Random Graphs and Networks

Rogayeh Dastranj Tabrizi
email: rda18@sfu.ca
Office: WMC 3607
Office Hours: Thursdays 12pm-2pm

Department of Economics
Simon Fraser University

Thanks to Matthew Jackson for access to his teaching resources.

16. Februar 2015

Random Networks

All the examples we have seen so far suggest the need for models of how and why networks forms as they do.

- Most basic network formation model: A completely random process is responsible for link formation in the networks.
- By comparing real observed networks to random ones, we can identify which elements of social structures are not random.
- Study properties of random networks such as
  - Distribution of links across different nodes,
  - Connectedness of the network in terms of paths between nodes,
  - Average and maximum path lengths and number of isolated nodes.

Erdos and Renyi Networks

Erdos and Renyi (1959-1960) provided seminal studies of purely random networks:

- Consider a set of nodes: $N = \{1, ..., n\}$.
- Let a link between i and j be formed with probability $0 < p < 1$.
- This formation is independent across links.
- This is binomial model of link formation and serves as a benchmark model.

A Simple Example

Assume $n = 3$ and links are formed with probability p. Then

- A complete network is formed with probability p^3.
- A network with two links is formed with probability $p^2(1-p)$.
- A network with one link is formed with probability $p(1-p)^2$.
- An empty network is formed with probability $(1-p)^3$.

General Case

Any network with m links on n nodes is formed with probability:

$$p^m(1-p)^{\frac{n(n-1)}{2}-m}$$

Some Statistics

- Degree distribution of a random network describes the probability that any given node will have degree d:

$$\left(\begin{array}{c} n-1 \\ d \end{array} \right) p^d (1-p)^{n-1-d} = \frac{(n-1)!}{d!(n-1-d)!} p^d (1-p)^{n-1-d}$$

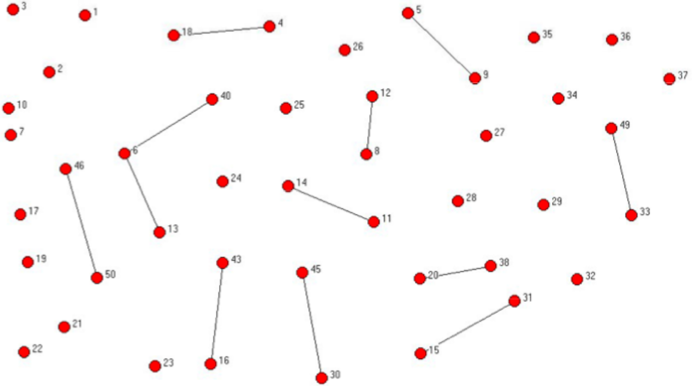
- For large n and small p this binomial expression is approximated by Poisson Distribution:

$$\frac{[(n-1)p]^d e^{-(n-1)p}}{d!}$$

- The class of random networks for which each link is formed independently with equal probability is called Poisson Random Networks.

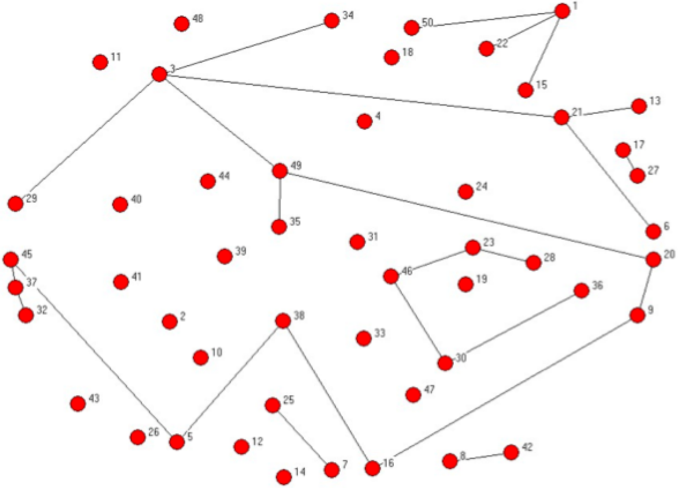
Example: $n = 50, p = 0.01$

Based on approximation of Poisson distribution with $n = 50,\ p = 0.01$ we should expect $0.61 = e^{-(n-1)p}$ of nodes be isolated (i.e. $d = 0$).



