to 1$ as $n \to \infty$, completing the proof.

Converse

- We next show claim (2), i.e., if $p(n) = \lambda \frac{\log(n)}{n}$ with $\lambda > 1$, then $\mathbb{P}(\text{connectivity}) \to 1$, or equivalently $\mathbb{P}(\text{disconnectivity}) \to 0$.
- From Eq. (4), we have $\mathbb{E}[X] = n \cdot n^{-\lambda} \to 0$ for $\lambda > 1$.
- This implies probability of having isolated nodes goes to 0. However, we need more to establish connectivity.
- The event "graph is disconnected" is equivalent to the existence of k nodes without an edge to the remaining nodes, for some $k \leq n/2$.
- We have

$$\mathbb{P}(\{1, \ldots, k\} \text{ not connected to the rest}) = (1-p)^{k(n-k)},$$

and therefore,

$$\mathbb{P}(\exists \text{ k nodes not connected to the rest}) = \binom{n}{k} (1-p)^{k(n-k)}.$$

Converse (Continued)

- Using the union bound [i.e. $\mathbb{P}(\cup_i A_i) \leq \sum_i \mathbb{P}(A_i)$], we obtain

$$\mathbb{P}(\text{disconnected graph}) \leq \sum_{k=1}^{n/2} \binom{n}{k} (1-p)^{k(n-k)}.$$

- Using Stirling's formula $k! \sim \left(\frac{k}{e}\right)^k$, which implies $\binom{n}{k} \leq \frac{n^k}{(\frac{k}{e})^k}$ in the preceding relation and some (ugly) algebra, we obtain

$$\mathbb{P}(\text{disconnected graph}) \to 0,$$

completing the proof.

Phase Transitions — Connectivity Threshold

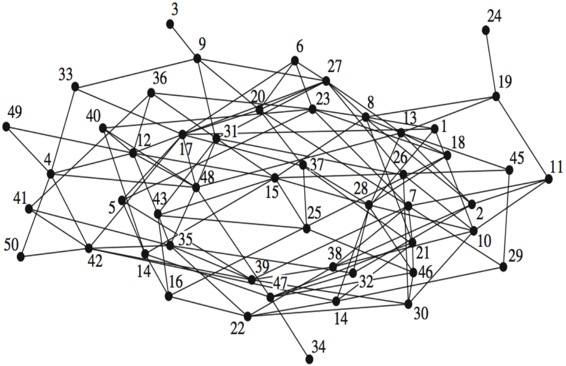


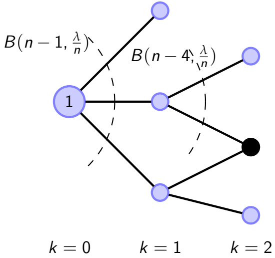
Figure: Emergence of connectedness: a random network on 50 nodes with $p = 0.10$.

Giant Component

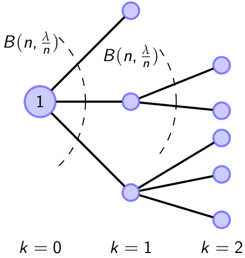
- We have shown that when $p(n) << \frac{\log(n)}{n}$, the Erdös-Renyi graph is disconnected with high probability.
- In cases for which the network is not connected, the component structure is of interest.
- We have argued that in this regime the expected number of isolated nodes goes to infinity. This 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 $p(n) = \frac{\lambda}{n}$.
- Let $B(n, \frac{\lambda}{n})$ denote a binomial random variable with n trials and success probability $\frac{\lambda}{n}$.
- Consider starting from an arbitrary node (node 1 without loss of generality), 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_k^G and Z_k^B 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

$$Z_k^G \leq Z_k^B \qquad \text{for all } k.$$

- From branching process analysis (see Lecture 3 notes), we have

$$\mathbb{E}[Z_k^B] = \lambda^k,$$

  (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

$$\mathbb{E}[S_1] = \sum_k \mathbb{E}[Z_k^G] \leq \sum_k \mathbb{E}[Z_k^B] = \sum_k \lambda^k = \frac{1}{1-\lambda}.$$

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

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

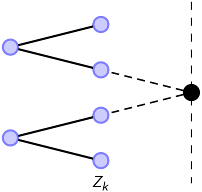
Here $|S_i|$ is the size of the component that contains node i.

- 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.

Emergence of the Giant Component—4

- 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

$$\mathbb{E}[\text{number of conflicts at stage } k+1] \leq np^2 \mathbb{E}[Z_k^2] = n\frac{\lambda^2}{n^2}\mathbb{E}[Z_k^2].$$



- We assume for large n that Z_k is a Poisson random variable and therefore $\text{var}(Z_k) = \lambda^k$. This implies that

$$\mathbb{E}[Z_k^2] = \text{var}(Z_k) + \mathbb{E}[Z_k]^2 = \lambda^k + \lambda^{2k} \approx \lambda^{2k}.$$

- 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 $\mathbb{P}(\text{there exists a component with size } \geq c\sqrt{n} \text{ nodes}) \to 1$ as $n \to \infty$.
- Moreover, between any two components of size $\sqrt{n}$, the probability of having a link is given by

$$\mathbb{P}(\text{there exists at least one link}) = 1 - (1 - \frac{\lambda}{n})^n \approx 1 - e^{-\lambda},$$

  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

$$\mathbb{P}(\text{node } n \text{ not in the giant component}) = 1 - q \equiv \rho.$$

- 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

$$\rho = \sum_d P_d \rho^d \equiv \Phi(\rho).$$

- Similar to the analysis of branching processes, we can show that 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 π.
- 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 πn 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:
- $p(1-\pi)n < 1$:

$$\mathbb{E}[\text{size of epidemic as a fraction of the society}] \leq \frac{\log(n)}{n} \approx 0.$$

- $1 < p(1-\pi)n < \log((1-\pi)n)$:

$$\mathbb{E}[\text{size of epidemic as a fraction of the society}]$$
$$= \frac{qq(1-\pi)n + (1-q)\log((1-\pi)n))}{n} \approx q^2(1-\pi),$$

where q denotes the fraction of nodes in the giant component of the graph with $(1-\pi)n$ nodes, i.e., $q = 1 - e^{-q(1-\pi)np}$.

- $p > \frac{\log((1-\pi)n)}{(1-\pi)n}$:

$$\mathbb{E}[\text{size of epidemic as a fraction of the society}] = (1-\pi).$$