Social Network Analysis

Sakthi Balan M

LNMIIT

Social Network Analysis

Network Models

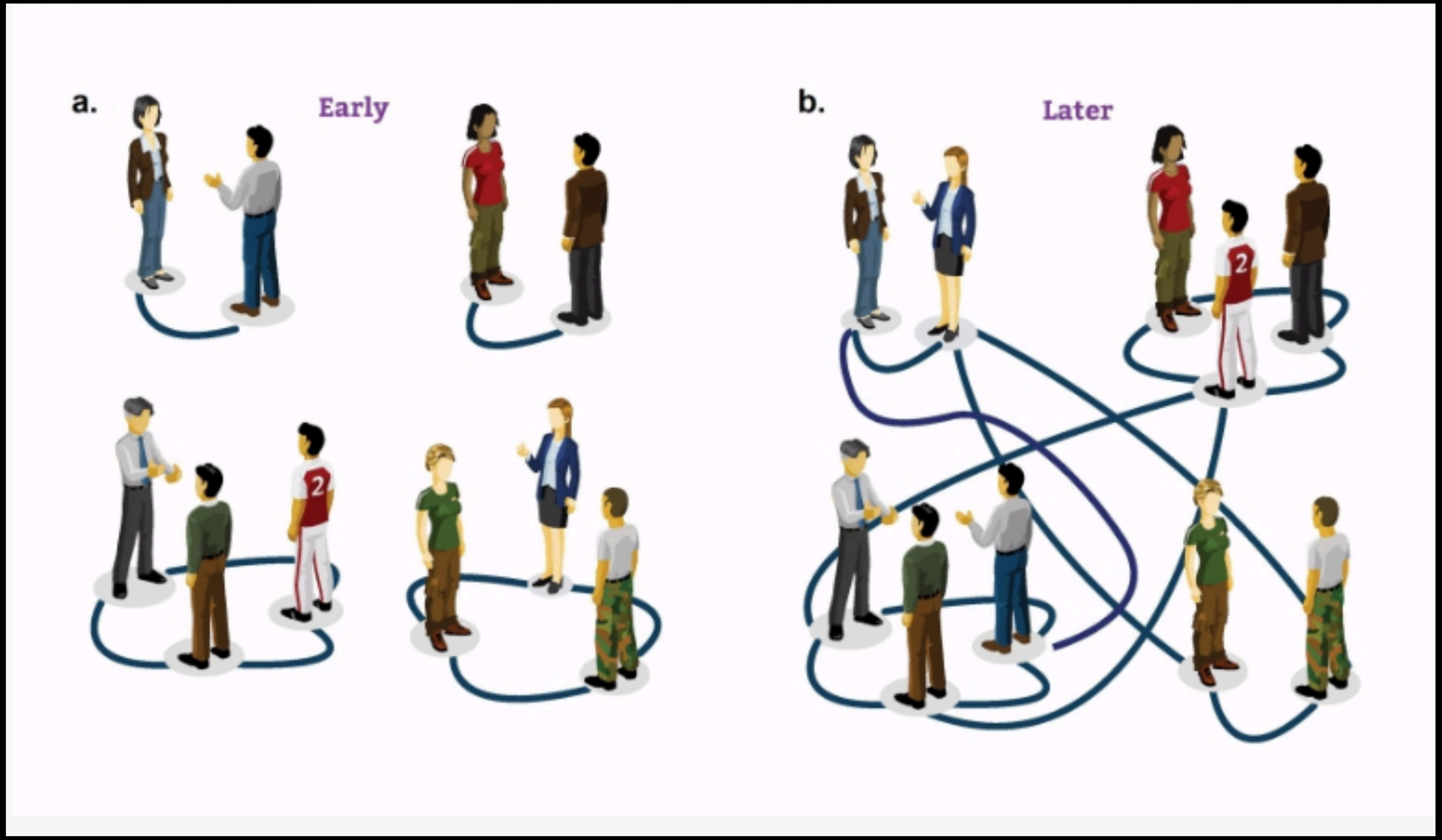
LNMIIT

Some pictures have been used from the following books:
(This is only for learning purpose)

Social Media Mining by Zafaranai et al,
Network Science by Barabasi
Networks, Crowd, Behaviour by Kleinberg et al

Acquaintance Network

- a party for a hundred guests who initially do not know each other
- offer them wine and cheese
- they chat in groups of two to three
- mention to Mary, one of your guests, that the red wine in the unlabeled dark green bottles is a rare vintage
- How long you need to wait to see red wine vanish!?



