Complex Network Theory

Lecture 5

Random Network Models

Instructor: S. Mehdi Vahidipour

(vahidipour@kashanu.ac.ir)

Spring 2018

Thanks A. Rezvanian

A. Barabasi, L.Adamic and J. Leskovec

Outline

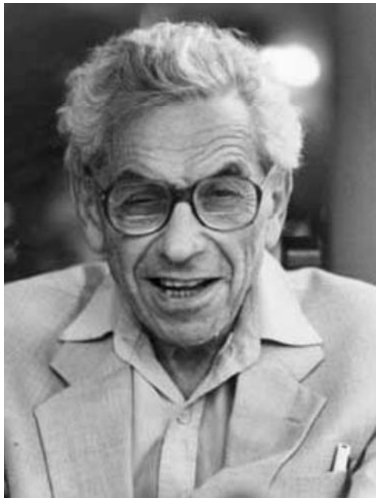
- Overview of class topics
  - Random graphs
  - Erdos-Renyi models
  - Properties of the Erdos-Renyi random graphs
  - Real networks and Erdos-Renyi random networks
  - Random geographic networks
  - Random clustered networks

- Next class:
  - Small-world networks

Birth of Random Graph Theory

- Two centuries later mathematicians moved from studying the properties of various graphs to asking:
  - How do real networks form?
  - What are the laws governing their appearance and structure?
- The first answer, came in 1950, when two Hungarian mathematicians made a revolution in graph theory:

  Paul Erdös and Alfréd Rényi



Paul Erdős (1913-1996)

Erdös number

- Erdös published over 1,500 papers with 507 coauthors.
- Mathematicians introduced the Erdös number to keep track of their distance from Erdös.
  - Erdös has Erdös number zero.
  - Those who coauthored a paper with him have Erdös number one.
  - Those who wrote a paper with an Erdös coauthor have Erdös number two, and so on.
- Examples:
  - Noam Chomsky, the famous linguist, has four.
  - Bill Gates, founder of Microsoft, who has published little science, has an Erdos number of four.