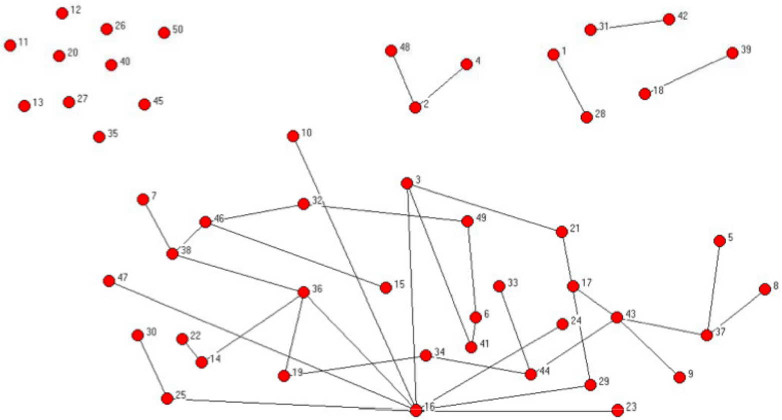
Example: $n = 50, p = 0.02$

Number of isolated nodes $0.37 = e^{-(n-1)p}$:

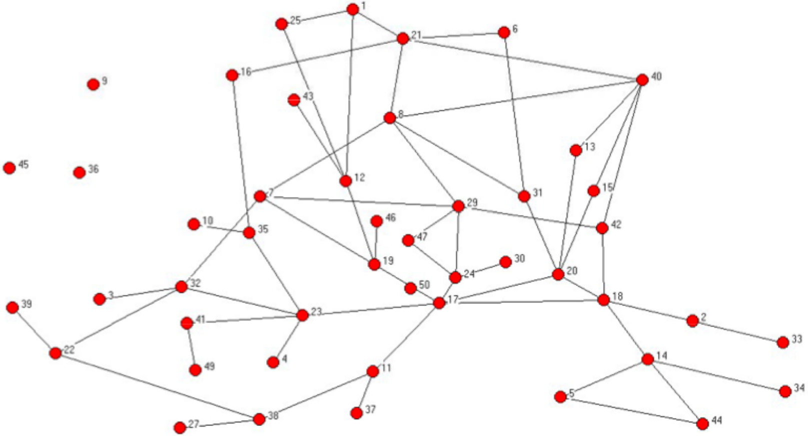


Example: $n = 50, p = 0.03$

$p = 0.02$ is the threshold for emergence of cycles and a giant component:

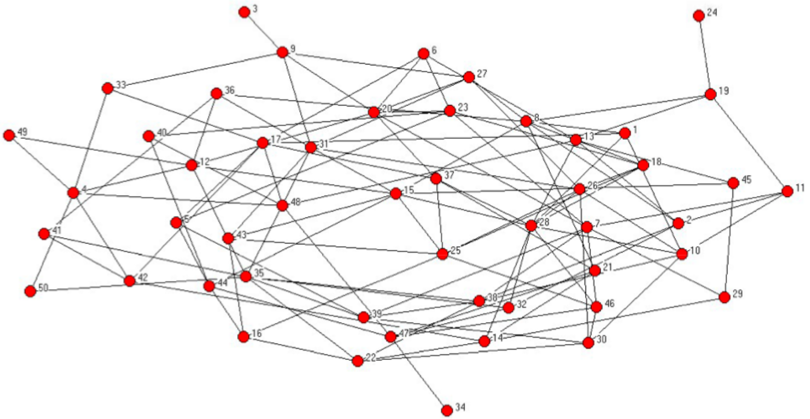


Example: $n = 50, p = 0.05$



Example: $n = 50, p = 0.10$

.08 is the threshold for connection



In Summary

- With small p, the network is a forest (i.e. no cycles).
  - The chance of there being a cycle is relatively low with such a small probability of link formation.
  - There are several components in the network, where one of the components is much larger than others.
- There is Phase Transition in the network with multiple isolated network to a path connected network with no isolated node when the average degree becomes greater than $\log(n)$.
  - Probability of having isolated nodes $(d = 0)$ is: $e^{-(n-1)p}$ when average degree $(n-1)p$ is not too large.
  - Then probability of having one isolated node on average is :

