# SOCIAL NETWORK ANALYTICS

## Random Networks

**Random Graphs as Models of Networks**

**Prakash C O**

Department of Computer Science and Engineering

# SOCIAL NETWORK ANALYTICS

## Random Networks

### Random Graphs as Models of Networks

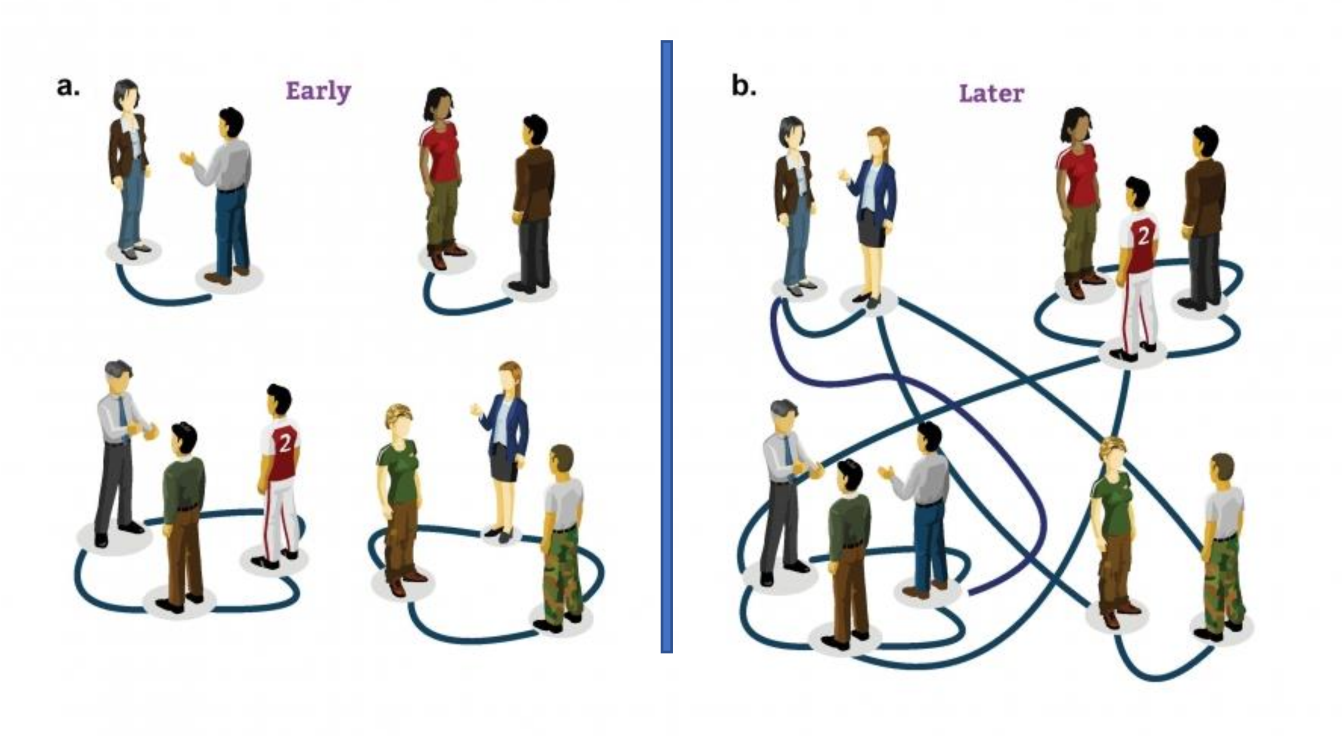
**Prakash C O**

Department of Computer Science and Engineering

## Models for network growth

- Network science aims to build models that reproduce the properties of real networks.
- From a modeling perspective, a network is a relatively simple object, consisting of only nodes and links.
- The real challenge, however, is to decide where to place the links between the nodes so that we reproduce the complexity of a real system.
- **In this respect the philosophy behind a random network is simple:** We assume that this goal is best achieved by placing the links randomly between the nodes. That takes us to the definition of a random network.

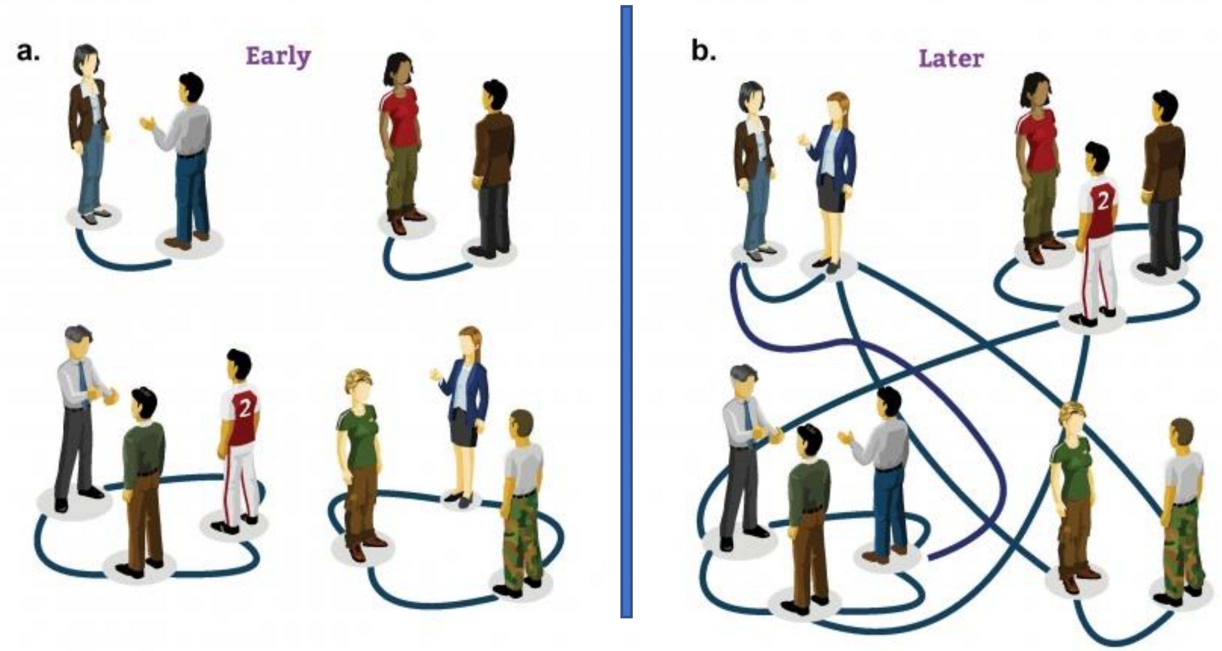
## Models for network growth




### From a Cocktail Party to Random Networks

**We start with the most basic assumption on how friendships can be formed: friendships between individuals are formed randomly.**
**The random graph model follows this basic assumption.**