Erdös number

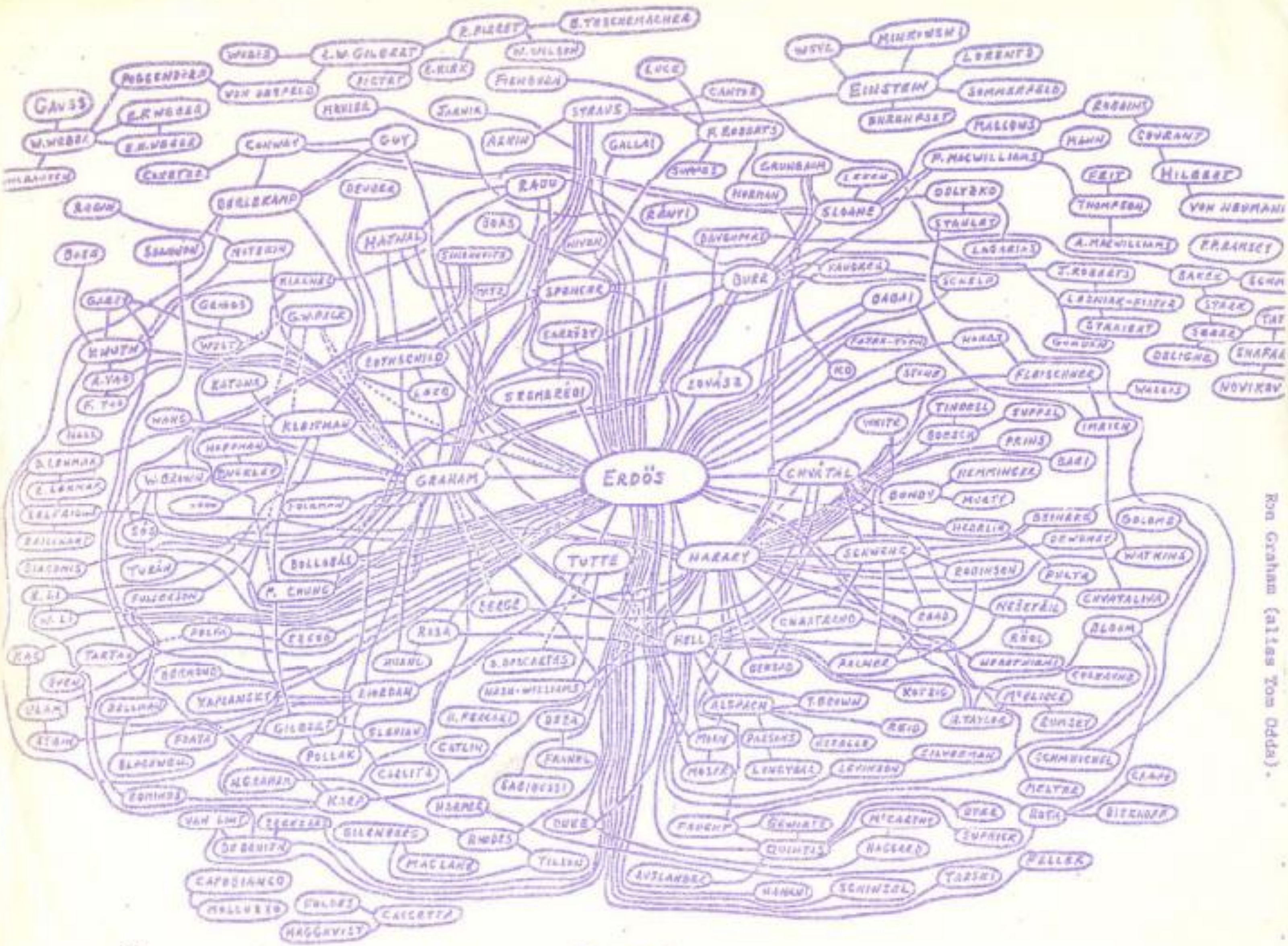
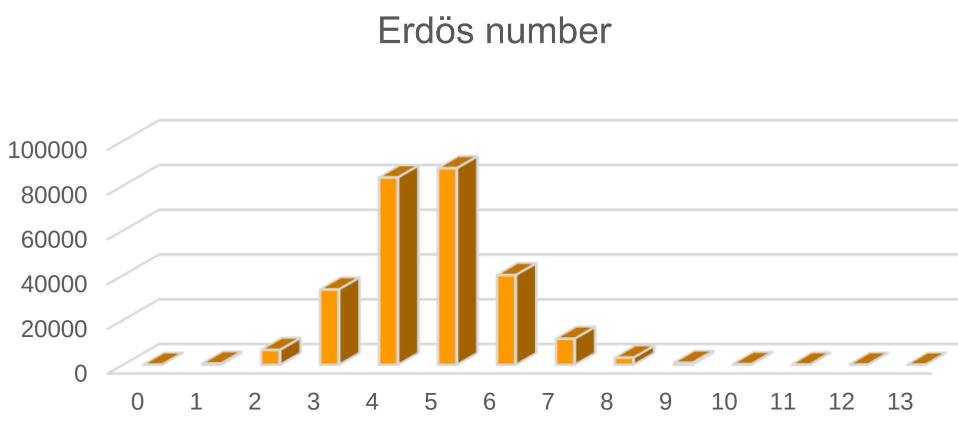


Figure 1

To appear in Topics in Graph Theory (F. Harary, ed.) New York Academy of Sciences (1979).

