RANDOM GRAPHS MODELS

WHY USING RANDOM GRAPH MODELS

- Several good reasons:
  - Study some properties in a "controlled environment"
    - How does property X behaves when increasing property Y ?
  - Compare an observed network with a randomized version
    - Is observed property X "exceptional", or any similar network with same property Y and Z ?
  - Explain a given phenomenon
    - Such simple mechanism can reproduce property X and Y
  - Generate synthetic datasets
    - Testing an algorithm on 100 variations of the same network

WHY USING RANDOM GRAPH MODELS

- Deterministic models
  - Repeated regular patterns (lattices)
- Generative models
  - The probability of an edge between 2 nodes depend on their properties
    - Erdos Renyi, Configuration model, etc.
- Mechanistic models
  - The network is created following a mechanism, a set of rules
    - Preferential attachment, Forest fire, etc.

Fundamental network models

Central quantities in network analysis

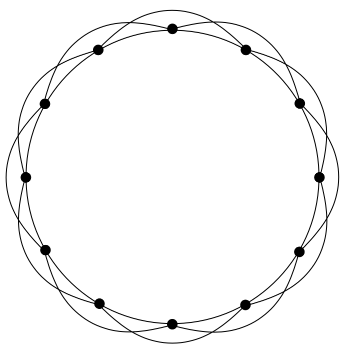
- Degree distribution: $P(k)$
- Clustering coefficient: C
- Average path length: $<d>$

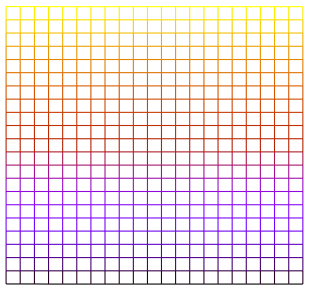
Network	Degree distribution	Path length	Clustering coefficient
Real world networks	broad	short	large

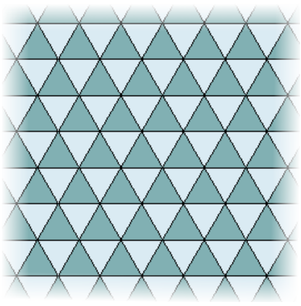
Regular lattices

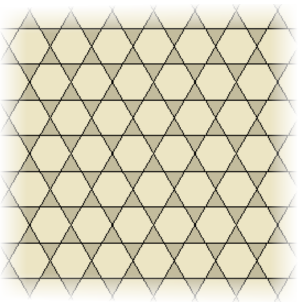
- Graphs where each node has the **same degree** k
- Translational symmetry in n directions