LNMIIT

The Random Network Model

- Most networks do not have the regularity of a crystal lattice or the predictable design like a spider web
- Real networks look as if they were spun randomly
- So what is random networks and how to build random networks?
- The real challenge in building is to decide where to place the links between the nodes so that we reproduce the complexity of a real system

LNMIIT

The Random Network Model

- A random network consists of N nodes where each node pair is connected with probability p

- There are two definitions of a random network:

  - 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

  - G(N, p) Model: Each pair of N labeled nodes is connected with probability p, a model introduced by Gilbert

LNMIIT

The Random Network Model

- 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.

Two mathematicians, Pál Erdős and Alfréd Rényi, have extensively studied in understanding the properties of these networks. To honor them a random network is called as the Erdős-Rényi network

LNMIIT

Number of Links in G(N,p)

The probability that a random network has exactly L links is the product of three terms:

- The probability that L of the attempts to connect the $N(N-1)/2$ pairs of nodes have resulted in a link, which is p^L.
- The probability that the remaining $N(N-1)/2 - L$ attempts have not resulted in a link, which is $(1-p)^{N(N-1)/2-L}$.
- A combinational factor,

$$\binom{\frac{N(N-1)}{2}}{L} \qquad (3.0)$$

counting the number of different ways we can place L links among $N(N-1)/2$ node pairs.

We can therefore write the probability that a particular realization of a random network has exactly L links as

$$\boxed{p_L = \binom{\frac{N(N-1)}{2}}{L} p^L (1-p)^{\frac{N(N-1)}{2}-L} \qquad (3.1)}$$

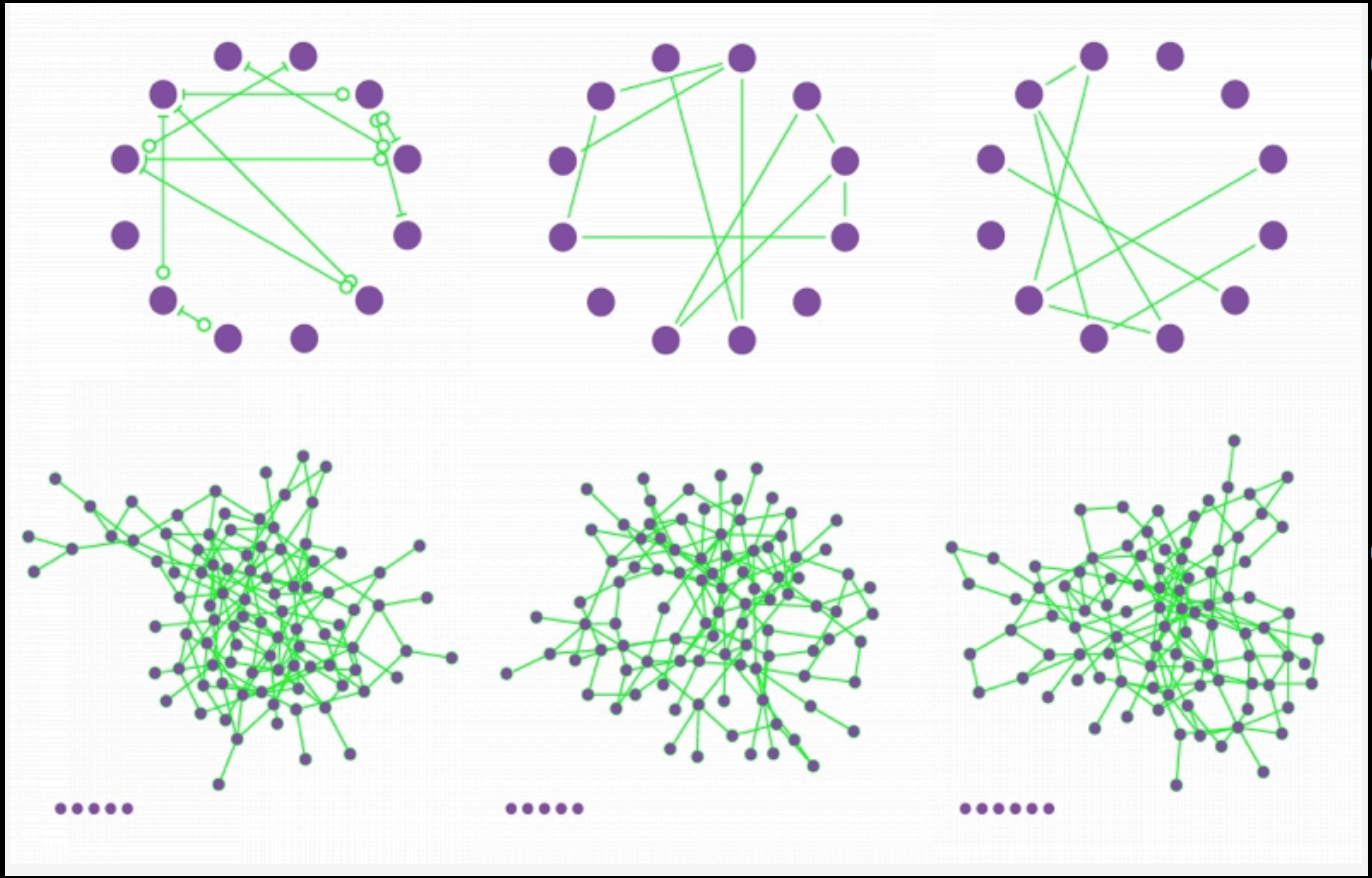
As (3.1) is a binomial distribution (BOX 3.3), the expected number of links in a random graph is

