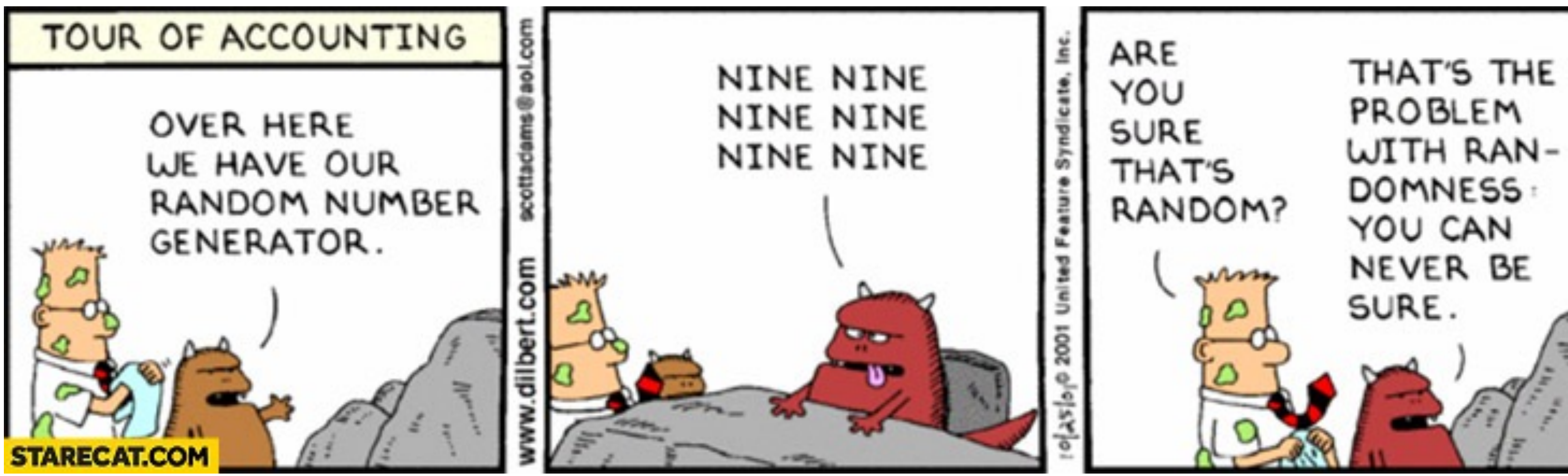
DS8104: Network Science

Class 4: Random Graphs



https://alexandergates.net https://uvads8104.github.io/

Random graphs

Hypothesis testing: compare the observed value to a **randomized null model**

Goal: We want to quantify what network properties are surprising and what is expected.

Random graph models allow comparisons between properties of real-world networks and random assumptions

Random graph history

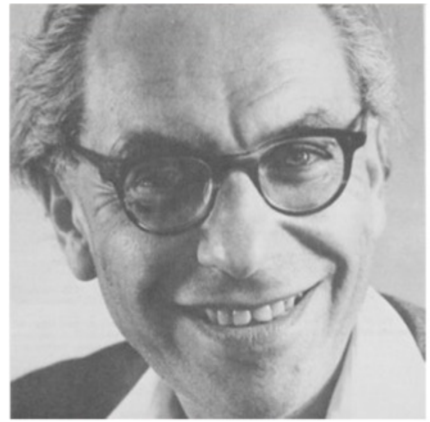
First used in: *JL Moreno, HH Jennings, 1938. Statistics of social configurations. Sociometry, 1(3/4):342–374.*

They manually generated (7) random graph instances to compare the observed network.

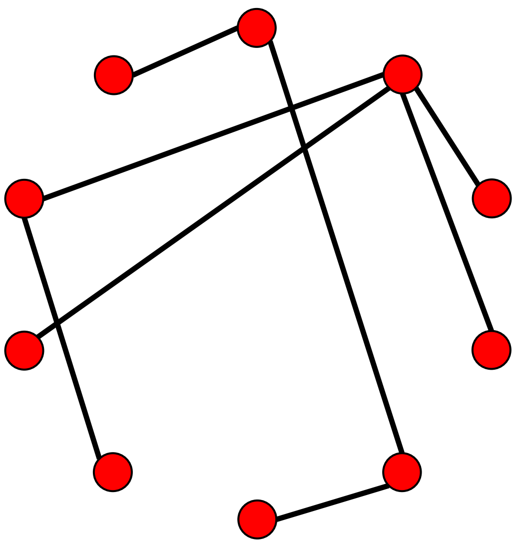
Much more detailed properties derived in:
R Solomonoff, A Rapoport, 1951. Connectivity of random nets. Bull. Math. Biophysics 13:107–117.