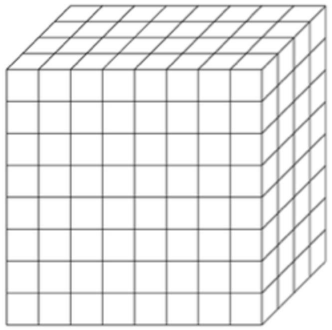
1D | 2D lattices | 3D lattices



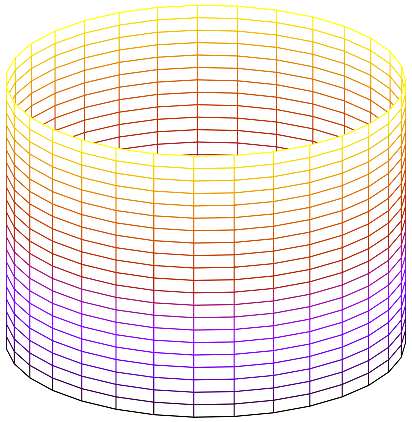


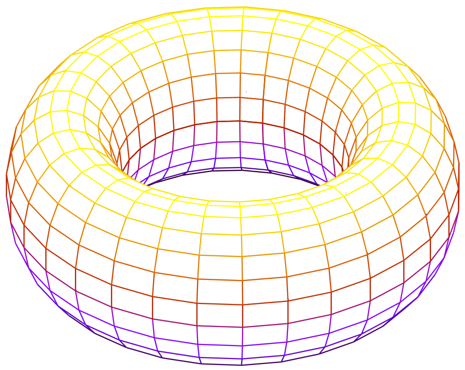






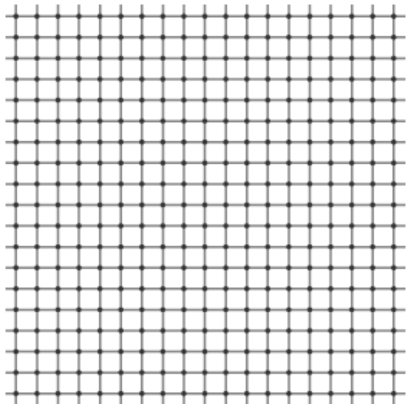
k=4 | k=4 | k=6 | k=4 | k=6



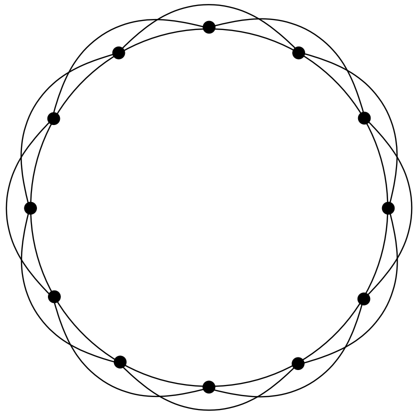


Regular lattices

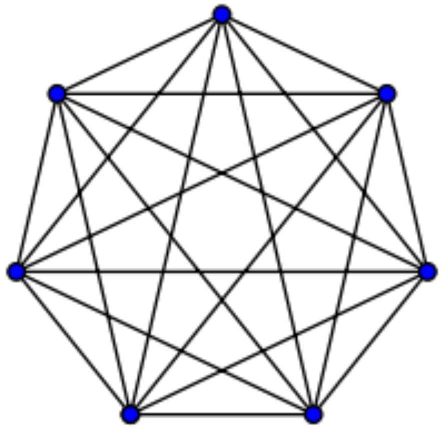
Clustering coefficient



C=0



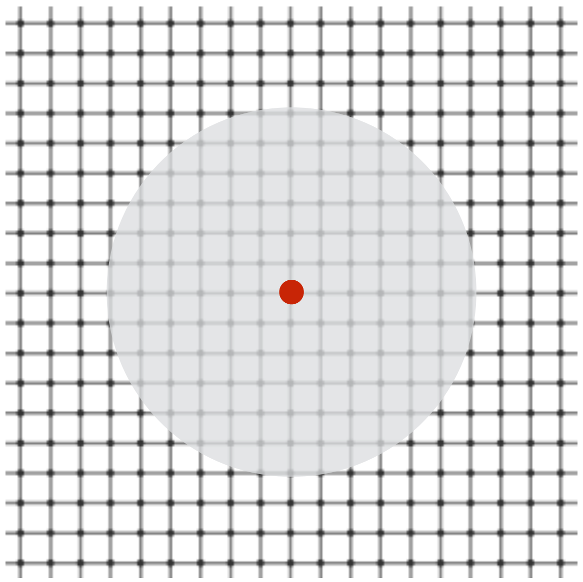
C=3/6



C=1

- Clustering coefficient depends on the structure (can be large or not)
- It is constant for each node

Path length



- Average path length grows quickly with n when k << n
- In a *large* graph with *realistic* average degrees, will be large