## Models for network growth

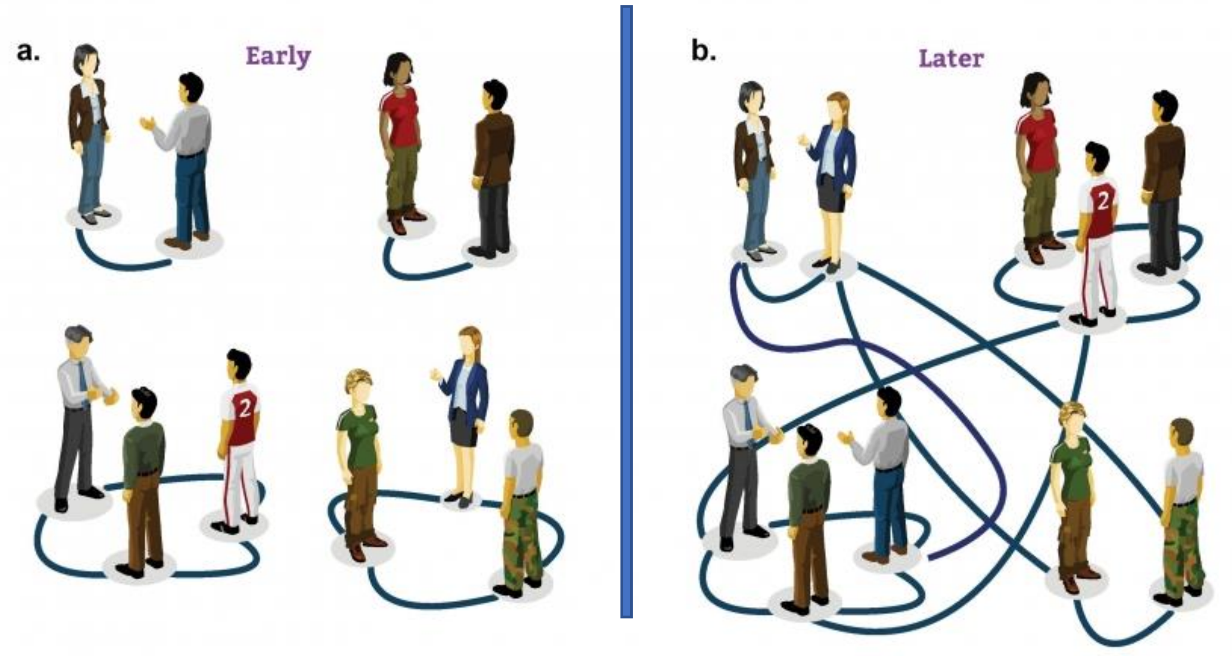



### From a Cocktail Party to Random Networks

The emergence of an acquaintance network through random encounters at a cocktail party.

- Early on the guests form isolated groups.
- As individuals mingle, changing groups, an invisible network emerges that connects all of them into a single network.

## Models for network growth

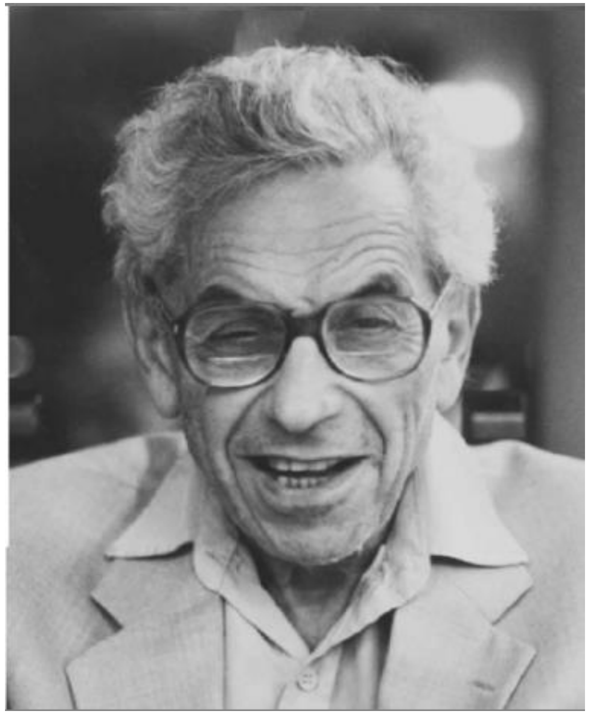



### From a Cocktail Party to Random Networks

Friendships in real-world networks are far from random. By assuming random friendships, we simplify the process of friendship formation in real-world networks, hoping that these random friendships ultimately create networks that exhibit common characteristics observed in real-world networks.

## Random Networks - A Brief History

**Pál Erdős (1913-1996)**



**Alfréd Rényi (1921-1970)**

In a series of seminal papers in the 1950s and 1960s, Paul Erdos and Alfred Renyi proposed and studied one of the earliest theoretical models of a network, the random graph (Erdos and Renyi, 1959, 1960, 1961).

## Random Networks - A Brief History

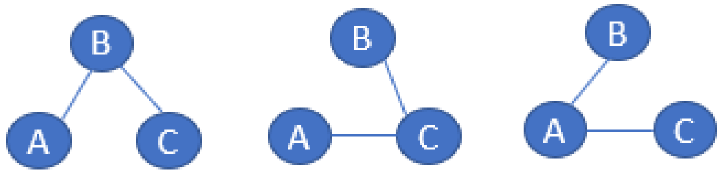
- **Anatol Rapoport (1911-2007), a Russian immigrant to the United States, was the first to study random networks.**
  In a paper written with Ray Solomonoff in 1951, Rapoport demonstrated that if we increase the average degree of a network, we observe an abrupt transition from disconnected nodes to a graph with a giant component.

- **The study of random networks reached prominence thanks to the fundamental work of Pál Erdős and Alfréd Rényi.**
  In 1959, aiming to describe networks seen in communications and the life sciences, Erdos and Renyi suggested that such systems could be effectively modelled by connecting their nodes with randomly placed links.

**There are two definitions of a random network:**

## 1. G(N, L) Model

- **N labeled nodes are connected with L randomly placed links**. Erdős and Rényi used this definition in their string of papers on random networks.
- In the G(N, L) model, a graph is chosen uniformly at random from the collection of all graphs which have N nodes and L edges.
- For example, in the G(3, 2) model, each of the three possible graphs on three vertices and two edges are included with probability 1/3.




- **The G(N, L)) model fixes the total number of links as L** and the average degree of a node is simply ‹k› = 2L/N

PES UNIVERSITY ONLINE

## ➢G(N, L) Model Illustration in NetLogo

1. http://ccl.northwestern.edu/netlogo/

**Go to File/Models Library/Code Examples/Random Network Example**

PES UNIVERSITY ONLINE

**There are two definitions of a random network:**

## 2. G(N, p) Model or G(n, p) Model

- **Each pair of N labeled nodes is connected with probability p.**
  This model was first proposed independently by Edgar Gilbert and Solomonoff and Rapoport.
- **The G(N, p) model fixes the probability *p* that two nodes are connected.**
- The network characteristics are easier to calculate in the $G(N, p)$ model.
  We will explore more on $G(N, p)$ model, not only for the ease that it allows us to calculate key network characteristics, but also because in real networks the number of links rarely stays fixed.

## Defining Random Networks

### ➢G(N, p) Model Illustration in NetLogo

1. http://ccl.northwestern.edu/netlogo/

   **Go to File/Models Library/Code Examples/Random Network Example**

## Random Network Construction

- **A random network consists of N nodes where each node pair is connected with probability p.**
- **To construct a random network we follow these steps:**
  1. **Start with N isolated nodes.**
  2. **Select a node pair and generate a random number between 0 and 1.**

     **If the number exceeds p, connect the selected node pair with a link,**

     **otherwise leave them disconnected.**
  3. **Repeat step (2) for each of the N(N-1)/2 node pairs.**

  **The network obtained after this procedure is called a random graph or a random network.**
- Erdős and Renyi, have played an important role in understanding the properties of these networks. In their honor a random network is called the Erdős-Renyi network.

