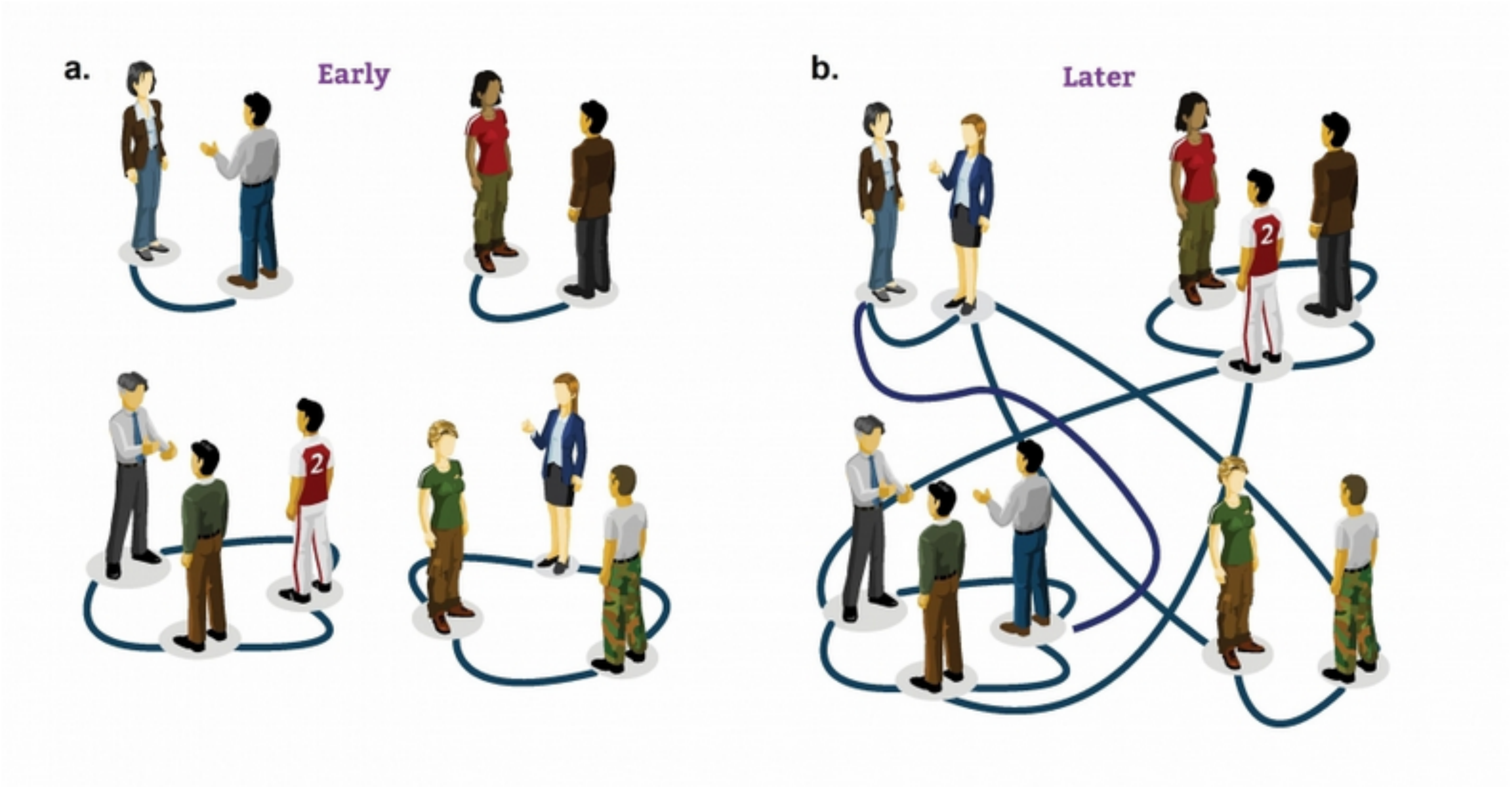
Random Networks

Topics

- What is a Random Network Model
- Degree Distribution (Binomial v. Poisson)
- Real networks are not Binomial or Poisson
- Evolution of a Random Network
- Small Worlds and Six Degrees of Separation
- Real Networks are Not Random

Random Networks

Cocktail Party: As individuals mingle, changing groups, an invisible network emerges that connects all of them into a single network