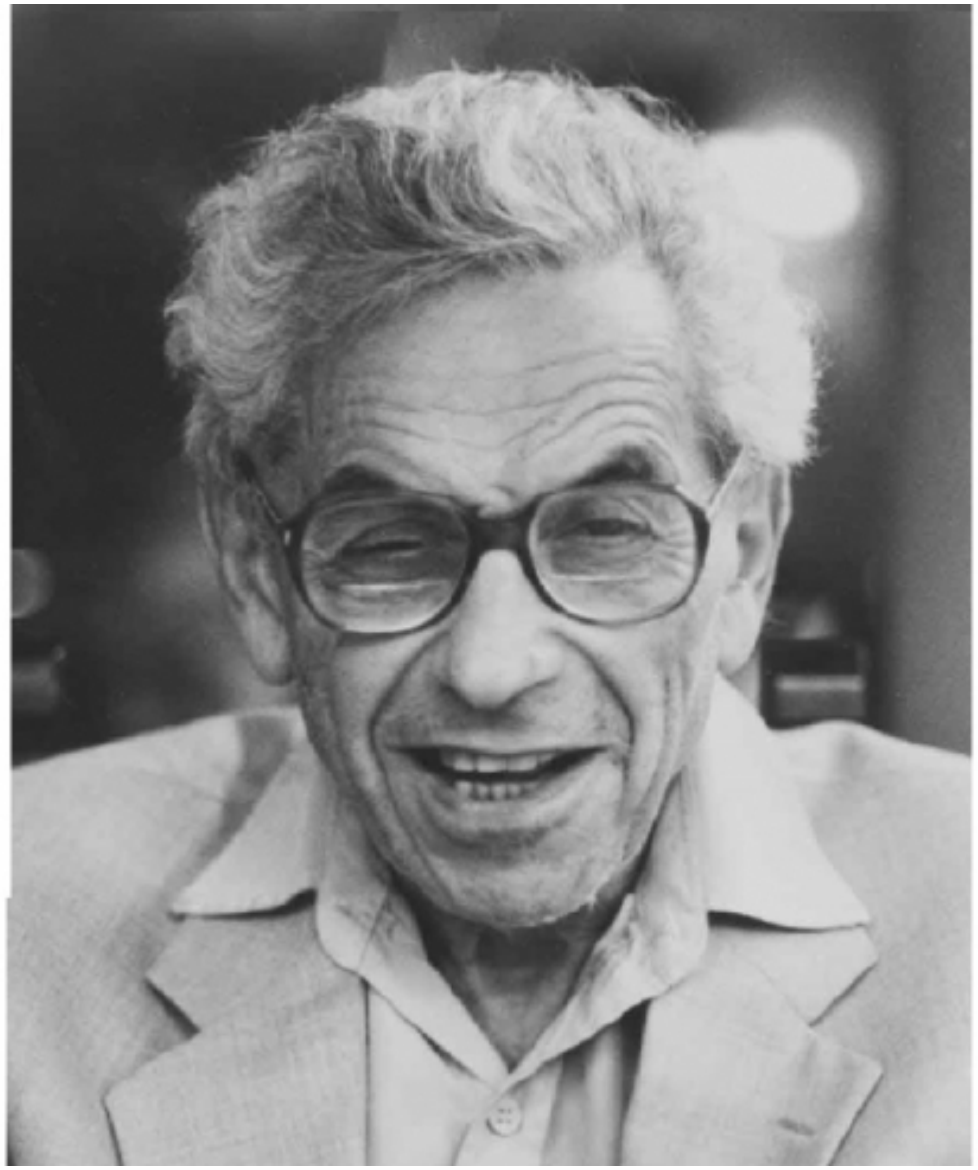
Regular lattices

Network	Degree distribution	Path length	Clustering coefficient
Real world networks	broad	short	large
Regular lattices	constant	long	can be large

PROBABILISTIC MODEL

The Erdős-Rényi Random Graph model (ER)

Random Graphs



Pál Erdős
(1913-1996)

Alfréd Rényi
(1921-1970)

"If we do not know anything else than the number *n* of nodes and the number *L* of links, the simplest thing to do is to put the links at random (no correlations)"

P. Erdős and A. Rényi. On random graphs, I. Publicationes Mathematicae (Debrecen), 6:290-297, 1959.
P. Erdős and A. Rényi. On the evolution of random graphs. Publ. Math. Inst. Hung. Acad. Sci., 5:17-61, 1960.