## Defining Properties of Random Networks

- G(n,p) is the ensemble of graphs of n vertices in which each graph appears with the probability appropriate to its number of edges.
- Often one wishes **to express properties of G(n,p) not in terms of p but in terms of the average degree z of a vertex.**

  The average number of edges on the graph as a whole is ***(1/2)n(n-1)p***, and the average number of ends of edges is twice this, since each edge has two ends. So **the average degree of a vertex** is

$$z = \frac{n(n-1)p}{n} = (n-1)p \simeq np,$$

where the last approximate equality is good for large n. Thus, once we know n, any property that can be expressed in terms of p can also be expressed in terms of z.

## Random Network Properties

- The Erdos-Renyi random graph has a number of desirable properties as a model of a network.

- In particular it is found that many of its ensemble average properties can be calculated exactly in the limit of large n (Bollobas, 1985; Janson et al., 1999).

- For example, one interesting feature, which was demonstrated in the original papers by Erdos and Renyi, is that **the ER model shows a phase transition with increasing z at which a giant component forms.**

$$z = \frac{n(n-1)p}{n} = (n-1)p \simeq np,$$

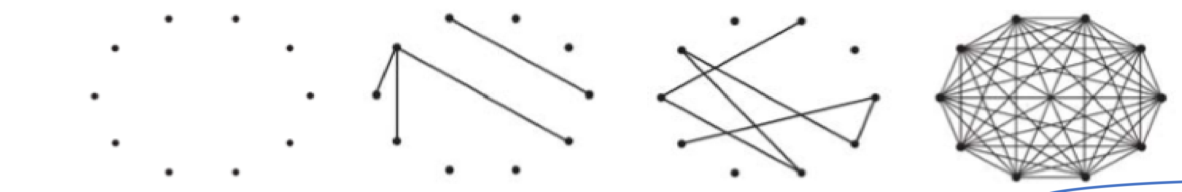
- In a 1960 paper, Erdős and Rényi described **the behavior of *G*(*n*, *p*) very precisely for various values of *p***. Their results included that:

$$z = \frac{n(n-1)p}{n} = (n-1)p \simeq np,$$

  - If ***np* < 1**, then a graph in $G(n, p)$ will almost surely have **no connected components of size larger than O(log(*n*))**.
  - If ***np* = 1**, then a graph in $G(n, p)$ will almost surely have **a largest component** whose size is of order $n^{2/3}$.
  - **If *np* > 1,** then a graph in $G(n, p)$ will almost surely have **a unique giant component containing a positive fraction of the vertices.** No other component will contain more than O(log(*n*)) vertices.

## Random Network

- We can tune the behavior of the random graph model by selecting the appropriate p value.
- In G(n,p),
  - when p = 0, the size of the largest connected component is 0 (no two pairs are connected), and
  - when p = 1, the size is n (all pairs are connected).
- **Table below provides the size of the largest connected component** (*slc* values in the table) for **random graphs with 10 nodes and different p values**.



| | | | | |
|---|---|---|---|---|
| $p$ | 0.0 | 0.055 | 0.11 | 1.0 |
| z | 0.0 | **0.495** | ≈1 | 9.0 |
| $ds$ | 0 | 2 | 6 | 1 |
| $slc$ | 0 | 4 | 7 | 10 |
| $l$ | 0.0 | 1.5 | 2.66 | 1.0 |

p=1/(n-1)=1/9=0.11

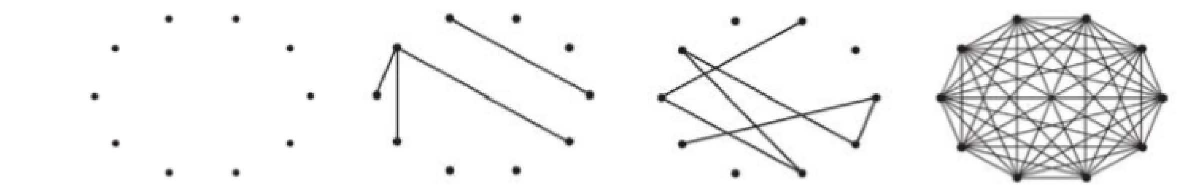
z=(n-1)p

**p - Probability**
**z - Average degree**
**ds -Diameter size**
**slc - Size of the largest component**
**l - Average path length**

## Random Network

**As shown, in Table,**

- **As p gets larger, the graph gets denser.**
- **When p is very small, the following is found:**
  1. No giant component is observed in the graph.
  2. Small isolated connected components are formed.
  3. The diameter is small because all nodes are in isolated components, in which they are connected to a handful of other nodes.



| | | | | |
|---|---|---|---|---|
| $p$ | 0.0 | 0.055 | 0.11 | 1.0 |
| z | 0.0 | **0.495** | $\approx 1$ | 9.0 |
| $ds$ | 0 | 2 | 6 | 1 |
| $slc$ | 0 | 4 | 7 | 10 |
| $l$ | 0.0 | 1.5 | 2.66 | 1.0 |

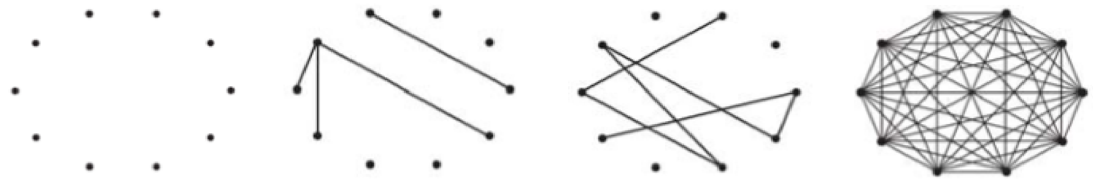
**p - Probability**
**z - Average degree**
**ds -Diameter size**
**slc - Size of the largest component**
**l - Average path length**

## Random Network

**As shown, in Table,**

➢ **As p gets larger**, the following occurs:

1. A giant component starts to appear.
2. Isolated components become connected.
3. The diameter values increase.

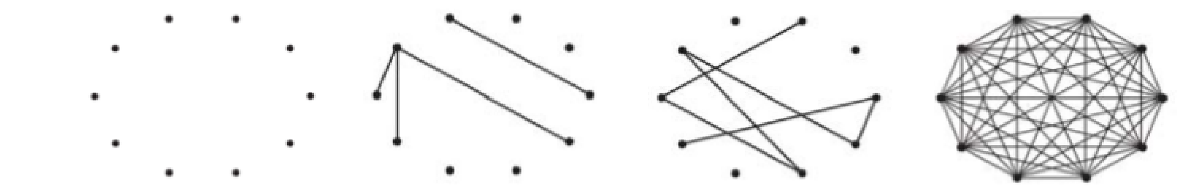


| | | | | |
|---|---|---|---|---|
| $p$ | 0.0 | 0.055 | 0.11 | 1.0 |
| z | 0.0 | **0.495** | $\approx 1$ | 9.0 |
| $ds$ | 0 | 2 | 6 | 1 |
| $slc$ | 0 | 4 | 7 | 10 |
| $l$ | 0.0 | 1.5 | 2.66 | 1.0 |