Constructing a Random Network (Random Graph) also called Erdős-Rényi Network

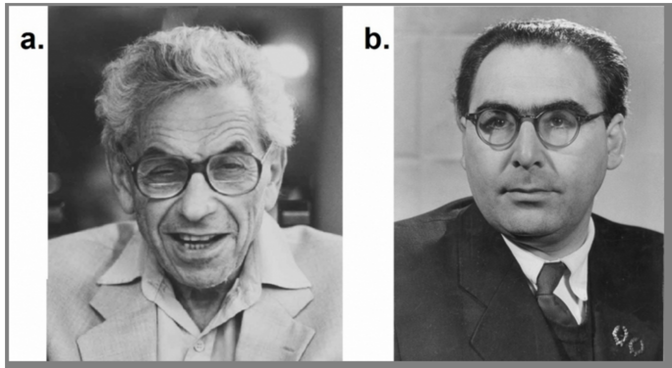
1. Start with N isolated nodes.
2. Select a node pair and generate a random number between 0 and 1.
3. If the number exceeds p, connect selected node pair with a link; otherwise leave them disconnected.
4. Repeat step 2 for each of the C_2^N= N!/((N-2)!2!) = N(N-1)/2 node pairs.

G(*N, L*) Model

- Given number of N nodes connected with given number of L randomly placed links.
- Average degree of a node <k> = 2L/N

G(*N, p*) Model

- Each pair of N nodes is connected with probability p.
- N is given, but number of links L is not given, and depends upon both N and p.
- Expected number of links in the entire graph <L> = pN(N-1)/2
- Maximum number of links in a random graph = L_{max} = C_2^N = N(N-1)/2
- Average degree per node <k> = 2<L>/N = p(N-1)
- G(*N, p*) is more realistic, since in real networks number of links rarely remains fixed.



Average Degree <k>

Total number of links possible in the entire network = $L_{max} = N(N-1)/2$

Probability of creating a link = p

Probability of not creating a link = 1-p

Number of links, say L

Number of non-connected node pairs = N(N-1)/2 - L

In how many ways can we select L links out of L_{max} links = $C_L^{N(N-1)/2}$

$P_L = C_L^{N(N-1)/2}\, p^L\, (1-p)^{N(N-1)/2}$

P_L is a binomial distribution, which gives

- Average number of links in a network: $\langle L \rangle = pN(N-1)/2$
- Average degree for a given node: $\langle k \rangle = 2\langle L \rangle/2 = p(N-1)$

Random Graph as a Binomial Distribution

Two outcomes: Node of pair is linked, or not linked.

Probability that a pair is linked = p (given).

Number of links for any given node, say = k

Probability of a pair not being linked = (1-p)

Number of nodes not being linked = N-1-k

- One node can only connect to N-1 remaining nodes
- With k links, the node has already reached out to k other nodes

Number of ways we can select k links out of N-1 possible links = C_k^{N-1}

Where, C_k^{N-1} = (N-1)! / ((N-1-k)! k!)

Thus we have a binomial distribution which has exactly two outcomes each with its fixed probability

Degree distribution for random graph can be expressed as:

$P_k = C_k^{N-1} p^k (1-p)^{N-1-k}$

Erdős - Rényi as a Poisson Distribution

Assuming that most real networks are sparse, i.e. <k> << N

Binomial can be approximated as a Poisson distribution for the region where average degree is very small compared to total number of nodes.

$$P_k = e^{-<k>} \frac{<k>^k}{k!}$$

Both Binomial and Poisson distributions describe the same quantity, hence they have similar properties.