ER Random Graphs

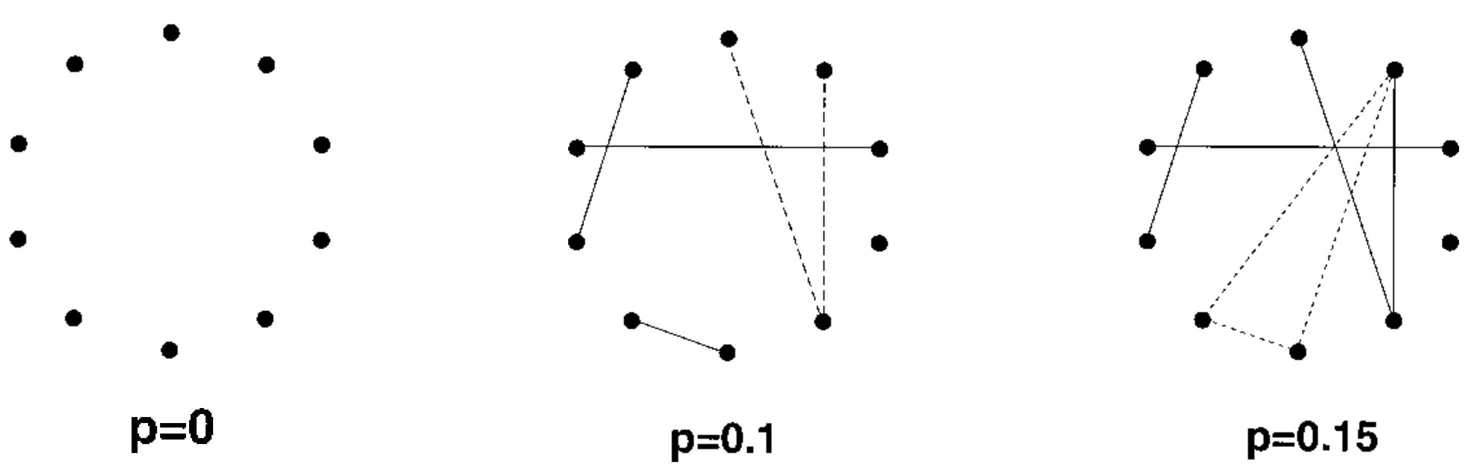
Erdős-Rényi model: simple way to generate random graphs

- The $G(n,L)$ definition
  1. Take n disconnected nodes
  2. Add L edges uniformly at random

  Alternatively:
  - pick uniformly randomly a graph from the set of all graphs with n nodes and L links

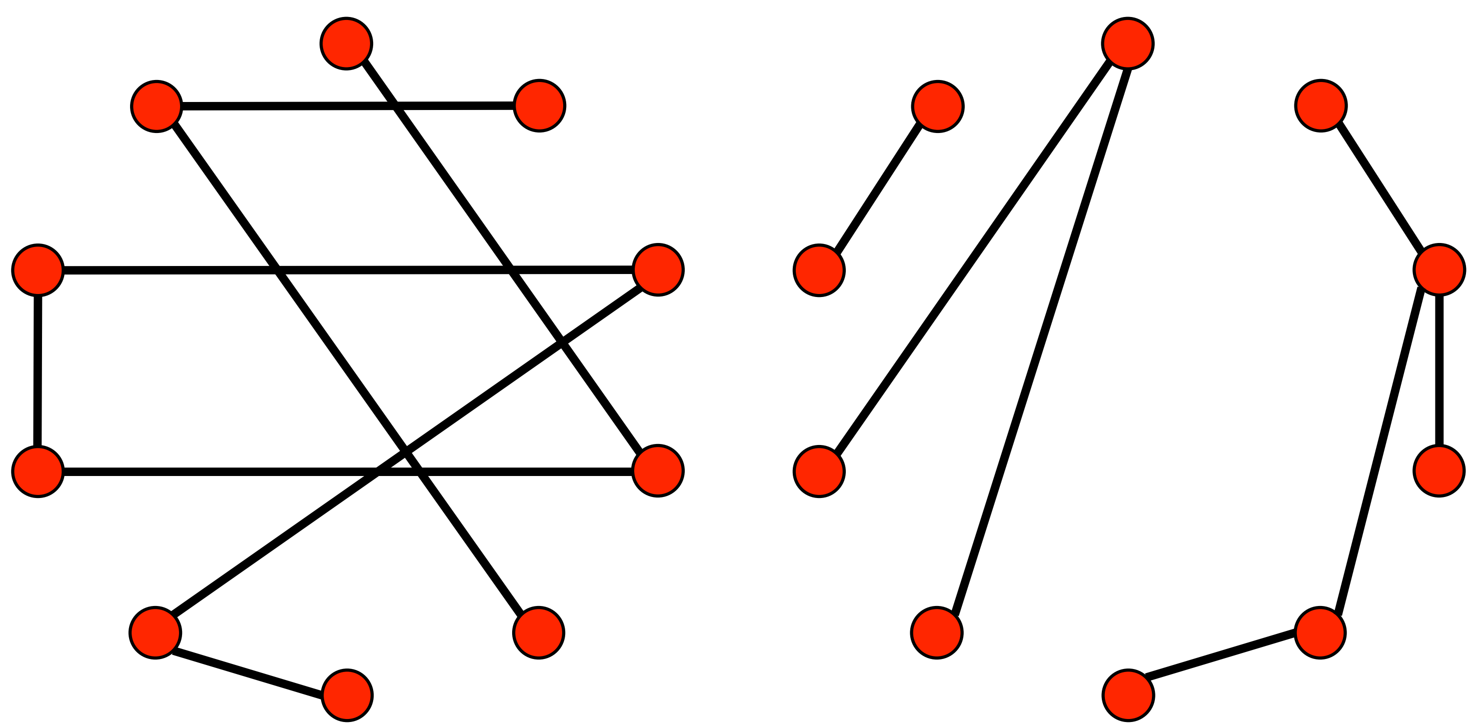
- The $G(n,p)$ definition
  1. Take n disconnected nodes
  2. Add an edge between any of the nodes independently with probability p