**p - Probability**
**z - Average degree**
**ds -Diameter size**
**slc - Size of the largest component**
**l - Average path length**

## Random Network

- At this point (p = 1/(n-1)=0.11 in Table below), nodes are connected to each other via long paths.
- **As p continues to get larger**, the random graph properties change again. For larger values, **the diameter starts shrinking** as nodes get connected to each other via different paths.
- **The point where diameter value starts to shrink in a random graph is called phase transition.**



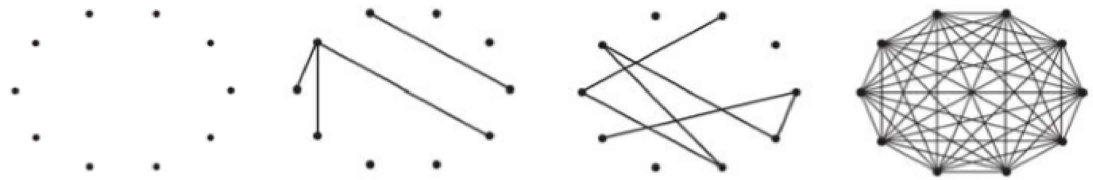
| | | | | |
|---|---|---|---|---|
| $p$ | 0.0 | 0.055 | 0.11 | 1.0 |
| z | 0.0 | **0.495** | ≈1 | 9.0 |
| $ds$ | 0 | 2 | 6 | 1 |
| $slc$ | 0 | 4 | 7 | 10 |
| $l$ | 0.0 | 1.5 | 2.66 | 1.0 |

**p - Probability**
**z - Average degree**
**ds -Diameter size**
**slc - Size of the largest component**
**l - Average path length**

## Random Network

- At the point of phase transition, the following two phenomena are observed:
  1. **The giant component, which just started to appear, starts to grow.**
  2. **The diameter, which just reached its maximum value, starts decreasing.**

- **It is proven that in random graphs phase transition occurs when z = 1; that is, p = 1/(n−1).**



| | | | | |
|---|---|---|---|---|
| $p$ | 0.0 | 0.055 | 0.11 | 1.0 |
| z | 0.0 | 0.495 | ≈1 | 9.0 |
| $ds$ | 0 | 2 | 6 | 1 |
| $slc$ | 0 | 4 | 7 | 10 |
| $l$ | 0.0 | 1.5 | 2.66 | 1.0 |

**p - Probability**
**z - Average degree**
**ds -Diameter size**
**slc - Size of the largest component**
**l - Average path length**

## Random Graph Generation demo - NetLogo

- RandomGraph (NetLogo 5.3.1)
- ErdosRenyiDegDist (NetLogo 5.3.1)

## Random Graph Generation demo - NetworkX

- RandomGraphGeneration.py

# SOCIAL NETWORK ANALYTICS

## Gaint Component model and
## Gaint Component model demo in NetLogo

**Prakash C O**

Department of Computer Science and Engineering

## Gaint Component model

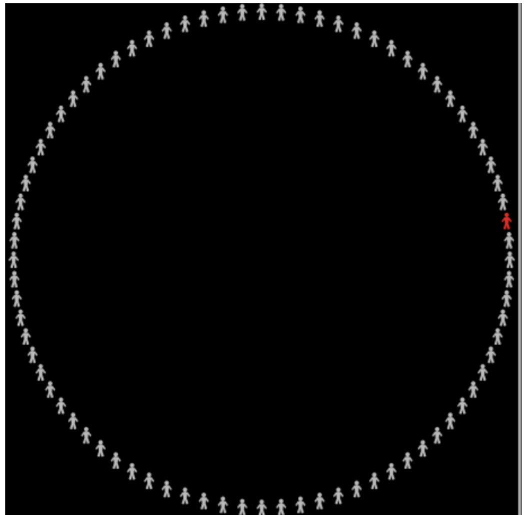
- In a network, a "**component**" is a group of nodes (people) that are all connected to each other, directly or indirectly.

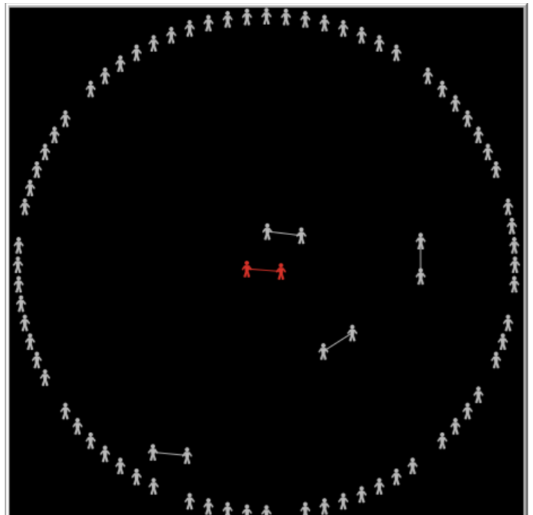
  So if a network has a "**giant component**", that means almost every node is reachable from almost every other.

- Gaint Component model (in **NetLogo**) shows how quickly a giant component arises if you grow a random network.

## Giant component    How it works in NetLogo

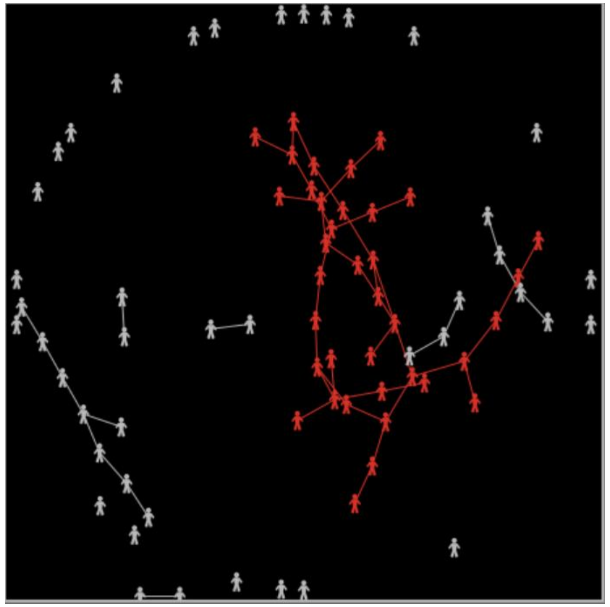
- Initially we have nodes but no connections (edges) between them.
- At each step, we pick two nodes at random which were not directly connected before and add an edge between them.
- All possible connections between them have exactly the same probability of occurring.





## Giant component

### How it works in NetLogo



- **As the model runs, small chain-like "components" are formed,** where the members in each component are either directly or indirectly connected to each other.