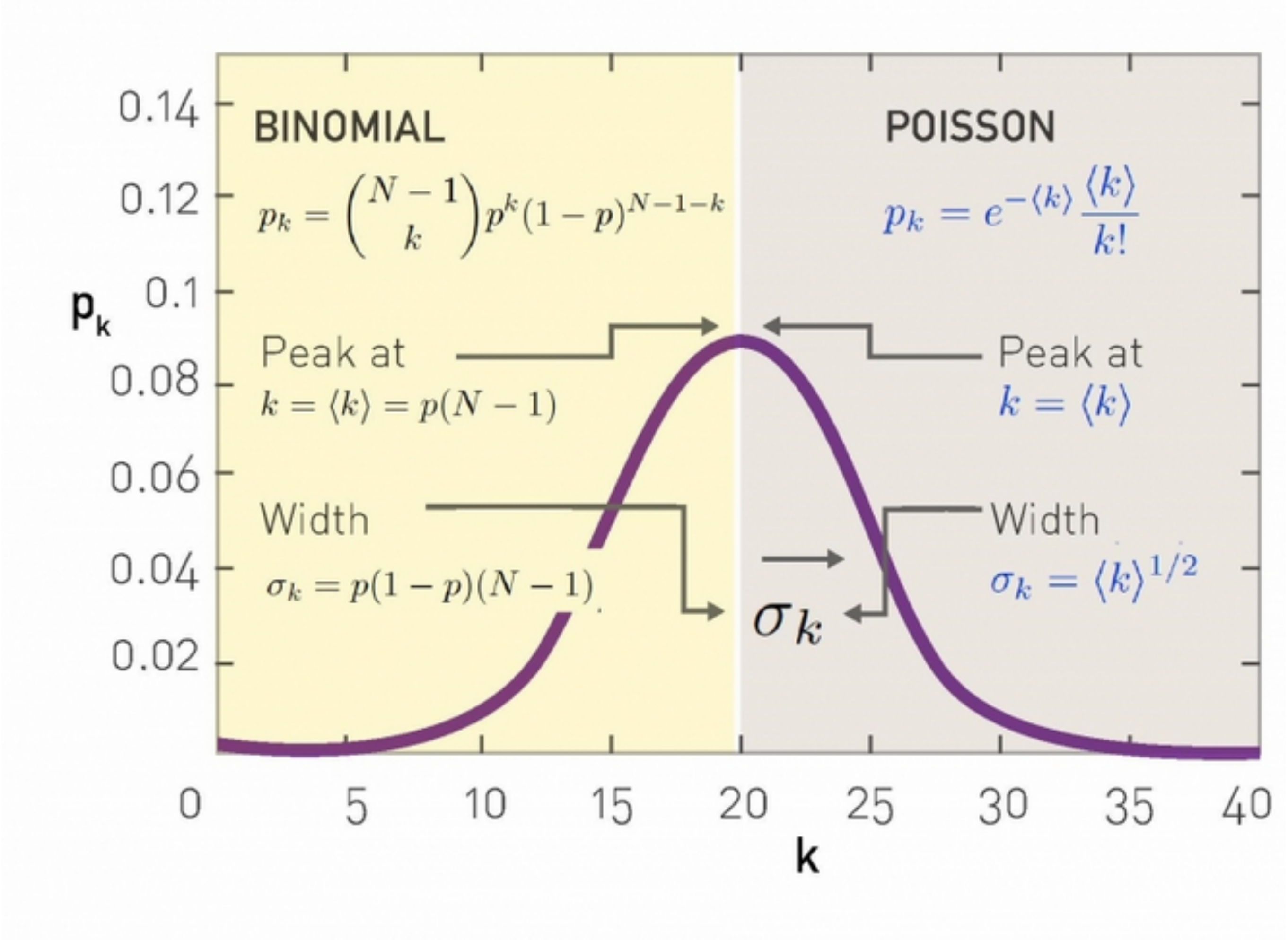
Erdős – Rényi: Degree Distribution

Binomial depends upon both p and N

Poisson depends only on <k>

They both peak at <k>

Poisson is much more simple, and hence more useful.