  Alternatively:
  - pick with probability $p^L (1-p)^{\binom{n}{2}-L}$ a network from the set of all networks with size n



Random Graphs

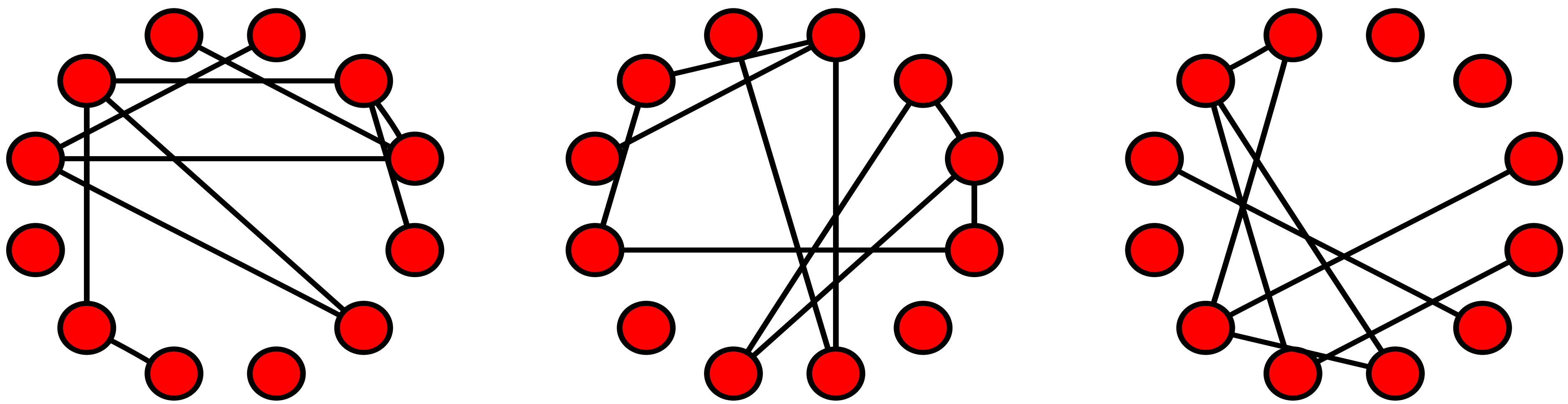
In the G(n,p) variant, the number of edges may vary