The distribution of Erdös numbers

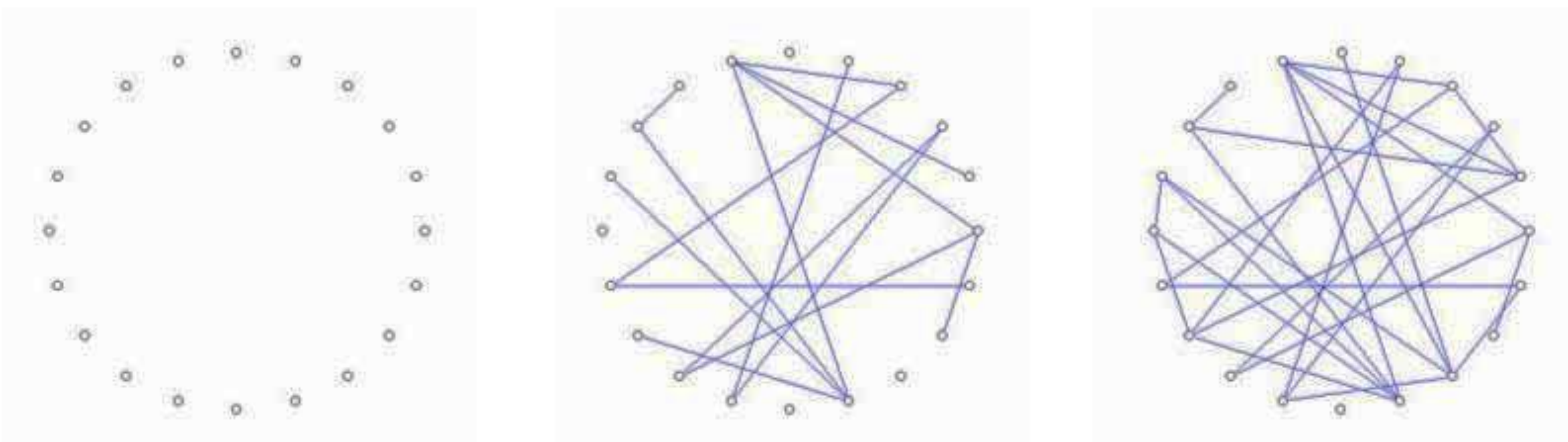
Erdös number 0 --- 1 person
Erdös number 1 --- 504 people
Erdös number 2 --- 6593 people
Erdös number 3 --- 33605 people
Erdös number 4 --- 83642 people
Erdös number 5 --- 87760 people
Erdös number 6 --- 40014 people
Erdös number 7 --- 11591 people
Erdös number 8 --- 3146 people
Erdös number 9 --- 819 people
Erdös number 10 --- 244 people
Erdös number 11 --- 68 people
Erdös number 12 --- 23 people
Erdös number 13 --- 5 people



- Thus the **median Erdös number is 5**; the **mean is 4.65**, and the **standard deviation is 1.21**.

Erdös Renyi Model

- **G(n, p) model ($G_{n,p}$)**: a graph is thought to be constructed by connecting nodes randomly.
- Each edge is included in the graph with probability **p**, with the presence or absence of any two distinct edges in the graph being independent.



$p = 0$ (a) $p = 0.1$ (b) $p = 0.2$ (c)

Erdös-Renyi Random Graphs

- For generation of Erdös-Renyi network, one of the following methods is used:

1. The $G_{n,p}$ model
   - **input:** the number of vertices n, and a parameter p, $0 \le p \le 1$
   - **process:** for each pair (i,j), generate the edge (i,j) independently with probability p
2. Related, but not identical: The $G_{n,m}$ model
   - process: select m edges uniformly at random
   - $G(n,p)$ behave similar to $G(n, m)$ with $m = pn(n-1)/2$

Properties of G(n,p)

- The total probability of drawing a graph with *m* edges from of G(n,p)

$$P(m) = \binom{\binom{n}{2}}{m} p^m (1-p)^{\binom{n}{2}-m}$$

- which is just the standard binomial distribution. Then the mean value of *m* is

$$\langle m \rangle = \sum_{m=0}^{\binom{n}{2}} m P(m) = \binom{n}{2} p$$

- the mean degree in a graph with exactly m edges is ⟨k⟩=2m/n, and hence the mean degree in G(n, p) is

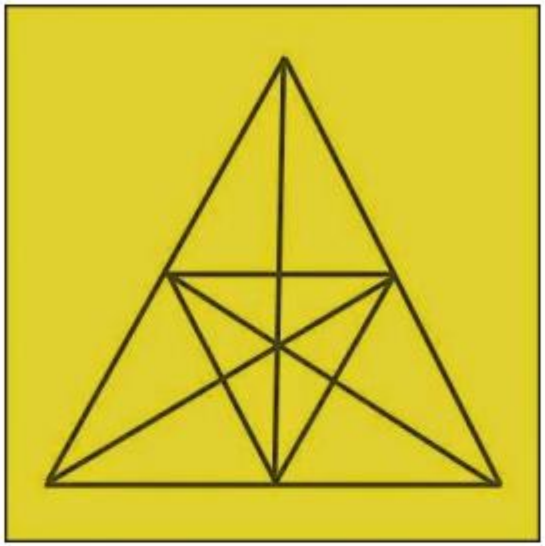
$$\langle k \rangle = \sum_{m=0}^{\binom{n}{2}} \frac{2m}{n} P(m) = \frac{2}{n}\binom{n}{2} p = (n-1)p$$

