1.022 Introduction to Network Models

Amir Ajorlou

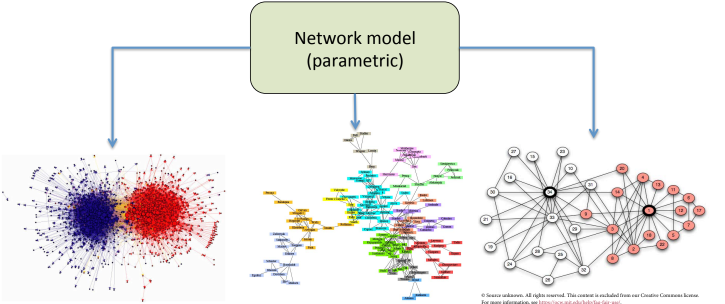
Laboratory for Information and Decision Systems
Institute for Data, Systems, and Society
Massachusetts Institute of Technology

Lecture 8

Why models for networks?

- Usual trade-off between losing details in an idealized representation while gaining insights into the simplified problem

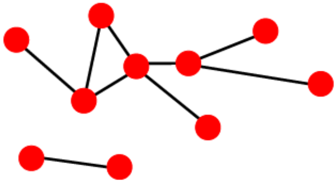
- Simple representations of complex networks
- Derive properties mathematically
- Predict properties and outcomes
- Common features of different real networks

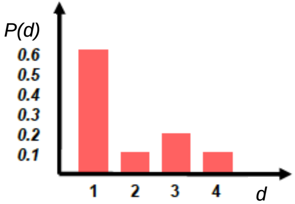


- Let $N(d)$ denote the number of vertices with degree d

  $\Rightarrow$ Fraction of vertices with degree d is $\mathrm{P}[d] := \frac{N(d)}{|V|}$

- The collection $\{\mathrm{P}[d]\}_{d \geq 0}$ is the degree distribution of G
  - Histogram formed from the degree sequence (bins of size one)



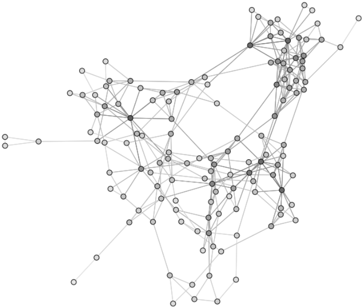


- $\mathrm{P}[d]$ = probability that randomly chosen node has degree d

  $\Rightarrow$ Summarizes the local connectivity in the network graph

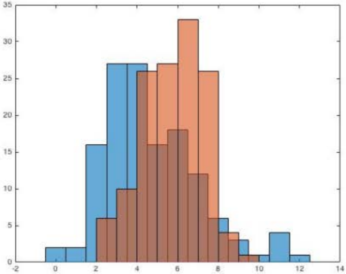
Friendship paradox

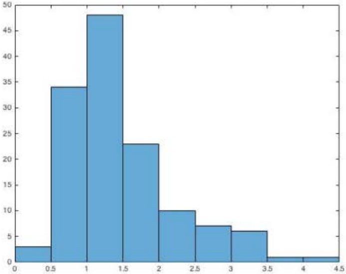
- First observed by Feld (1991)
  - ⇒ "Why Your Friends Have More Friends Than You Do"
  - ⇒ American Journal of Sociology
- Example: Network of 135 households from a rural Indian village
  - ⇒ Banerjee, Chandrasekhar, Duflo, and Jackson (2013)



Friendship paradox

- Jackson (2016)





- Left: empirical distribution of households' degrees (blue) and the distribution of average neighbors' degrees (red)
- Right: empirical distribution of ratio of average neighbors' degree over own degree

Friendship paradox

- Network of firm-level input-output linkages in Japan
- Carvalho, Nirei, Saito, and Tahbaz-Salehi (2016)

	Disaster Area Firms	Firms in the Rest of Japan
Log Sales	11.54	11.74
	(1.52)	(1.64)
Customers' Log Sales	14.83	14.51
	(2.37)	(2.45)
Suppliers' Log Sales	14.30	14.60
	(2.21)	(2.49)

Carvalho, Vasco M., Makoto Nirei, Yukiko U. Saito, et al. "Supply Chain Disruptions: Evidence from the Great East Japan Earthquake." December 2016.

- Firms' customers and suppliers are on average larger than the average firm

Friendship paradox

Theorem [Jackson, 2016] In any given undirected network, the average degree of neighbors at least as high as the average degree:

$$\frac{1}{n}\sum_{i:d_i>0}\frac{\sum_{j\in N_i} d_j}{d_i} \geq \frac{1}{n}\sum_{i=1}^{n} d_i.$$

Furthermore, the inequality is strict if and only if at least two linked agents have different degrees.