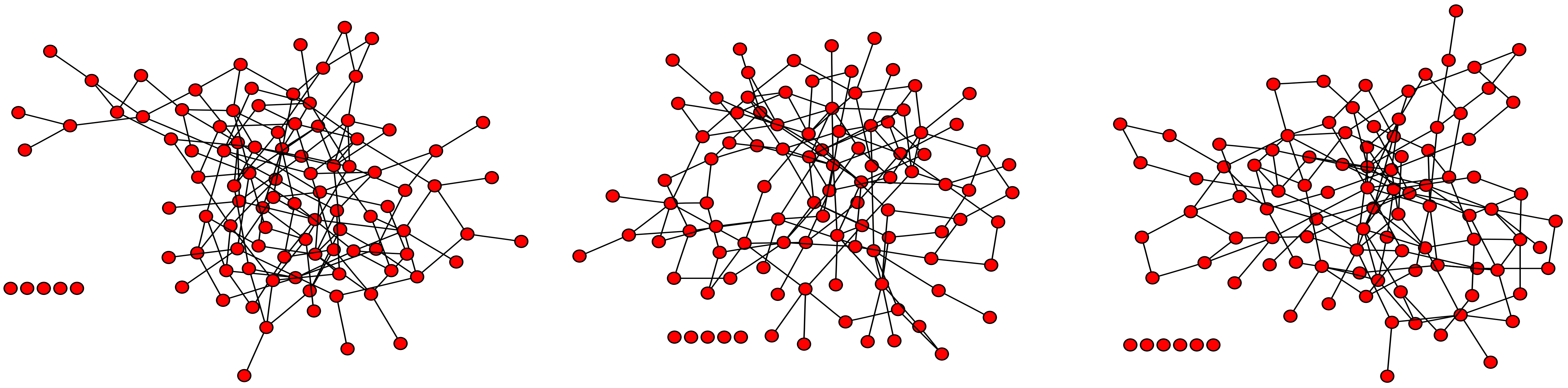
n=10
p=1/6

ER Random Graphs

p=1/6
N=12



p=0.03
N=100



Random Graphs

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

Binomial distribution:

Discrete probability distribution of the number of successes(**x**) in a sequence of **N** independent experiments, with success probability **p**

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

Reminder: Binomial coefficient:

Number of ways, disregarding order, that **k** objects can be chosen from among **n** objects

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Random Graphs

Binomial distribution

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

N: Number of experiments

Pairs of nodes

$$N = \binom{n}{2} = \frac{n(n-1)}{2}$$

P(L): probability to have exactly **L** links in a network of **n** nodes (with **p** the probability to have an edge)

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

Random Graphs

Properties of Binomial distribution

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

Mean $$< x > = pN$$

variance $$\sigma^2 = Np(1-p)$$

Random Graphs

Expected number of links <L>

$$< L > = pN = p\frac{n(n-1)}{2}$$

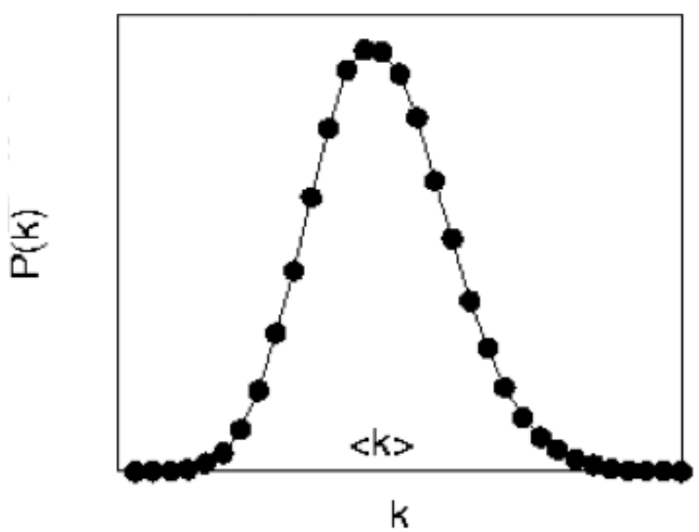
Expected average degree <k>

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

Variance

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

Degree distribution - Random Graphs



For each node,
independent probabilities to take each neighbor
=> Binomial distribution

P(k): probability to have exactly **k** links among **n** (total # of nodes), with **p** the (overall) probability to have an edge

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

Characteristics:

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

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

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

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>.