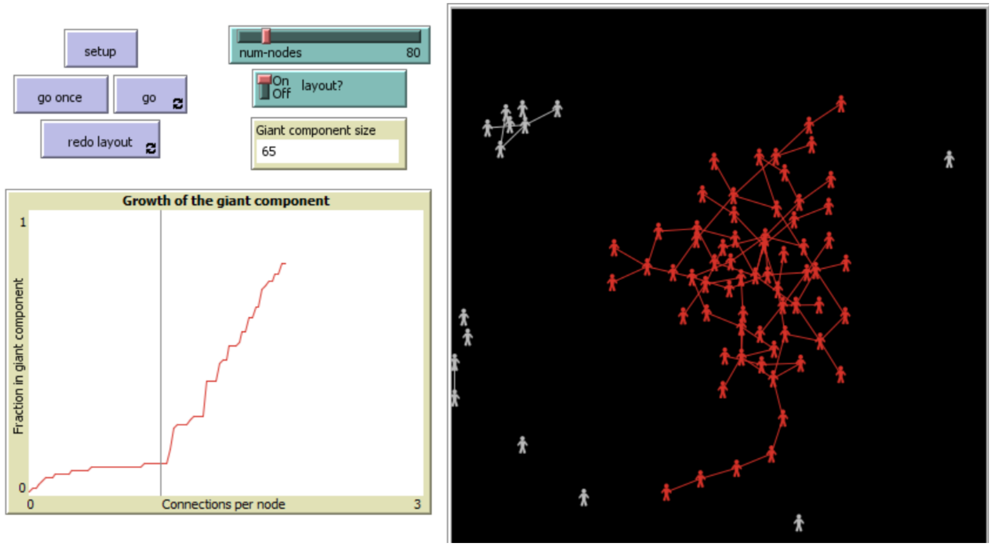
- If an edge is created between nodes from two different components, then those two components merge into one.

- **The component with the most members at any given point in time is the "giant" component and it is colored red.**

## Giant component How it works in NetLogo

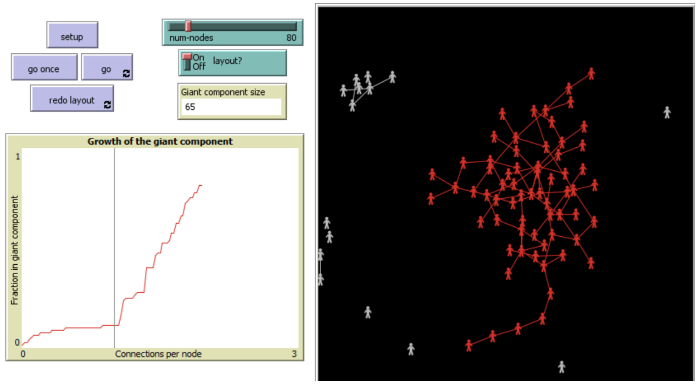
- Gaint Component model (in NetLogo demo) showed that the largest connected component of a network, **rapidly grows after the average number of connections per node equals 1.**

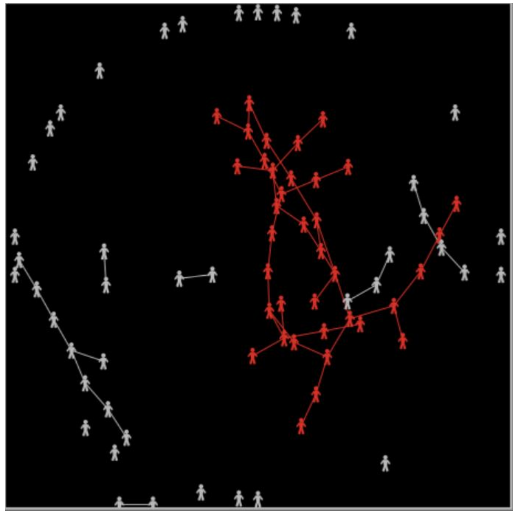
- **When n×p (i.e., average degree) tends to a constant >1** (as n grows), the graph will almost surely contain a **"giant" connected component, absorbing a considerably large fraction of the nodes.**

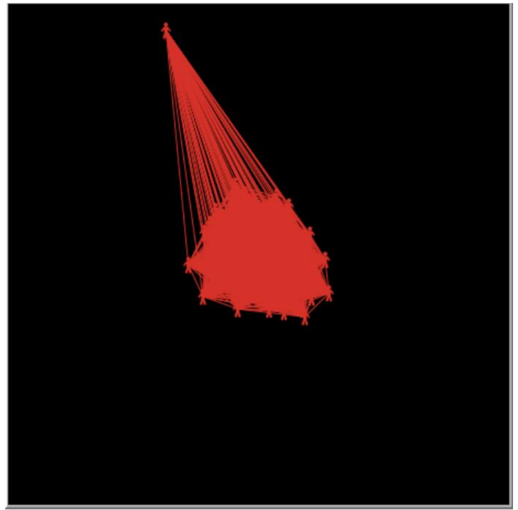


### Giant component How it works in NetLogo

- When n×p tends to a constant >1, the graph will almost surely contain a **"giant" connected component, absorbing a considerably large fraction of the nodes.**
  In other words, **the average number of connections has a "critical point" where the network undergoes a "phase transition" from a rather unconnected world of a bunch of small, fragmented components, to a world where most nodes belong to the same connected component**.





## Emergence of the giant component

**The Existence (or not) of a Giant Component is Important in Social Networks**

- If there is no giant component then communication can only take place within small groups of people
- If there is a giant component then a large fraction of network can all communicate with one another

## Random Networks

- **Giant Component Illustration in NetLogo**

- http://ccl.northwestern.edu/netlogo/

  **Go to File/Models Library/Networks/Gaint Component**

- **Demo using NetworkX**

- **plot_giant_component.py**

- https://networkx.github.io/documentation/stable/auto_examples/drawing/plot_giant_component.html

## Quiz:

➢ What is the average degree z at which the giant component starts to emerge?

a) 0

b) 1

c) 3/2

d) 3

PES UNIVERSITY ONLINE

## Assignment – Paper Reading

- Duncan J. Watts. Six Degrees: The Science of a Connected Age (W.W. Norton & Company, New York, 2003), pages 43-47.
- P. Erdos and A. Renyi. **On random graphs**. Publ. Math. Debrecen, 6:290-297, 1959.
- S. Janson, D.E. Knuth, T. Luczak, and B. Pittel. The birth of the giant component. Random Structures & Algorithms 4, 3 (1993), pages 233-358.

# SOCIAL NETWORK ANALYTICS

## Random Network - Degree Distributions

**Prakash C O**

Department of Computer Science and Engineering

- **The random graph differs from real-world networks in some fundamental ways.**
  - **First, real-world networks show strong clustering or network transitivity, where Erdos and Renyi's model does not.**

A network is said to show clustering if the probability of two vertices being connected by an edge is higher when the vertices in question have a common neighbor.

In many real-world networks the clustering coefficient is found to have a high value, anywhere from a few percent to 50 percent or even more.

## Random Networks

In Table-1 we compare clustering coefficients for a number of real-world networks with their values on a random graph with the same number of vertices and edges.