Erdos-Renyi model

Pál Erdös
(1913-1996)



Alfréd Rényi
(1921-1970)



p=1/6 N=10

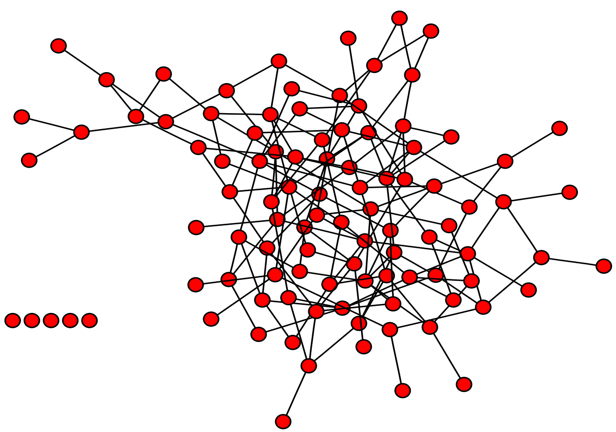
<k> ~ 1.5

G(N, L): *L* links are dropped randomly without replacement between *N* nodes

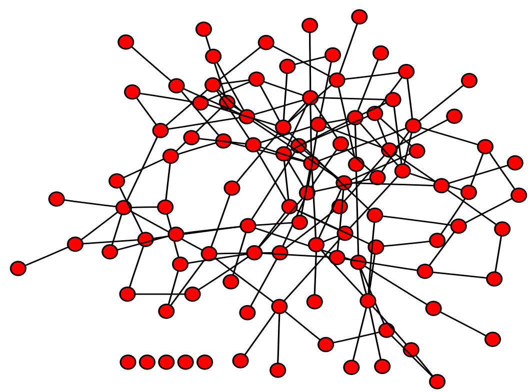
G(N, p): Each edge has probability *p* to be created between *N* nodes

Erdos-Renyi model

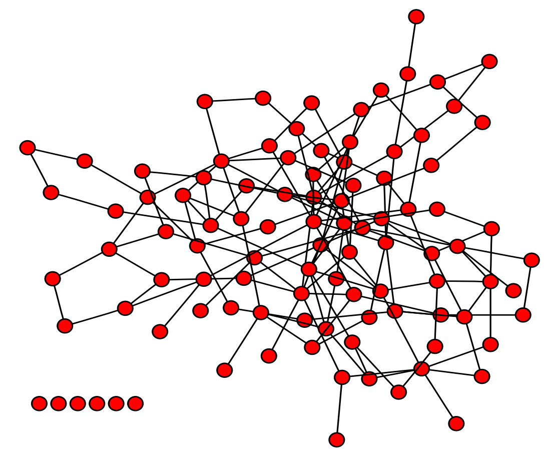
p=0.03
N=100



L=300



L=286



L=295

Erdos-Renyi model

P(*L*): the probability to have exactly *L* links in a network of *N* nodes and probability *p*:

The maximum number of links in a network of N nodes.

$$P(L) = \underbrace{\begin{pmatrix} \begin{pmatrix} N \\ 2 \end{pmatrix} \\ L \end{pmatrix}}_{} p^L (1-p)^{\frac{N(N-1)}{2}-L}$$

Number of different ways we can choose L links among all potential links.

Binomial distribution...

Binomial distributions

$$P(x) = \binom{N}{x} p^x (1-p)^{N-x}$$

$$< x >= Np$$

$$< x^2 >= p(1-p)N + p^2 N^2$$

$$\sigma_x = (< k^2 > - < k >^2)^{1/2} = [p(1-p)N]^{1/2}$$

http://keral2008.blogspot.com/2008/10/derivation-of-mean-and-variance-of.html

Erdos-Renyi model

P(*L*): the probability to have exactly *L* links in a network of *N* nodes and probability *p*:

$$P(L) = \binom{\binom{N}{2}}{L} p^L (1-p)^{\frac{N(N-1)}{2}-L}$$

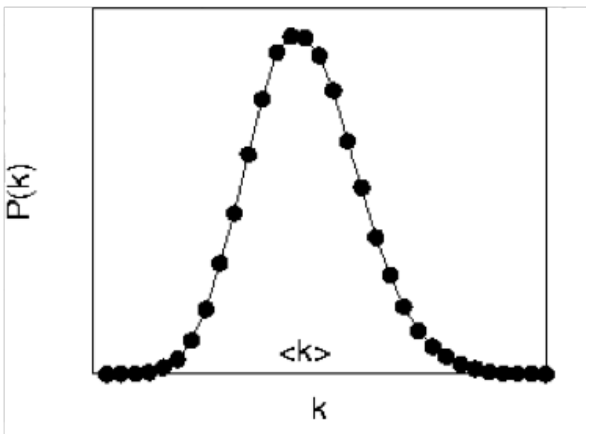
The average number of links <L> in a random graph:

$$<L> = \sum_{L=0}^{\frac{N(N-1)}{2}} LP(L) = p\frac{N(N-1)}{2}$$

$$<k> = 2L/N = p(N-1)$$

$$\sigma^2 = p(1-p)\frac{N(N-1)}{2}$$

Erdos-Renyi model



$$P(k) = \binom{N-1}{k} p^k (1-p)^{(N-1)-k}$$

Select *k* nodes from *N-1*

probability of having *k* edges

probability of missing *N-1*-kedges

$$<k> = p(N-1)$$

$$\sigma_k^2 = p(1-p)(N-1)$$

$$\frac{\sigma_k}{<k>} = \left[\frac{1-p}{p}\frac{1}{(N-1)}\right]^{1/2} \approx \frac{1}{(N-1)^{1/2}}$$

As the network size increases, the distribution becomes increasingly narrow—we are increasingly confident that the degree of a node is in the vicinity of <k>.