Slide from CCNR course, A. L. Barabási (2012)

Degree distribution - Random Graphs

For large **n** and small **k** (p,L), we can approximate the degree distribution using a poisson distribution of parameter (mean) $\lambda = < k >$

Poisson distribution

$$P(K) = \frac{\lambda^K e^{-\lambda}}{K!}$$

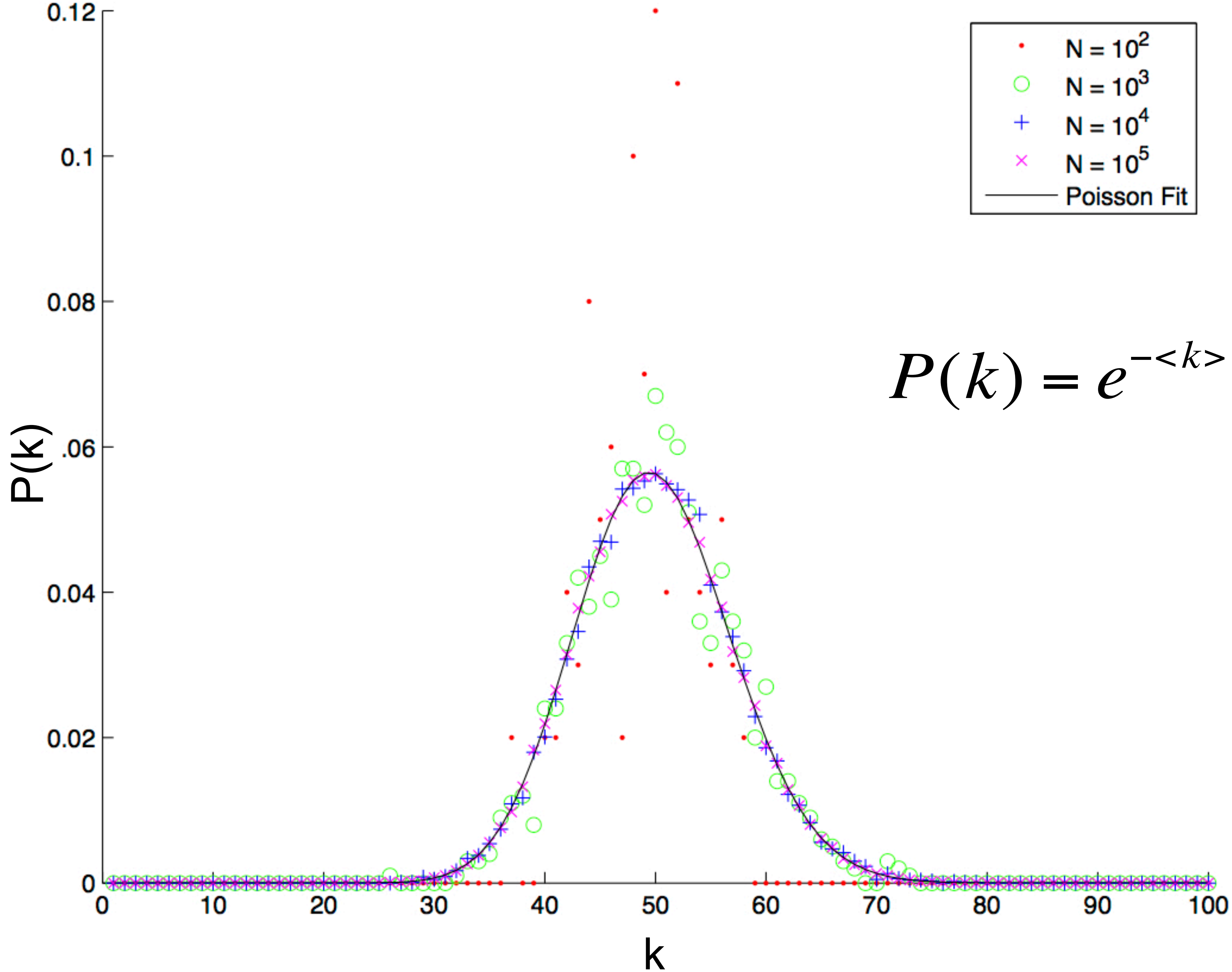
Distribution of degrees

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

standard deviation

$$\sigma = \sqrt{< k >}$$

Degree distribution - Random Graphs



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

Degree distribution - Random Graphs

Conclusion: degree distribution is **not**

-Heterogeneous

-Long tail

-Scale free

Clustering - Random Graphs

Local clustering of a node

Reminder, clustering coefficient

$$C_i \equiv \frac{2n_i}{k_i(k_i-1)}$$

where n_i is the number of links between the neighbours of node i

- Edges are independent and have the same probability p

$$n_i \cong p\,\underbrace{\frac{k_i(k_i-1)}{2}}_{\text{\# possible links btw neighbors}}$$