- The mean degree of a random graph is often denoted *c* in the literature

$$c = (n-1)p$$

Triangles in G(n, p=c/n)

- There are $\binom{n}{3}$ potential triangles
- Each is a triangle with probability $(c/n)^3$
- The expected number of triangles is $\binom{n}{3}(c/n)^3 \approx c^3/6$

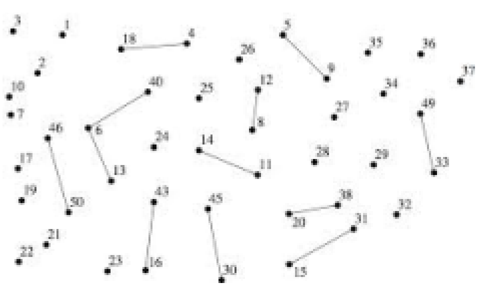


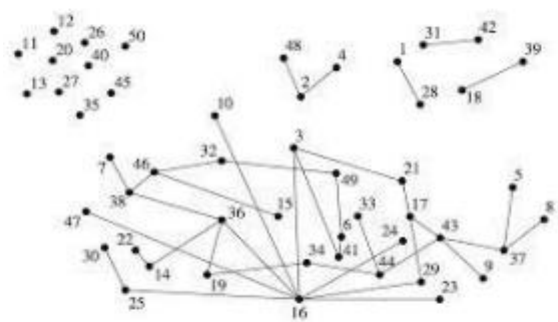
Components of G(n,p)

- The expected number of edges in undirected graphs:

  $p.n(n-1)/2$

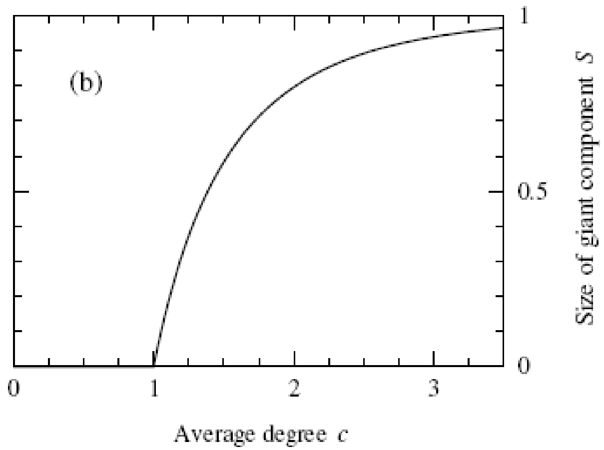
- If $p < 1/n$, then a graph in G(n, p) will almost surely have no connected components of size larger than $O(\log n)$
- If $p = 1/n$, then a graph in G(n, p) will almost surely have largest component whose size is of order $n^{2/3}$
- If $p > 1/n$, then a graph in G(n, p) will almost surely have a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