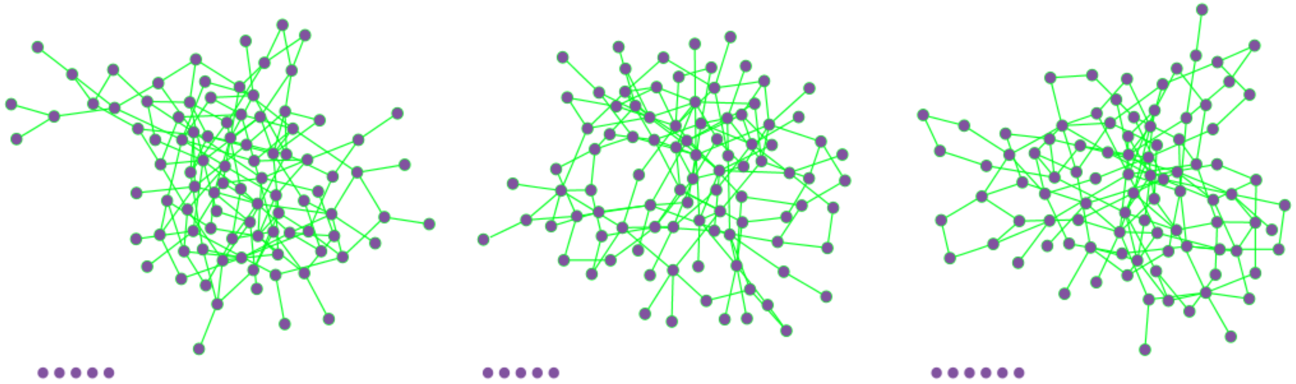
| network | $n$ | $z$ | clustering coefficient $C$ measured | clustering coefficient $C$ random graph |
|---|---|---|---|---|
| Internet (autonomous systems)[a] | 6 374 | 3.8 | 0.24 | 0.00060 |
| World-Wide Web (sites)[b] | 153 127 | 35.2 | 0.11 | 0.00023 |
| power grid[c] | 4 941 | 2.7 | 0.080 | 0.00054 |
| biology collaborations[d] | 1 520 251 | 15.5 | 0.081 | 0.000010 |
| mathematics collaborations[e] | 253 339 | 3.9 | 0.15 | 0.000015 |
| film actor collaborations[f] | 449 913 | 113.4 | 0.20 | 0.00025 |
| company directors[f] | 7 673 | 14.4 | 0.59 | 0.0019 |
| word co-occurrence[g] | 460 902 | 70.1 | 0.44 | 0.00015 |
| neural network[c] | 282 | 14.0 | 0.28 | 0.049 |
| metabolic network[h] | 315 | 28.3 | 0.59 | 0.090 |
| food web[i] | 134 | 8.7 | 0.22 | 0.065 |

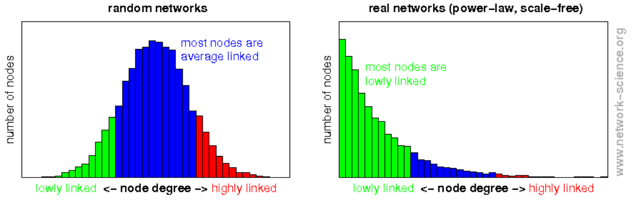
Table 1: Number of vertices $n$, mean degree $z$, and clustering coefficient $C$ for a number of different networks.

Ref: Random graphs as models of networks - M. E. J. Newman

## Random Networks

- **The random graph differs from real-world networks in some fundamental ways.**
  - A **second way in which random graphs differ from their real-world counterparts is in their degree distributions**.






- **In a given realization of a random network, there will be very few nodes with a very large degree and very few nodes with a very low degree**
- These differences are captured by the degree distribution $p_k$, which is the probability that a randomly chosen node has degree k.

**Degree Distribution - For a graph generated by G(n,p)**

- Each node has (n-1) tries to get edges
- Each try is a success with probability p
- The binomial distribution gives us the probability that a node has degree k:

$$P(\deg(v) = k) = p_k = \binom{n-1}{k} p^k (1-p)^{n-1-k}.$$

- **The probability $p_k$** that **a vertex** in an **Erdos Renyi random graph has degree exactly k** is given by the binomial distribution:

$$P(\deg(v) = k) = p_k = \binom{n-1}{k} p^k (1-p)^{n-1-k}$$

- The binomial distribution becomes Poisson distribution for large n and *z*=*np*=constant

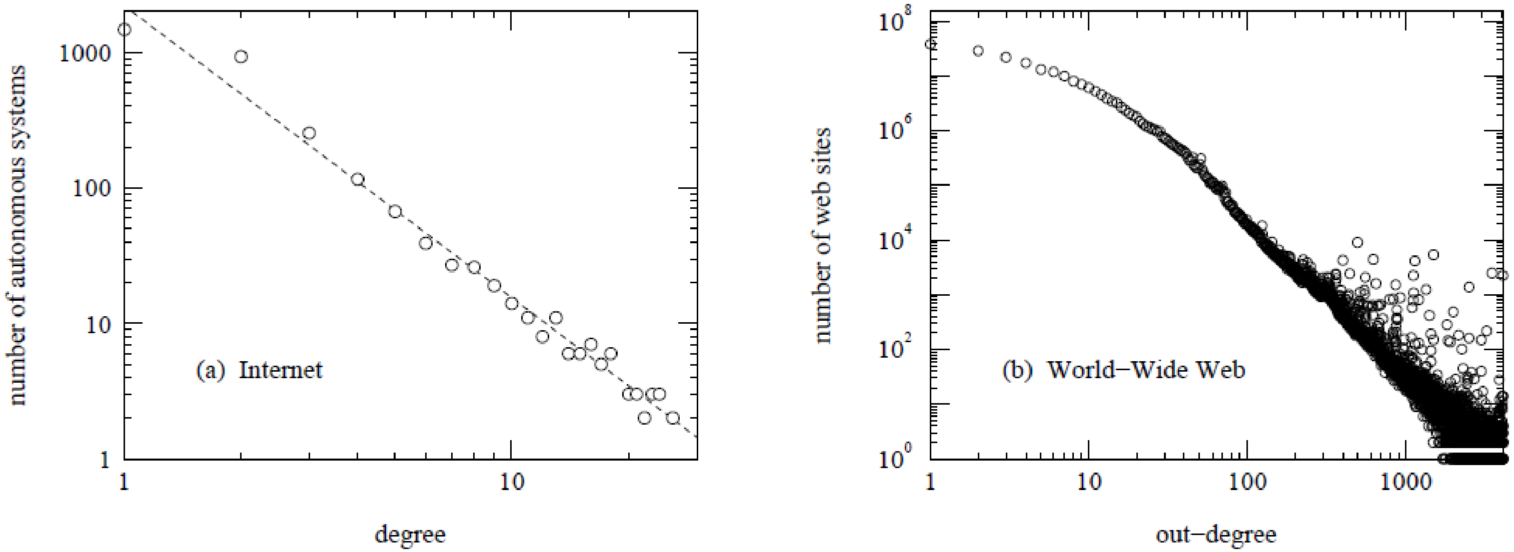
In the limit where $n \gg kz$, this becomes

$$p_k = \frac{z^k e^{-z}}{k!},$$

which is the well-known Poisson distribution. Both binomial and Poisson distributions are strongly peaked about the mean $z$, and have a large-$k$ tail that decays rapidly as $1/k!$.