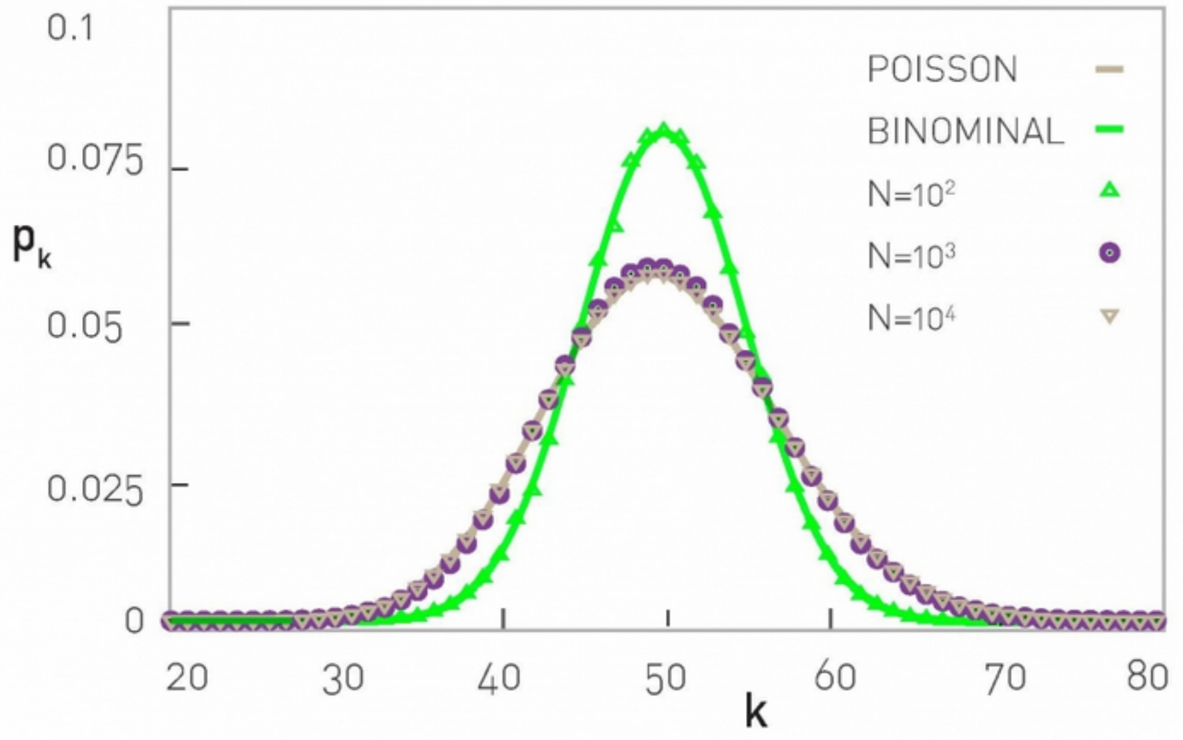
Erdős – Rényi: Binomial v. Poisson

For small random networks one needs to use Binomial.

- Green Binomial represents small size networks shown with green triangles well.

For large random networks one needs to use Poisson since large networks are independent of number of nodes.

- Gray Poisson represents large size networks shown with purple dots and gray triangles well.



Real networks are not Poisson

Poisson form significantly underestimates the number of high-degree nodes.

In a random network most nodes have comparable degrees.

- in the narrow vicinity of $<k>$

In real networks high-degree nodes co-exist with small-degree nodes.

Following networks are not random:

- Internet: few sites are extremely popular such as Google, Facebook.
- Social networks: celebrities are much more popular than any one else.

Emergence of a Giant Component

Critical Point: <k>=1 separates the two regimes:

- Subcritical: For <k> < 1 there is no one giant component
- Supercritical: For <k> > 1 there is a giant component, and other smaller components. Resembles real networks.

Connected regime: <k> > ln N.

- Single connected network: for sufficiently large p, the giant component absorbs all nodes and components (but the network is still relatively sparse)

Real Networks

All networks here have a giant component.

- They also have <k> >1

Giant component absorbs all nodes and components when <k> > Ln N

Most networks have <k> < ln N

Network	N	L	‹K›	lnN
Internet	192,244	609,066	6.34	12.17
Power Grid	4,941	6,594	2.67	8.51
Science Collaboration	23,133	94,437	8.08	10.05
Actor Network	702,388	29,397,908	83.71	13.46
Protein Interactions	2,018	2,930	[illegible]	7.61

<k> > LnN
Number of links explodes!


Random Graphs
Emergence of the Giant Component
Video
Click to play

Small Worlds

Small Worlds phenomenon also known as six degrees of separation.