Erdos-Renyi model

$$P(k) = \binom{N-1}{k} p^k (1-p)^{(N-1)-k}$$

$$<k> = p(N-1)$$

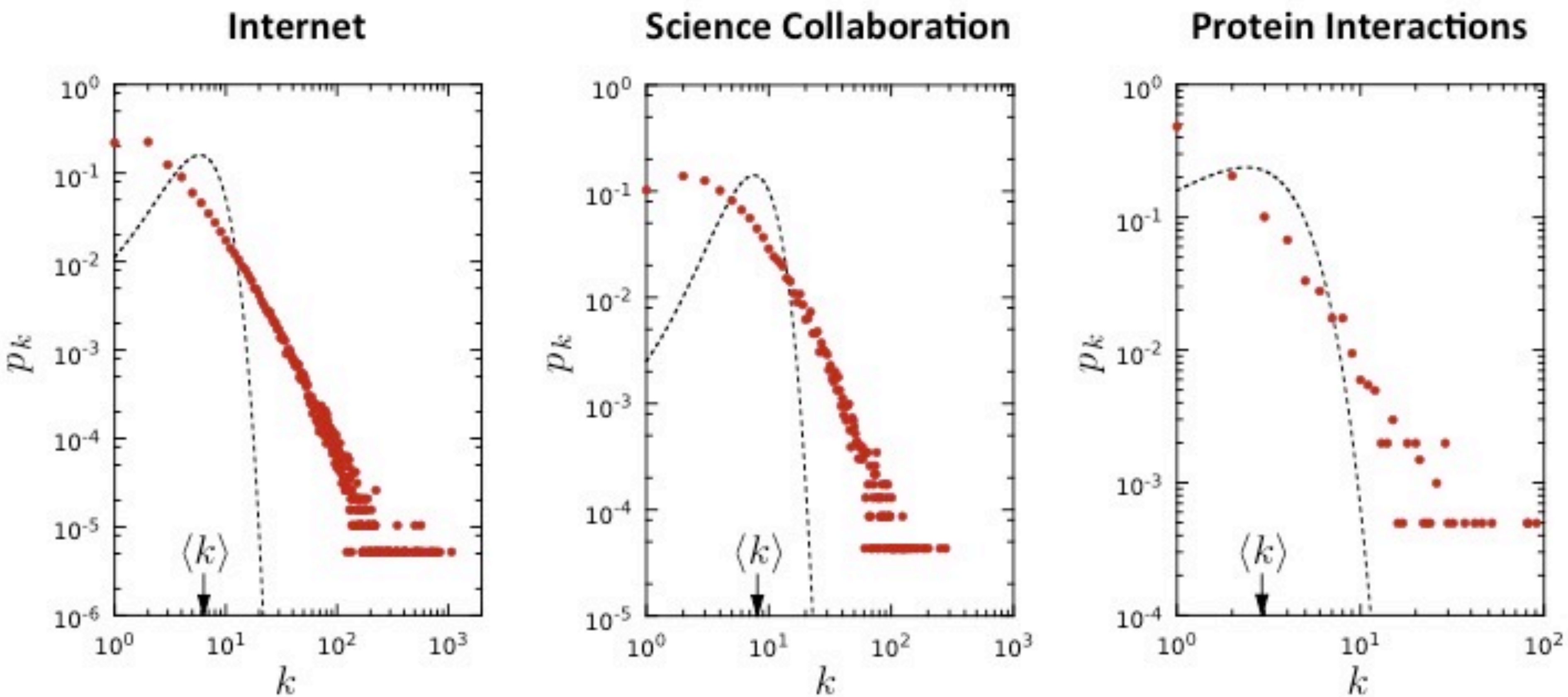
$$p = \frac{<k>}{(N-1)}$$

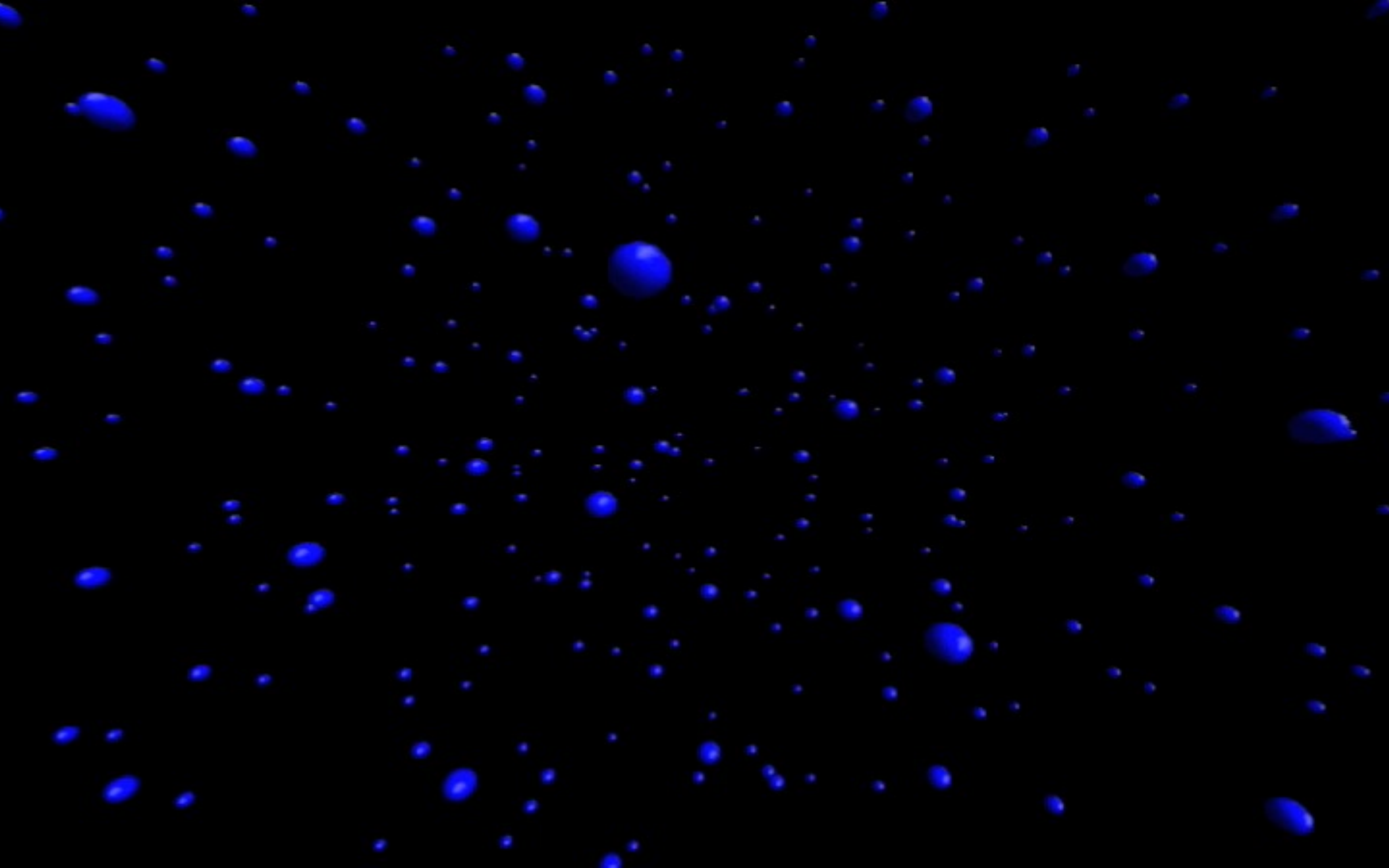
For large *N* and small *k (p)*, we arrive to the Poisson distribution:

$$P(k) = e^{-<k>} \frac{<k>^k}{k!}$$

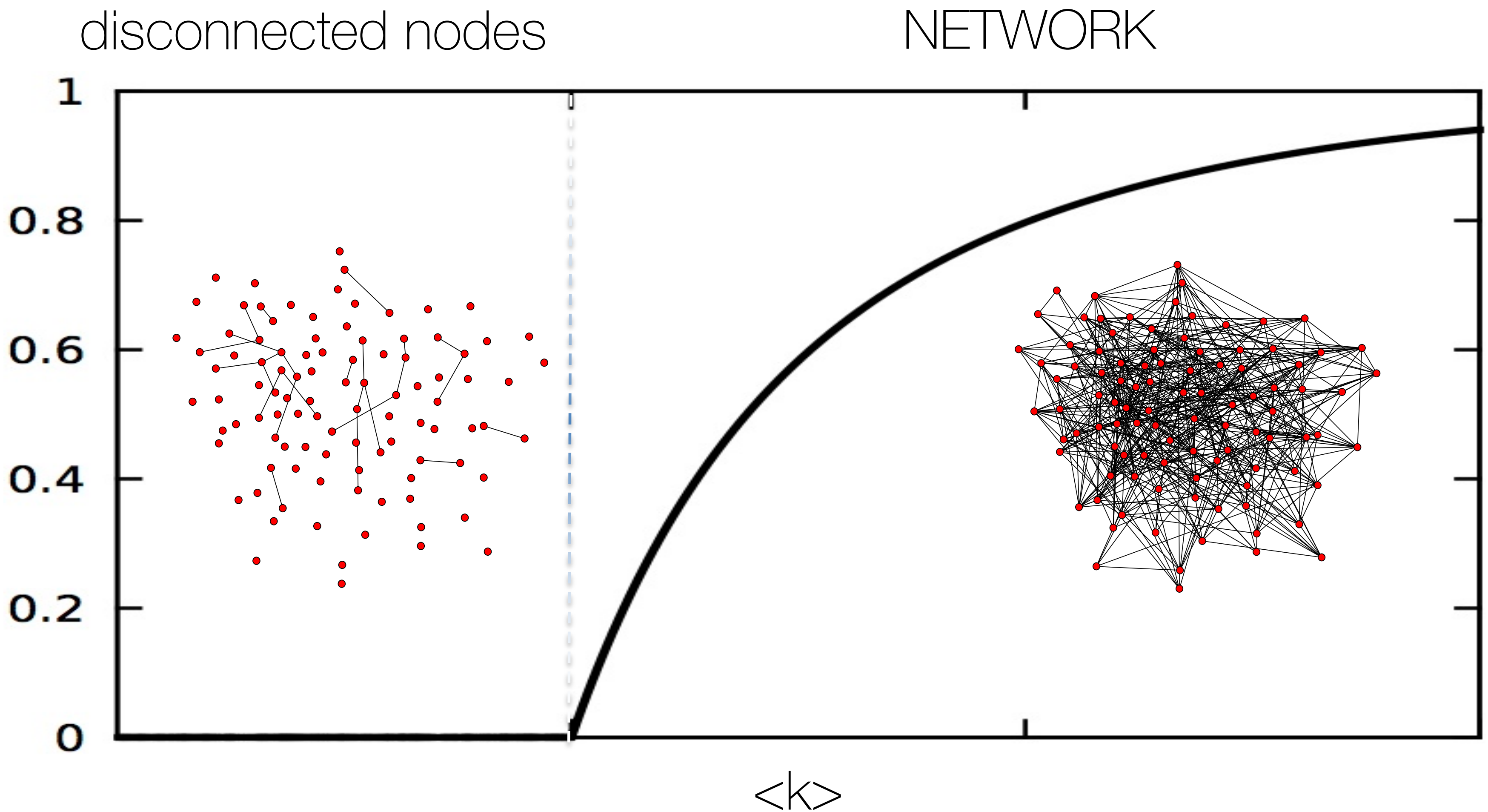
Real networks are not Poisson

$$P(k) = e^{-<k>} \frac{<k>^k}{k!}$$