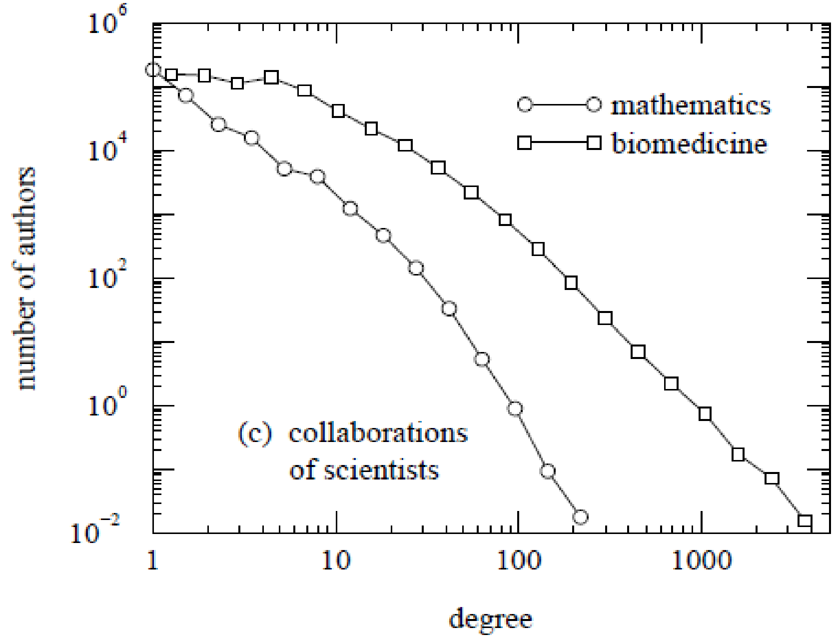
## Random Networks

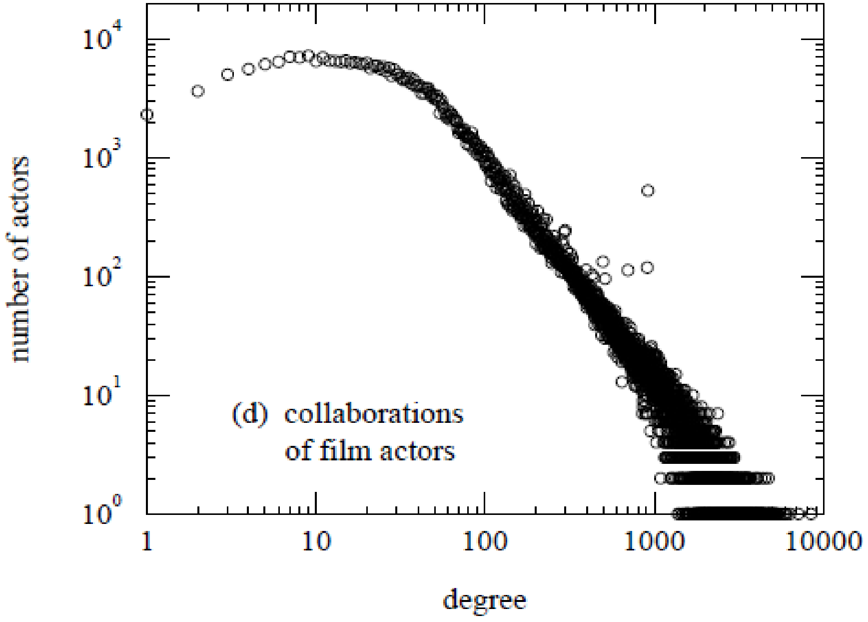
- **We can compare these predictions to the degree distributions of real networks** by constructing histograms of the degrees of vertices in the real networks.
- Measured degree distributions for a number of different networks.
    - Figure(a) Physical connections between autonomous systems on the Internet, circa 1997 (Faloutsos et al., 1999).
    - Figure(b) A 200 million page subset of the World-Wide Web, circa 1999 (Broder et al., 2000). The figure shows the out-degree of pages, i.e., numbers of links pointing from those pages to other pages.



## Random Networks

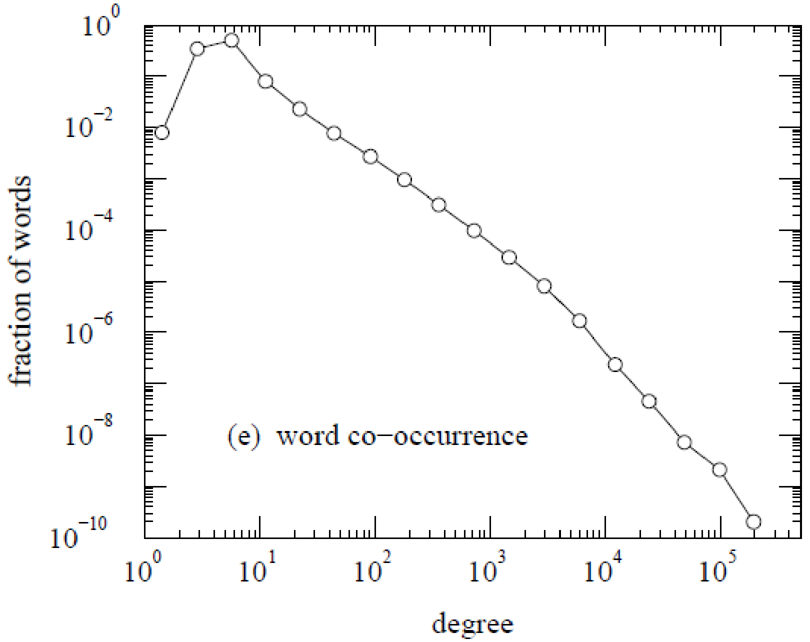
- Measured degree distributions for a number of different networks.
  - Figure(c) Collaborations between biomedical scientists and between mathematicians (Newman, 2001b,d).
  - Figure(d) Collaborations between film actors of lm actors (Amaral et al., 2000).

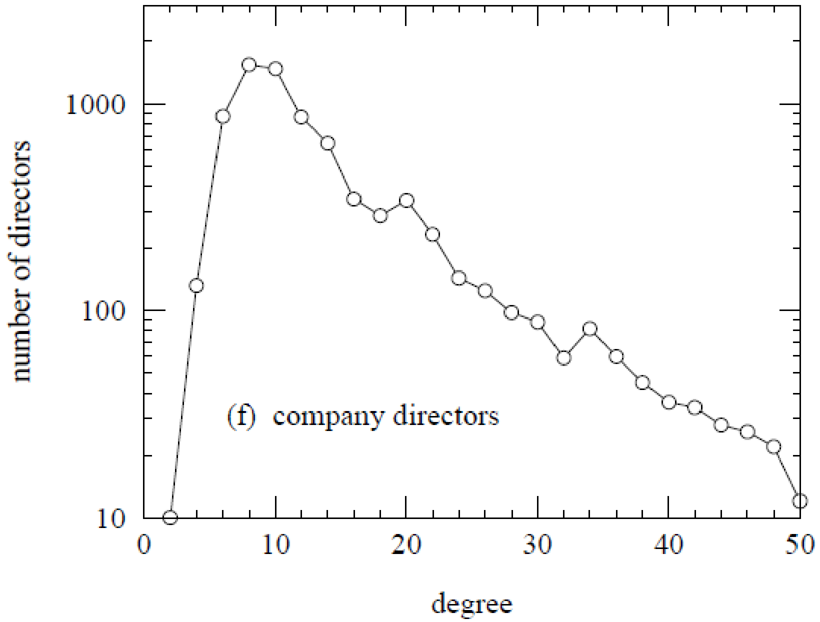


## Random Networks

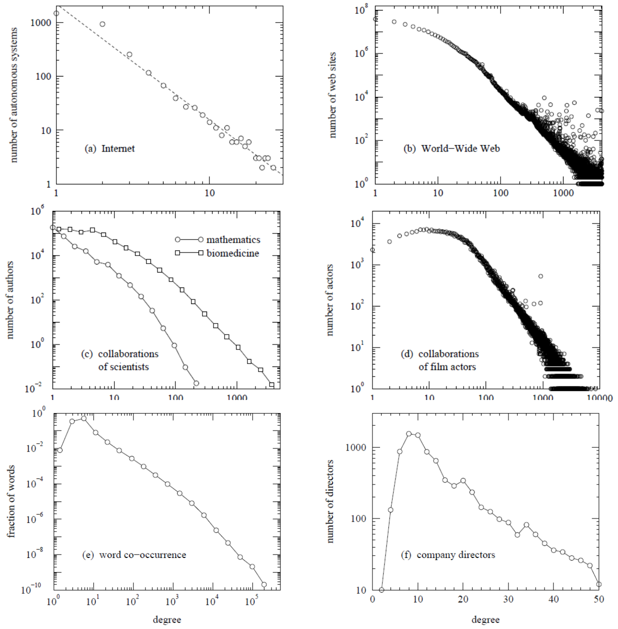
- Measured degree distributions for a number of different networks.
  - Figure(e) Co-occurrence of words in the English language (i Cancho and Sole, 2001).
  - Figure(f) Board membership of directors of Fortune 1000 companies for year 1999 (Newman et al., 2001).