Frigyes Karinthys (1929):

- Short story entitled ‘Lanczszemek’ (Chains): links a worker in Ford factory with himself

John Guare (1938):

- Coins the phrase ‘six degrees of separation’

Stanley Milgram (1967):

- Stock broker in Boston and student in Sharon, Massachusetts as targets
- Residents of Wichita and Omaha as respondents
- 64 of 296 letters made it back
- 5.2 median number of intermediaries

WWW (1999):

- Separation between two randomly chosen documents is 19

Facebook (2011):

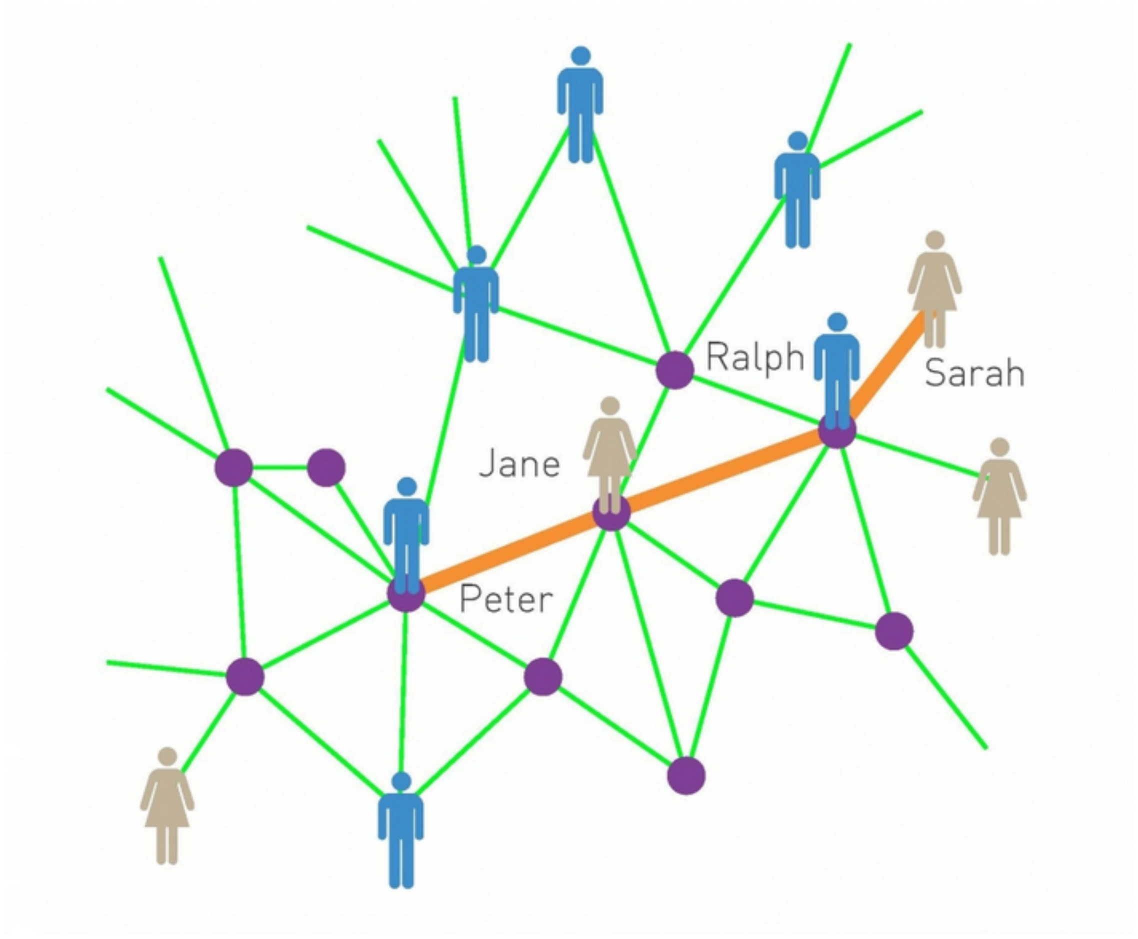
- 721 million active users
- 68 billion symmetric friendship links
- average distance of 4.74 between users

Six Degrees of Separation

Two individuals anywhere in the world can be connected through a chain of six or fewer acquaintances.

Sarah does not know Peter, but she knows Ralph, and Ralph knows Jane, and Jane knows Peter.

Distance between any two nodes in a network is unexpectedly small.



Six Degrees of Separation

Link: https://www.youtube.com/watch?v=TcxZSmzPw8k

Average Distance

$$\langle d \rangle \propto \frac{Ln\ N}{Ln \langle k \rangle}$$

Clustering coefficient C_i measures the density of links between node i's neighbors.