Erdos-Renyi model



Erdos-Renyi model

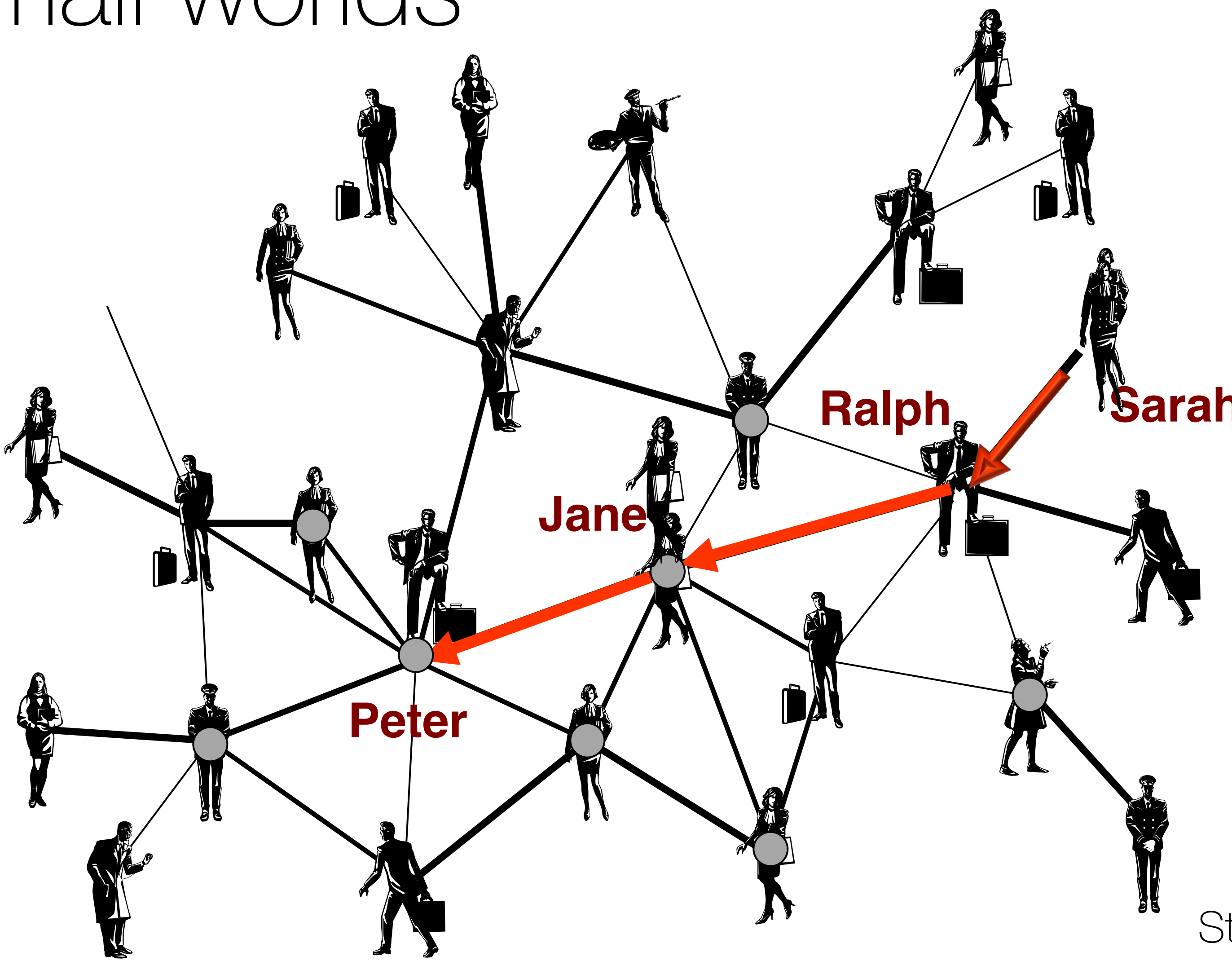
$<k_C>=1$ (Erdos and Renyi, 1959)