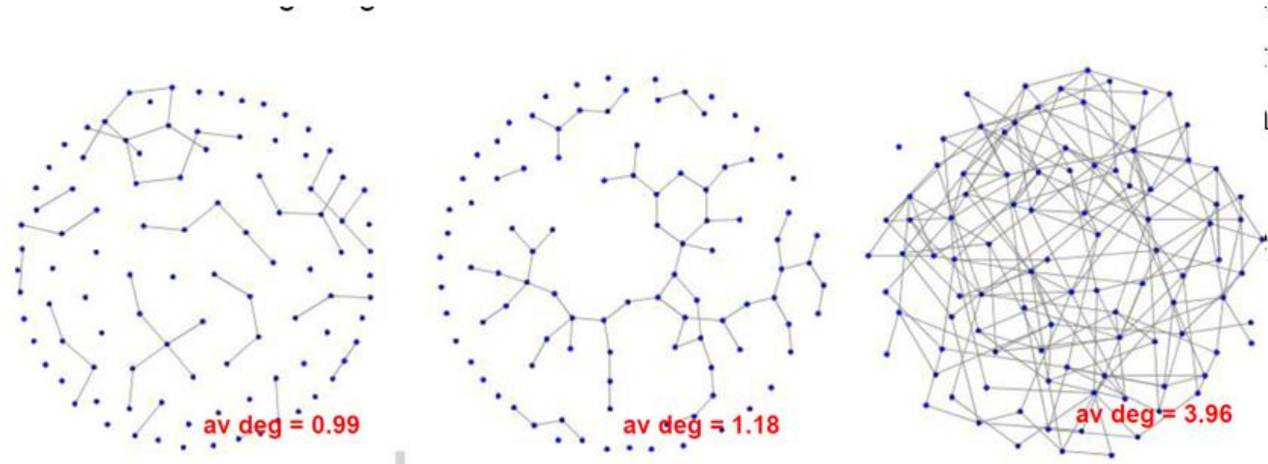




Threshold phenomena

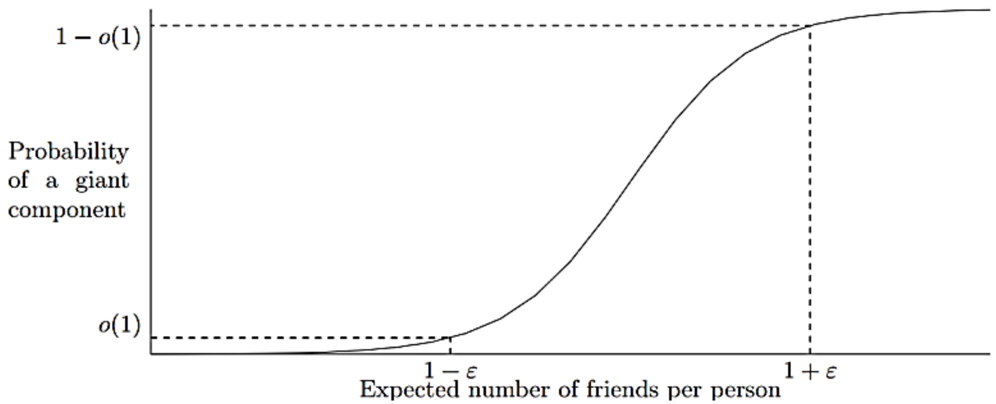
- Threshold phenomena: Many properties appear suddenly. That is, there exist a probability p_c such that for $p<p_c$ the property does not hold a.s. and for $p>p_c$ the property holds a.s.
- ***What do you expect to be a threshold phenomenon in random graphs?***
- Percolation threshold: how many edges need to added before the giant component appears?

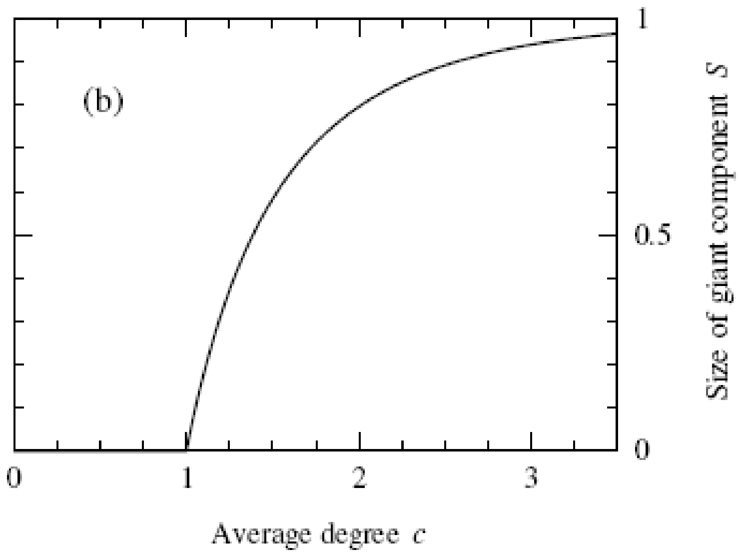




Isolated vertices in G(n,p)