Overall network clustering coefficient $\langle C \rangle$= Local Clustering Coefficient of a Node $\langle C_i \rangle = p = \frac{\langle k \rangle}{N}$

Random networks, it seems, explain small-world phenomenon (but fail in explaining hubs).

Watts - Strogatz Model

Regular lattice.

Random.

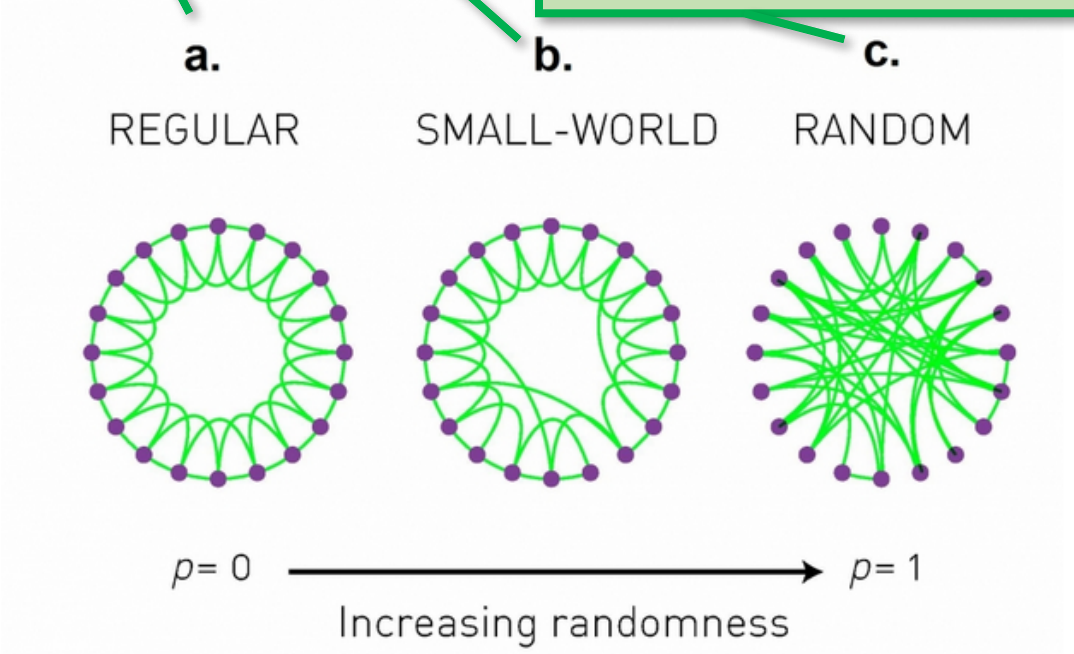
Watts - Strogatz:

- Also called **small-world model**
- Interpolates between regular lattice and random network
- Randomly assigns few more links to a regular lattice
- Predicts Poisson-like bounded degree distribution
- Doesn't explain hubs

High clustering, but lacks small world phenomenon.

High clustering, and also displays small world phenomenon.

Shows small world phenomenon but has low clustering.



Ten Reference Networks

LnN/Ln<k> attempts to predict average distance

Prediction by LnN/Ln<k> needs some adjustments

Network	N	L	‹k›	‹d›	d_{max}	ln*N*/ln‹*k*›
Internet	192,244	609,066	6.34	6.98	26	6.58
WWW	325,729	1,497,134	4.60	11.27	93	8.31
Power Grid	4,941	6,594	2.67	18.99	46	8.66
Mobile-Phone Calls	36,595	91,826	2.51	11.72	39	11.42
Email	57,194	103,731	1.81	5.88	18	18.4
Science Collaboration	23,133	93,437	8.08	5.35	15	4.81
Actor Network	702,388	29,397,908	83.71	3.91	14	3.04
Citation Network	449,673	4,707,958	10.43	11.21	42	5.55
E. Coli Metabolism	1,039	5,802	5.58	2.98	8	4.04
Protein Interactions	2,018	2,930	2.90	5.61	14	7.14

Real networks are not Random?

Small world phenomenon is reasonably explained by ln N / Ln <k>

Connectedness: real networks do not follow <k> > Ln N; still they are connected.

Real networks have hubs.

Clustering:

- Random networks - clustering is independent of k, and depends on <k> and also on N
- Real networks – clustering is independent of N, and depends on k

Questions? Thoughts? Visit the course's online discussion forum.