- In any given network,

$$\sum_{i:d_i>0}\frac{\sum_{j\in N_i} d_j}{d_i} = \sum_{i<j:i\in N_j}\left(\frac{d_i}{d_j}+\frac{d_j}{d_i}\right) = \sum_{i<j:i\in N_j}\left(\frac{(d_i-d_j)^2}{d_i d_j}+2\right).$$

- And as a result,

$$\sum_{i:d_i>0}\frac{\sum_{j\in N_i} d_j}{d_i} \geq \sum_{i<j:i\in N_j} 2 = \sum_{i=1}^{n} d_i.$$

The Erdős-Rényi (ER) model

- Very simple model $\Rightarrow$ Not applicable to many real networks

  $\Rightarrow$ Easy to get insights, studied in the 1950s

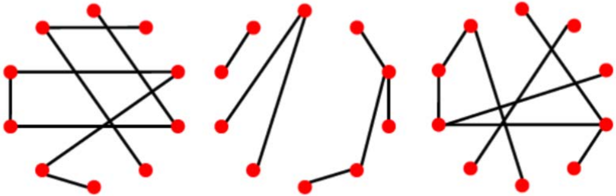
- Generating an ER random graph $G_{n,p}$

  $\Rightarrow$ 1) Choose a number of vertices n

  $\Rightarrow$ 2) Choose a probability p

  $\Rightarrow$ 3) For each possible edge, add it with probability p

- Examples of $G_{10,\frac{1}{6}}$ $\Rightarrow$ $|E|$ is a random variable



Degree distribution of $G_{n,p}$

- Q: Degree distribution $\mathrm{P}[d]$ of the Erdős-Rényi graph $G_{n,p}$?
- Define $\mathbb{I}\{(v,u)\} = 1$ if $(v,u) \in E$, and $\mathbb{I}\{(v,u)\} = 0$ otherwise.

  $\Rightarrow$ Fix v. For all $u \neq v$, the indicator RVs are i.i.d. Bernoulli(p)
- Let D_v be the (random) degree of vertex v. Hence,

$$D_v = \sum_{u \neq v} \mathbb{I}\{(v,u)\}$$

  $\Rightarrow$ D_v is binomial with parameters $(n-1, p)$ and

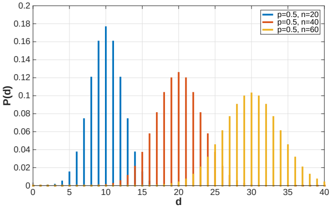
$$\mathrm{P}[d] = \mathrm{P}[D_v = d] = \binom{n-1}{d} p^d (1-p)^{(n-1)-d}$$

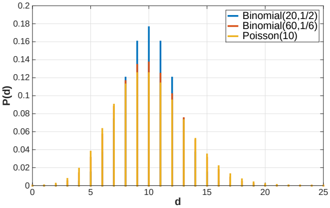
- In words, the probability of having exactly d edges incident to v

  $\Rightarrow$ Same for all $v \in V$, by independence of the $G_{n,p}$ model

- Q: How does the degree distribution look like for a large network?
- Recall D_v is a sum of $n-1$ i.i.d. Bernoulli(p) RVs

  $\Rightarrow$ Central Limit Theorem: $D_v \sim \mathcal{N}(np, np(1-p))$ for large n





- Makes most sense to increase n with fixed $\mathbb{E}[D_v] = (n-1)p = \mu$

  $\Rightarrow$ Law of rare events: $D_v \sim \text{Poisson}(\mu)$ for large n

  $\Rightarrow \mathrm{P}[D_v = d] = e^{-\mu}\frac{\mu^d}{d!}$

Behavior of $G_{n,p}$ for increasing p

- ER graphs exhibit phase transitions

  $\Rightarrow$ Sharp transitions between behaviors as $n \to \infty$

ER connectivity theorem

- A threshold function for the connectivity of $G_{n,p(n)}$ is $p(n) = \frac{\ln(n)}{n}$

- Let $p(n) = \lambda \frac{\ln(n)}{n}$ then

  $\Rightarrow$ If $\lambda < 1$ $\Rightarrow \mathbb{P}(connected) \to 0$ as $n \to infty$

  $\Rightarrow$ If $\lambda > 1$ $\Rightarrow \mathbb{P}(connected) \to 1$ as $n \to infty$

ER giant component theorem

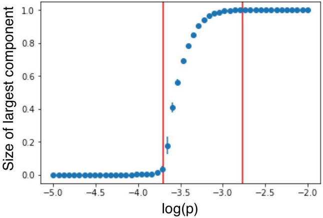
- A threshold function for the emergence of a giant component in $G_{n,p(n)}$ is $p(n) = \frac{1}{n}$

- Let $p(n) = \frac{\lambda}{n}$ then

  $\Rightarrow$ If $\lambda < 1$ $\Rightarrow$ Size of largest component $\sim \ln(n)$ as $n \to infty$

  $\Rightarrow$ If $\lambda > 1$ $\Rightarrow$ Size of largest component $\sim n$ as $n \to infty$

- You will test this in the homework
  - ⇒ Plot the relative size of largest component
  - ⇒ As a function of $log(p)$



- Movie: $G_{n,p}$ for increasing p
  - ⇒ https://www.youtube.com/watch?v=mpe44sTSoF8

1.022 Introduction to Network Models
Fall 2018