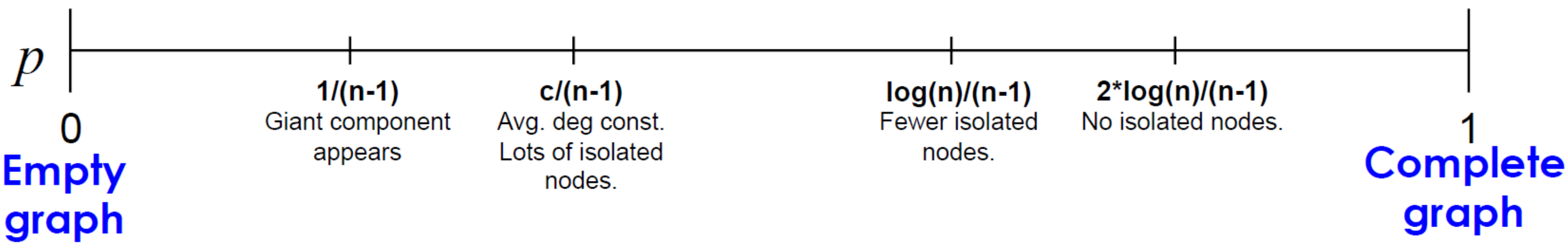
- p = Ln(n)/n is a sharp threshold for the presence of isolated vertices in a graph in G(n, p)
  - If $p < (1-\varepsilon)\ln(n)/n$, then a graph in G(n, p) will almost surely contain isolated vertices.
  - If $p > (1+\varepsilon)\ln(n)/n$, then a graph in G(n, p) will almost surely have no isolated vertices.
- Note: The behavior of random graphs are often studied in the case where n tends to infinity.





"Evolution" of a Random Graph

Graph structure of G_{np} as p changes:



Emergence of a Giant Component:

avg. degree $k=2E/n$ or $p=k/(n-1)$

- $k=1-\varepsilon$: all components are of size $\Omega(\log n)$
- $k=1+\varepsilon$: 1 component of size $\Omega(n)$, others have size $\Omega(\log n)$

Degree distribution

- A given vertex in the graph is connected with independent probability p to each of the n-1 other vertices. Thus the probability of being connected to a particular k other vertices and not to any of the others is **$p^k(1-p)^{n-1-k}$**
- There are $\binom{n-1}{k}$ ways to choose those k other vertices, and hence the total probability of being connected to exactly k others is

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

which is a binomial distribution

Degree distribution

- For the random network model of Erdos and Renyi, the histogram follows a Binomial distribution B(n,p).

$$p(k) = B(n;k;p) = \binom{n}{k} p^k (1-p)^{n-k}$$

- The binomial distribution converges towards the Poisson distribution
  - if n is sufficiently large and p is sufficiently small z = np can be used as an approximation.

$$p(k) = P(k;z) = \frac{z^k}{k!} e^{-z}$$

- Highly concentrated around the mean, with a tail that drops exponentially

Meaning of Poisson Distribution

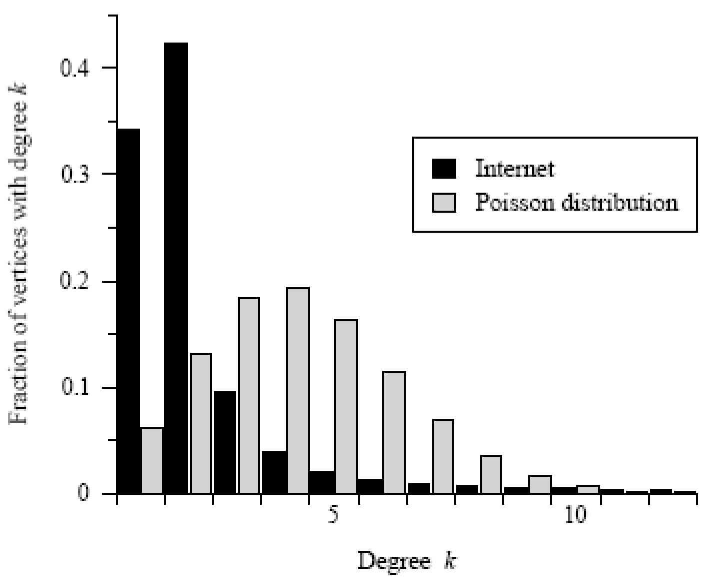
- Random graph theory predicts that if we assign social links randomly, we end up with an extremely democratic society.
- It predicts that it is exponentially rare to find someone who deviates from the average by having considerably more or fewer links than the average person.
- Erdos and Renyi's random universe predicts that:
  - most people have roughly the same number of acquaintances;
  - most neurons connect roughly to the same number of other neurons;
  - most companies trade with roughly the same number of other companies;
  - most Websites are visited by roughly the same number of visitors.
  - You don’t expect large hubs in the network !
- ! Are they true?