$$e^{-(n-1)p} = \frac{1}{n} \quad \Rightarrow \quad p(n-1) = \log(n)$$

Application

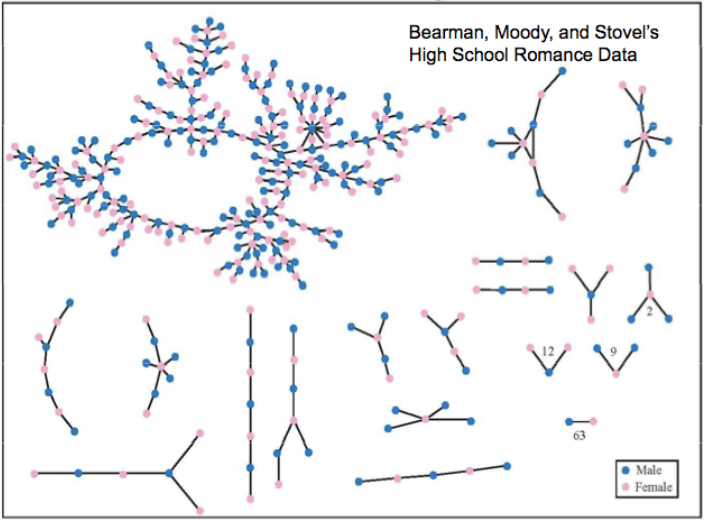
- Idea, disease, computer virus spreads via connections in the network
- Nodes are linked if one would “infect” the other
- Will an infection take hold?
- How many nodes/people will it reach?

Component Structure Is Important

- Reach of contagion is determined by the component structure
- Get nontrivial diffusion if someone in the giant component is infected/adopts
- Some nodes are immune and some links fail to transmit
- Size of the giant component determines likelihood of diffusion and its extent
- Random network models allow for giant component calculations

Component Structure Is Important

The Structure of Romantic and Sexual Relations at "Jefferson High School"



Giant Component

How big is the giant component?

- Assume q is fraction of nodes in largest component.
- Therefore the probability that a node is in giant component is q, and the probability that it is outside is $1 - q$.
- The chance that this node is **outside of the giant component** is the chance that all of its neighbours are outside of the giant component.
- probability that all of its neighbours are outside $(1 - q)^d$, where d is its degree.

Giant Component

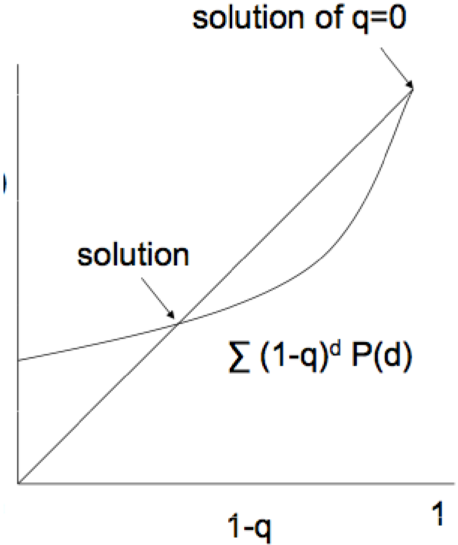
- So, probability $1 - q$ that a node is outside of the giant component is

$$1 - q = \sum_d (1 - q)^d P(d)$$

  Where $P(d)$ is the chance that the node has d neighbours. Therefore, the expression on the right hand side is the average of $(1 - q)^d$ across all nodes.

- Solve for q.

Giant Component



Conclusion

Studying random networks help us answer following questions:

- When do we get diffusion?
- What is the extent of diffusion?
- How does it depend on the particulars of the process as well as the network?
- Who is likely to be infected earliest?