The fact that at least one link per node is necessary to have a giant component is not unexpected.

It is somewhat unexpected, however that one link is sufficient for the emergence of a giant component.

It is equally interesting that the emergence of the giant cluster is not gradual, but follows what physicists call a second order phase transition at $<k>=1$.

Small worlds



Stanley Milgram, 1967

The Small-World Problem

By Stanley Milgram

HOW TO TAKE PART IN THIS STUDY

1. ADD YOUR NAME TO THE ROSTER AT THE BOTTOM OF THIS SHEET, so that the next person who receives this letter will know who it came from.

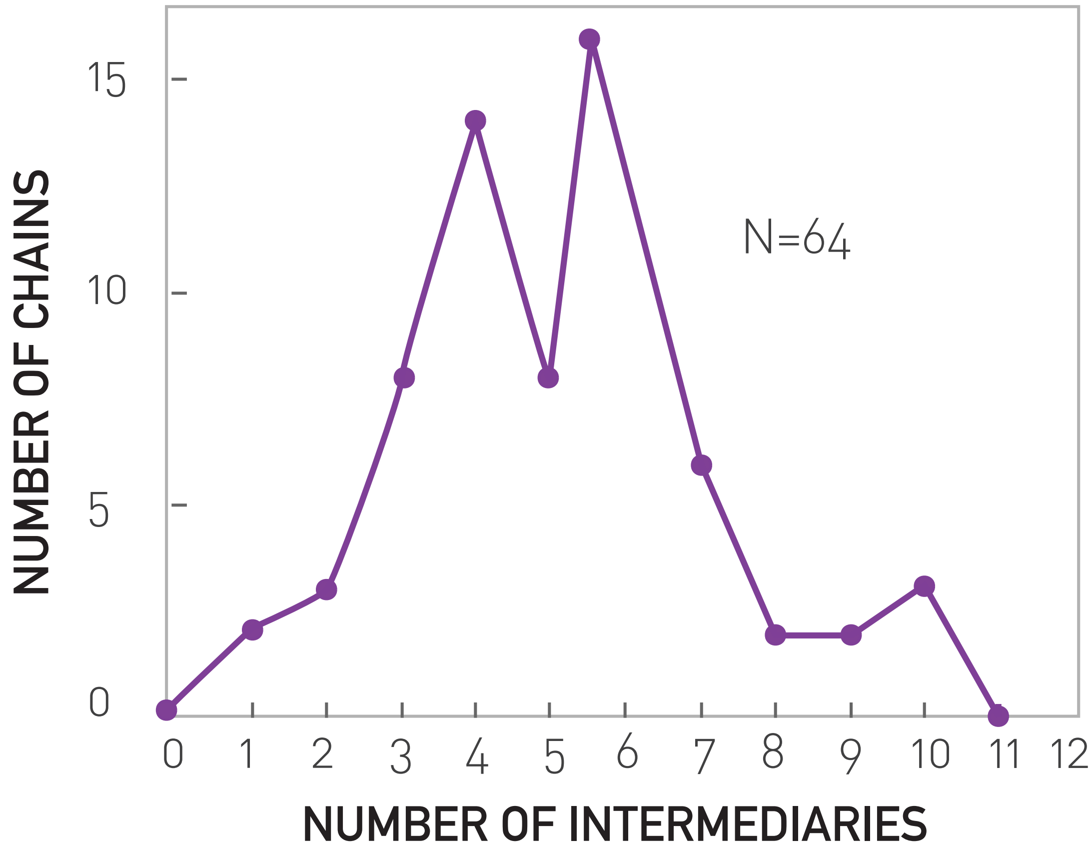
2. DETACH ONE POSTCARD. FILL IT AND RETURN IT TO HARVARD UNIVERSITY. No stamp is needed. The postcard is very important. It allows us to keep track of the progress of the folder as it moves toward the target person.