\

$$\langle L \rangle = \sum_{L=0}^{\frac{N(N-1)}{2}} L p_L = p\frac{N(N-1)}{2} \qquad (3.2)$$

LNMIIT

Random Networks



Same parameters
p=1/6 and N=12

p=0.03 and N=100

LNMIIT

Degree Distribution in a Random Network

- The probability that a node i has exactly k links:
    - $p_k = \binom{N-1}{k} p^k (1-p)^{N-1-k}$

This is a Binomial distribution with two parameters: N and p

Using Binomial distribution we can calculate $< k >$ **and** σ_k

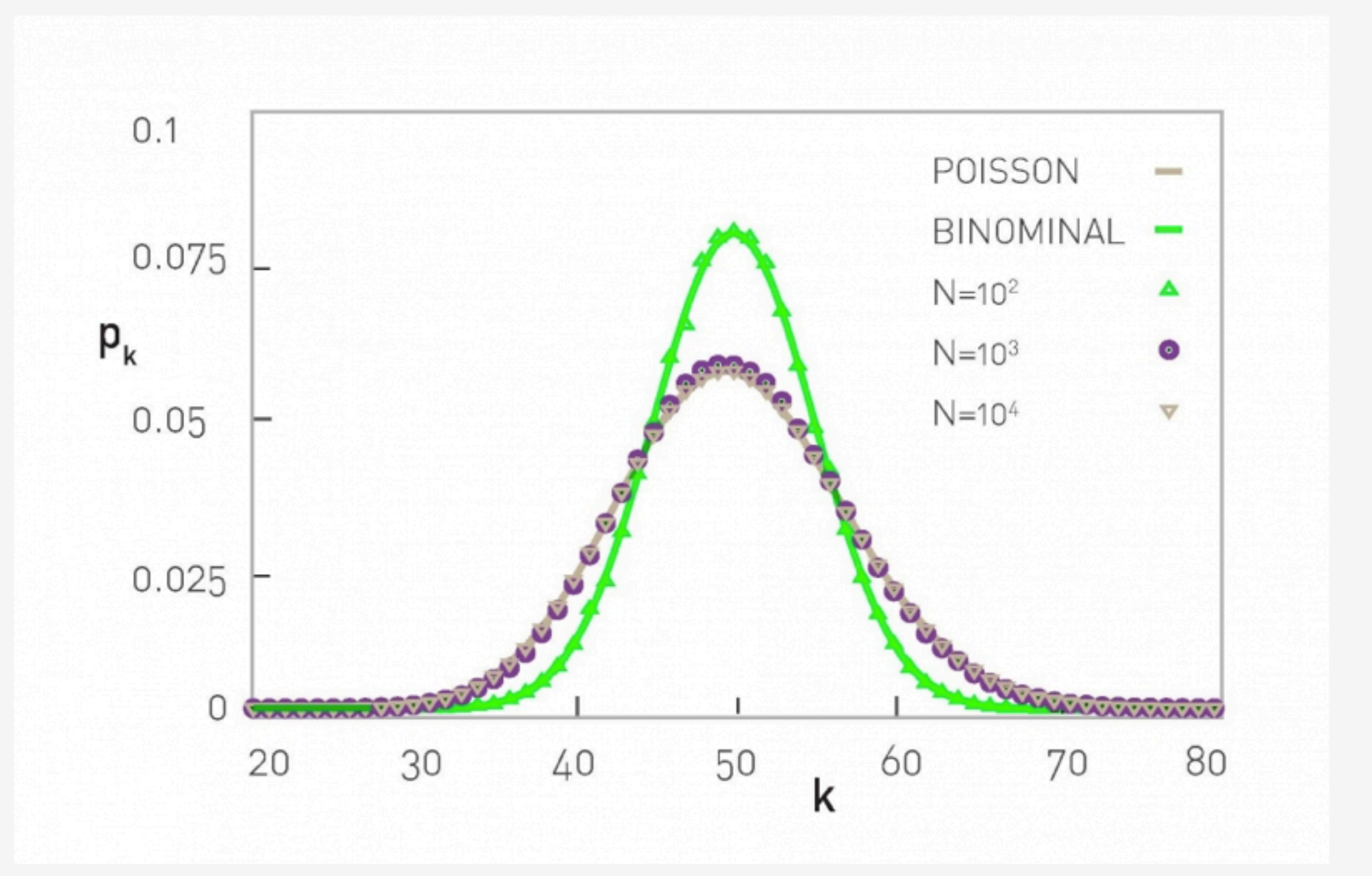
LNMIIT

Degree Distribution in a Random Network

- In most scenarios $\langle k \rangle \ll N$
- Hence the binomial distribution of the degrees can be approximated to poisson distribution with the following formula:

$$p_k = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$$

Binomial Vs Poisson when $\langle k \rangle \ll N$



- Both have a peak around ‹k›.
- If we increase p the network becomes denser, increasing ‹k› and moving the peak to the right.
- The width of the distribution (dispersion) is also controlled by p or ‹k›. The denser the network, the wider is the distribution, hence the larger are the differences in the degrees.

Real Networks are Not Poisson

- Degree of a node in a random network can vary between 0 and N-1
- How big are the differences between the node degrees in a random network?
- Can high degree nodes coexist with small degree nodes?

Real Networks are Not Poisson

- Sociologists estimate that a typical person knows about 1,000 individuals. So let us assume that $\langle k \rangle \approx 1{,}000$.
- Consider random society of $N \approx 7 \times 10^9$ of individuals.
- Expected to have kmax = 1,185 acquaintances (largest degree).
- Expected to have kmin = 816 acquaintances (smallest degree).
- kmax and min are not that different so is $\langle k \rangle$!
- The dispersion of a random network is $\sigma_k = \langle k \rangle^{1/2}$
- For $\langle k \rangle = 1{,}000$, $\sigma_k = 31.62$
- This means that the number of friends is between 968 and 1,032, a rather narrow window

LNMIIT

Real Networks are Not Poisson

- US president Franklin Delano Roosevelt's appointment book has about 22,000 names, individuals he met personally
- A study of the social network behind Facebook has documented numerous individuals with 5,000 Facebook friends, the maximum allowed by the social networking platform

Real Networks are Not Poisson

Box 3.4

Why are Hubs Missing?