## Random Networks

- As the figures (a),(b),(c),(d),(e),(f) shows, in most cases the degree distribution of the real network is very different from the Poisson distribution.
- Many of the networks, including **Internet** and **World-Wide Web graphs, appear to have power-law degree distributions,** which means that a small but non-negligible fraction of the vertices in these networks have very large degree.
- The **collaboration graphs**, appear to have power-law degree distributions with an exponential cutoff at high degree (Amaral et al., 2000; Newman, 2001a,b).
- The **graph of company directors**, seem to have degree distributions with a purely exponential tail (Newman et al., 2001).
- The **power grid network** is another example of a network that has an exponential degree distribution (Amaral et al., 2000).



## Limitations of Random Networks

1. Its degree distribution differs from that of real-world networks, which usually follow power-law.
2. The clustering coefficients of the Erdos-Renyi networks are too small.
3. Most importantly, real-world networks are not generated randomly.

## Summary

- Erdos-Renyi random graphs have been used as a benchmark for many graph-based algorithms and models,

  but now we know that they are quite unrealistic as a representation of a real-world network, because most real-world networks have a long-tail distribution of degrees (e.g., power-law).

- However, since other methods for generating networks are significantly more complex compared to random networks.

## Random Networks

- In the paper "***Random Graphs as Models of Networks -M. E. J. Newman***".
  - *M.E.J Newman* showed how to generalize the Erdos Renyi random graph to mimic the clustering and degree properties of real-world networks.

- In the paper "***Random graph models of social networks - M. E. J. Newman, D. J. Watts, and S. H. Strogatz***"
  - They described some new exactly solvable models of the structure of social networks, based on random graphs with arbitrary degree distributions.
  - They give models both for simple unipartite networks, such as acquaintance networks, and bipartite networks, such as affiliation networks.
  - **They compared the predictions of their models to data for a number of real-world social networks and find that in somecases, the models are in remarkable agreement with the data**, whereas **in others the agreement is poorer, perhaps indicating the presence of additional social structure in the network that is not captured by the random graph.**

## Quiz:

- As the size of the network increases, if you keep *p*, the probability of any two nodes being connected, the same, what happens to the average degree.

  1. Stays the same
  2. Increases
  3. Decreases

- C:\...\ErdosRenyiDegDist.nlogo

## Assignment

1. Why the Random Graph if it does not have the correct degree distribution for real-world networks?

2. Is it possible to create a model that matches real-world networks better than a random graph, but is still exactly solvable?

## References

1. “Networks – An introduction”, MEJ Neumann, Oxford University Press 2010.
2. Random Graphs as Models of Networks - M. E. J. Newman
3. Social Network Analysis: **Lada Adamic,** University of Michigan.
4. http://networksciencebook.com/
5. Wikipedia – Current Literature

# THANK YOU

**Prakash C O**

Department of Computer Science and Engineering

**copraka sha@pes.edu**

+91 98 8059 1946

## Models for network growth

- So far, we have examined models that define how a network of *n* nodes forms by adding edges. These models are static with respect to nodes
- Consider starting with a small network of few nodes and define how to grow it by adding new nodes.
- Why do networks grow?
  - citation networks (new papers written)
  - web graph (new pages created)
  - social networks (new individuals enter)
  - Collaboration networks (new authors are born)

- A **probability distribution** describes how the values of a random variable is distributed.
- **The binomial distribution is a discrete probability distribution**.
  - It describes the outcome of *n* independent trials in an experiment.
  - Each trial is assumed to have only two outcomes, either success or failure.
  - If the probability of a successful trial is *p*, then the probability of having *x* successful outcomes in an experiment of *n* independent trials is as follows.

$$f(x) = \binom{n}{x} p^x (1-p)^{(n-x)} \quad where\ x = 0, 1, 2, \ldots, n$$

## Random Networks

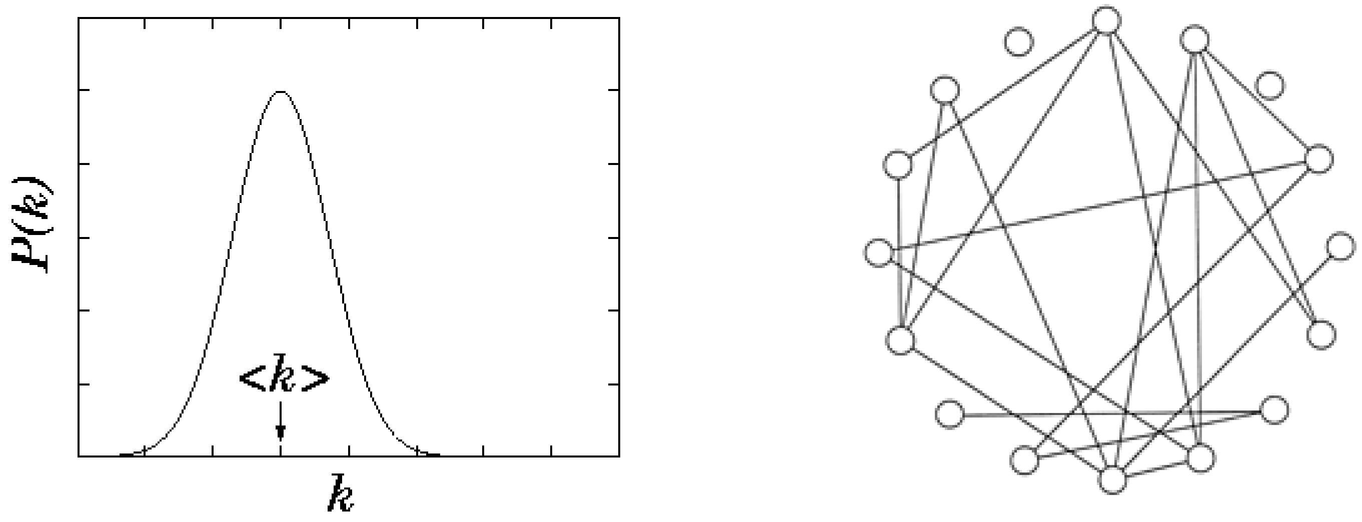



**Figure :** Degree distribution of a random graph, and an example of such a graph.

**There will be very few nodes with a very large degree and very few nodes with a very low degree**

Ref: Analysis of biological networks: Random Models - Roded Sharan

## Random Networks