3. IF YOU KNOW THE TARGET PERSON ON A PERSONAL BASIS, MAIL THIS FOLDER DIRECTLY TO HIM (HER). Do this only if you have previously met the target person and know each other on a first name basis.

4. IF YOU DO NOT KNOW THE TARGET PERSON ON A PERSONAL BASIS, DO NOT TRY TO CONTACT HIM DIRECTLY. INSTEAD, MAIL THIS FOLDER (POST CARDS AND ALL) TO A PERSONAL ACQUAINTANCE WHO IS MORE LIKELY THAN YOU TO KNOW THE TARGET PERSON. You may send the folder to a friend, relative or acquaintance, but it must be someone you know on a first name basis.

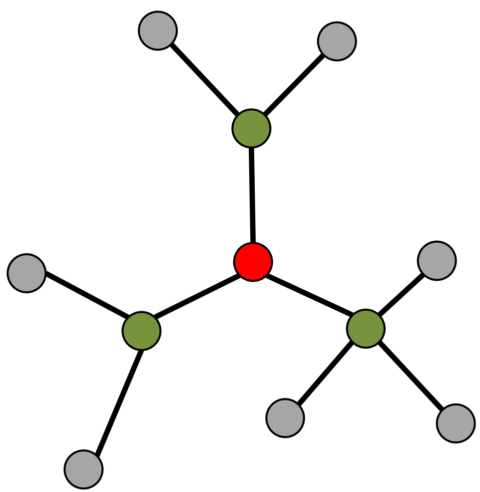
The Small-World Problem

By Stanley Milgram



Erdos-Renyi model

Random graphs tend to have a tree-like topology with almost constant node degrees.



$<k>$ nodes at distance one (d=1)

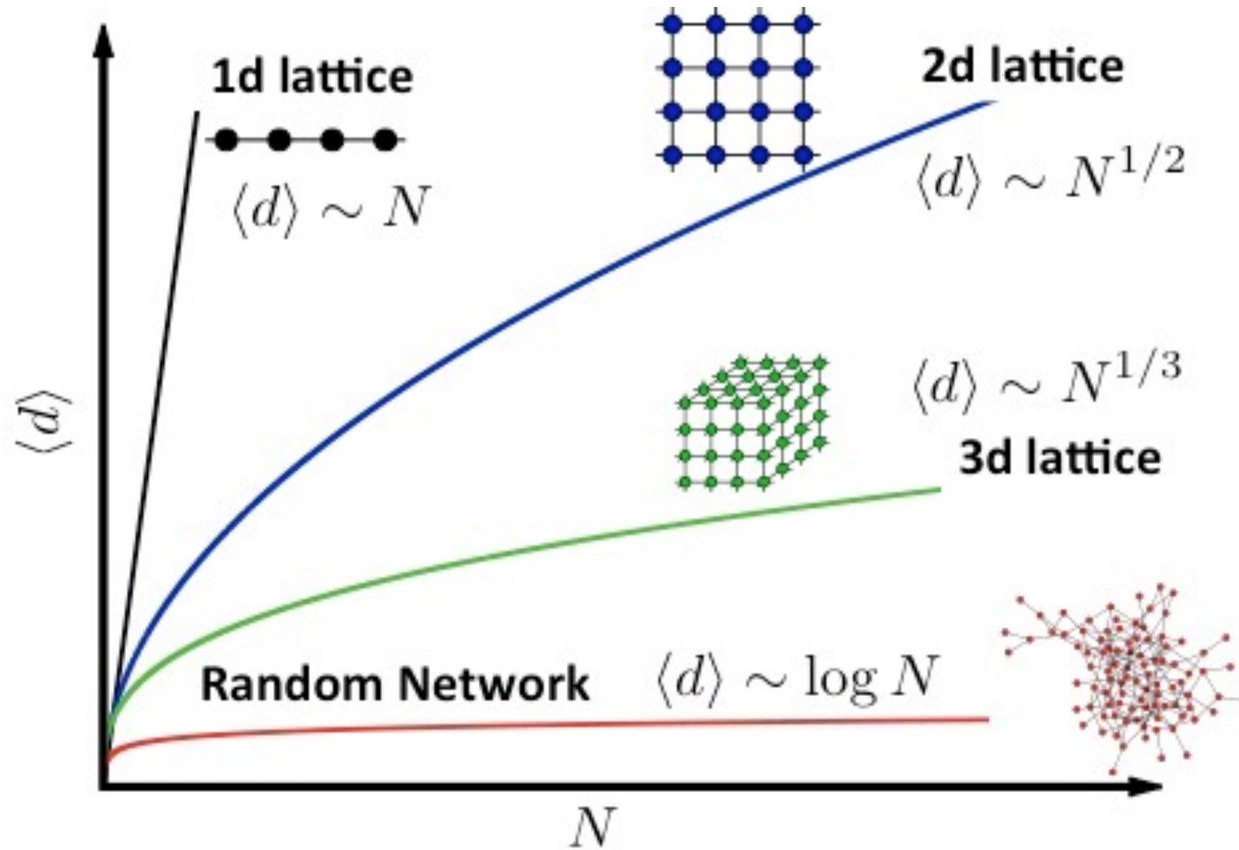
$<k>^2$ nodes at distance two (d=2)

...