We first note that the $1/k!$ term in (3.8) significantly decreases the chances of observing large degree nodes. Indeed, the Stirling approximation

$$k! \sim \left[\sqrt{2\pi k}\right] \left(\frac{k}{e}\right)^k$$

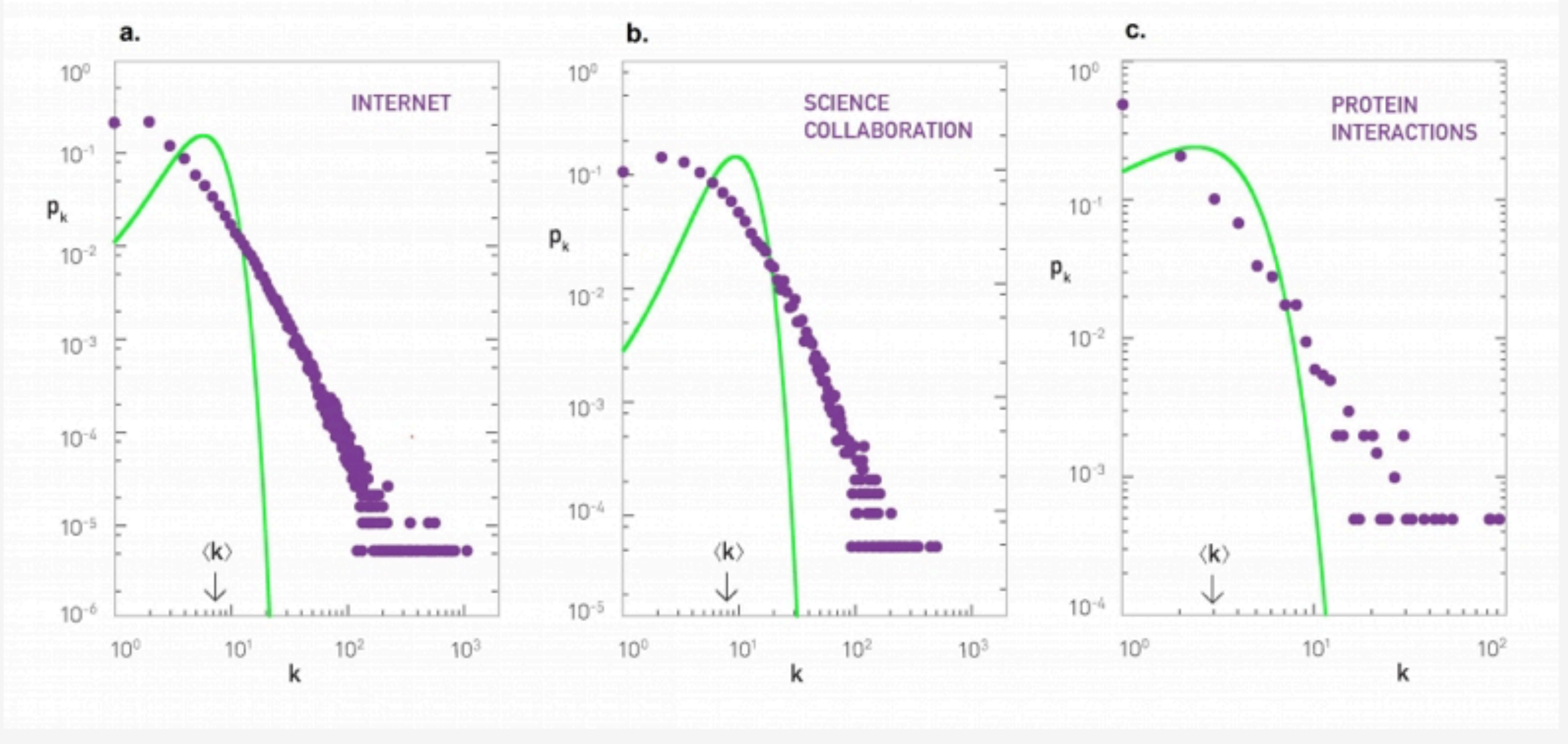
allows us rewrite (3.8) as

$$p_k = \frac{e^{-\langle k \rangle}}{\sqrt{2\pi k}} \left(\frac{e\langle k \rangle}{k}\right)^k \qquad (3.9)$$

For degrees $k > e\langle k \rangle$ the term in the parenthesis is smaller than one, hence for large k both k-dependent terms in (3.9), i.e. $1/\sqrt{k}$ and $(e\langle k \rangle/k)^k$ decrease rapidly with increasing k. Overall (3.9) predicts that in a random network the chance of observing a hub decreases faster than exponentially.

LNMIIT

Real Networks are Not Poisson



LNMIIT