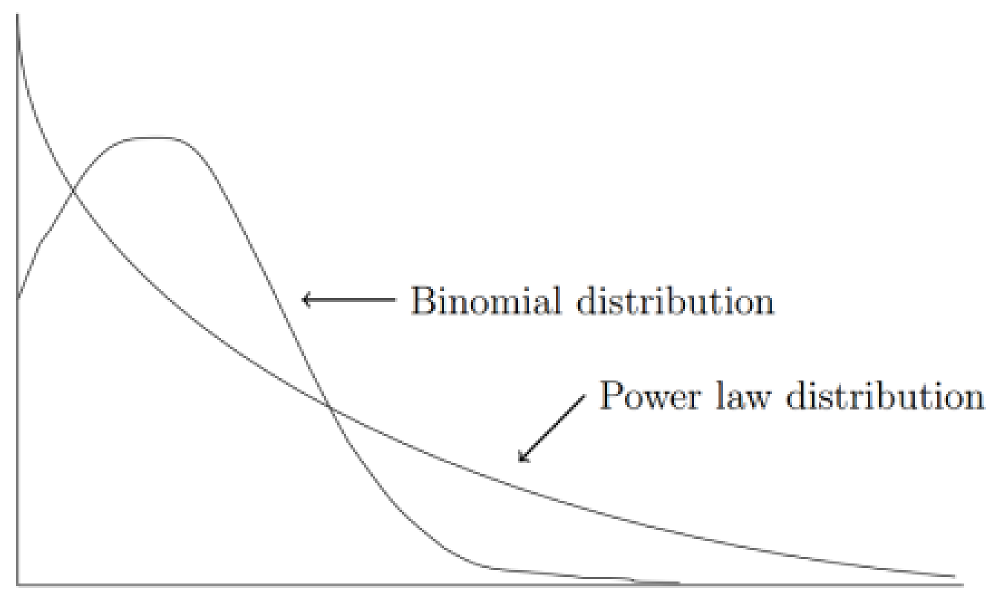
Number of links in real networks

- Nature repeatedly and extravagantly exceeds the one-link minimum.
- Sociologists estimate that we know between 200 and 5,000 people by name.
- Each company is inevitably linked to 100 of suppliers and customers; some of the biggest have links to millions.
- In our body, most molecules take part in far more than a single reaction—some, like water, in 100.

Real-world degree distributions

- Tail of a random variable = values far from mean (measured in number of standard variations)
- Tail of binomial distribution falls off exponentially fast
- Many graphs in applications have “heavy” tails
- Models more complex than G(n,p) needed for real-world applications





Erdos–Renyi random graphs and Real networks

- Beyond threshold
  - Random network theory tells us that as the average number of links per node increases beyond the critical one, the number of nodes left out of the giant cluster decreases exponentially.
  - That is, the more links we add, the harder it is to find a node that remains isolated.
- Density
  - Nature does not take risks by staying close to the threshold. It well surpasses it.
  - The networks around us are very dense networks from which nothing can escape and within which every node is navigable.
  - There are multiple paths between different nodes
- Summary

	Degree distribution	Clustering coefficient	Average diameter
Real networks	Power-law	High	Small
ER networks	Poisson	Low	Small

Other properties

- Average degree: $c=p(n-1)$
- Clustering coefficient
  - C = z/n , C=p (z=np)
- Average distance among any two points in a random graph
  - L= ln(n)
- Diameter (maximum length of the shortest paths)
  - D = log n / log z
- Phase transition
  - Starting from some vertex v perform a BFS walk
  - At each step of the BFS a Poisson process with mean z, gives At each step of the BFS a Poisson process with mean z, gives birth to new nodes
  - When $z<1$ this process will stop after $O(\log n)$ steps
  - When $z>1$ this process will continue for $O(n)$ steps

Readings

- Newman, Mark. **Networks: an introduction**. Oxford University Press, 2010. (Chapter 12)

- Van Steen, Maarten. "**Graph Theory and Complex Networks An Introduction**, 2010. (Chapter 7)

- Newman, Mark EJ. "**Random graphs as models of networks**." Handbook of Graphs and Networks: From the Genome to the Internet (2006).