$<k>^{dmax}$ nodes at distance d_{max}

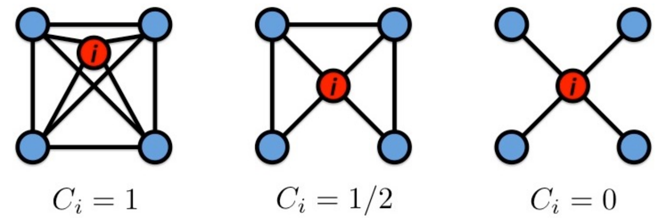
$$N = 1 + \langle k \rangle + \langle k \rangle^2 + \ldots + \langle k \rangle^{d_{\max}} = \frac{\langle k \rangle^{d_{\max}+1} - 1}{\langle k \rangle - 1} \approx \langle k \rangle^{d_{\max}} \Rightarrow d_{\max} = \frac{\log N}{\log \langle k \rangle}$$

Why are small worlds surprising?



Clustering coefficient

$$C_i \equiv \frac{2 < L_i >}{k_i(k_i - 1)}$$

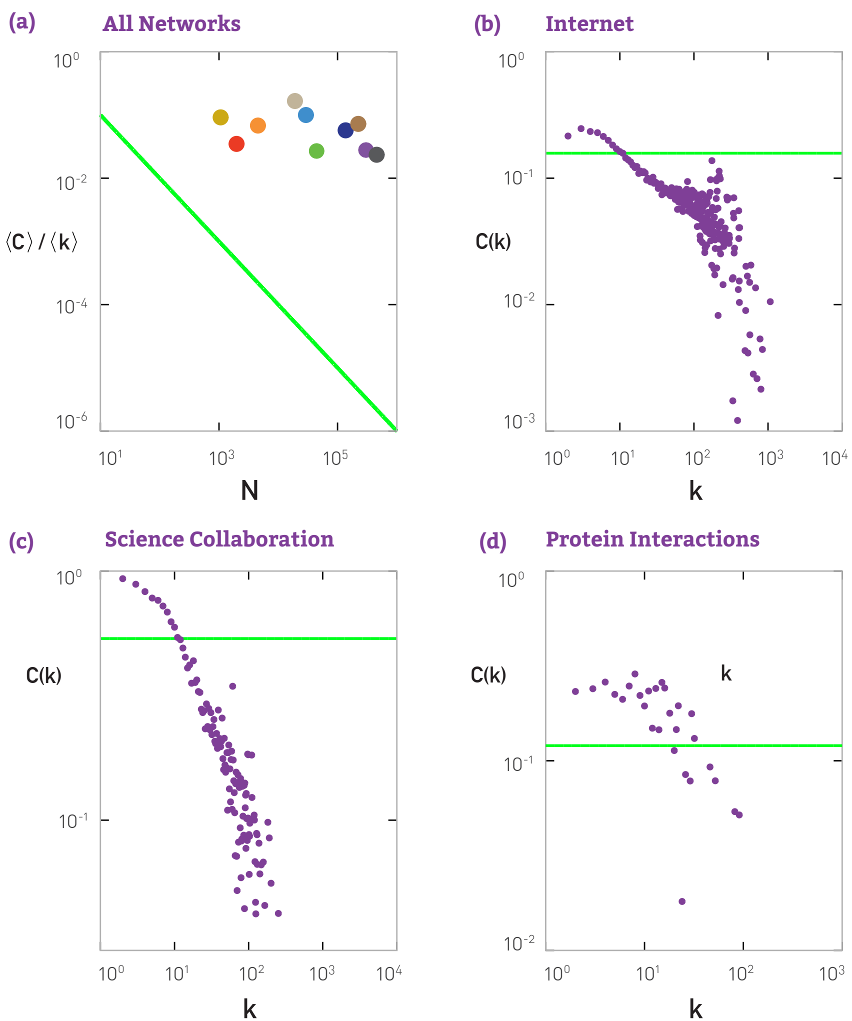


Edges are independent and have the same probability *p*

$$< L_i > \cong p \frac{k_i(k_i - 1)}{2} \quad \Rightarrow \quad C_i = \frac{2\langle L_i \rangle}{k_i(k_i - 1)} = p = \frac{\langle k \rangle}{N}.$$

The clustering coefficient of a random graph is small
(we don't expect many triangles!)

Clustering coefficient



$$C_i = \frac{2\langle L_i \rangle}{k_i(k_i - 1)} = p = \frac{\langle k \rangle}{N}.$$

C decreases with the system size N.

C is independent of a node's degree k.

Blogosphere

Collective dynamics of ‘small-world’ networks

Duncan J. Watts* **& Steven H. Strogatz**

Department of Theoretical and Applied Mechanics, Kimball Hall, Cornell University, Ithaca, New York 14853, USA