- Earlier we showed $p = \dfrac{<k>}{n-1}$

$$C_i = \frac{2<k>}{n-1}\,\frac{k_i\,(k_i-1)}{2}\,\frac{1}{k_i\,(k_i-1)} = \frac{<k>}{n-1} = p$$

- For fixed average degree C is decreasing as N goes large

➡ Low clustering coefficient

➡ It is vanishing with the system size

Clustering - ER Random Networks

- **Small clustering coefficient** $$C_i \equiv \frac{1}{N} < k >= p$$

Real-world networks

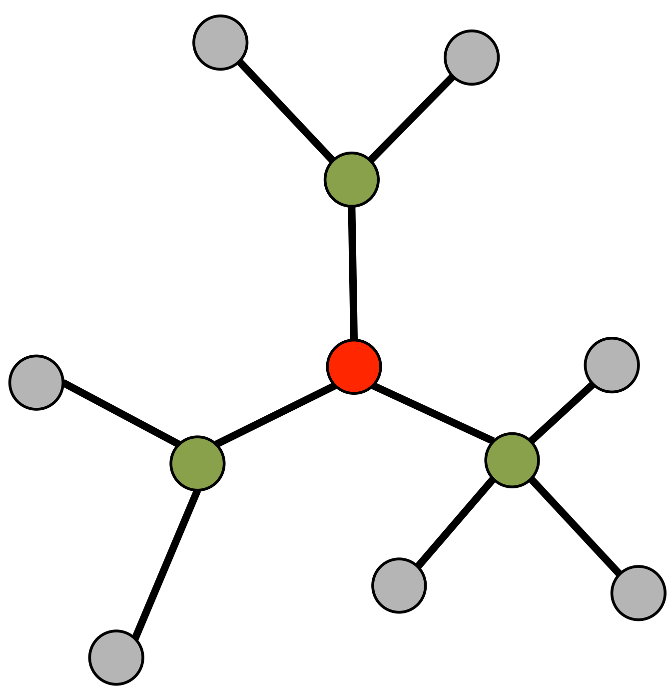
Network	Size	$\langle k \rangle$	ℓ	ℓ_{rand}	C	C_{rand}	Reference
WWW, site level, undir.	153 127	35.21	3.1	3.35	0.1078	0.00023	Adamic, 1999
Internet, domain level	3015–6209	3.52–4.11	3.7–3.76	6.36–6.18	0.18–0.3	0.001	Yook *et al.*, 2001a, Pastor-Satorras *et al.*, 2001
Movie actors	225 226	61	3.65	2.99	0.79	0.00027	Watts and Strogatz, 1998
LANL co-authorship	52 909	9.7	5.9	4.79	0.43	1.8×10^{-4}	Newman, 2001a, 2001b, 2001c
MEDLINE co-authorship	1 520 251	18.1	4.6	4.91	0.066	1.1×10^{-5}	Newman, 2001a, 2001b, 2001c
SPIRES co-authorship	56 627	173	4.0	2.12	0.726	0.003	Newman, 2001a, 2001b, 2001c
NCSTRL co-authorship	11 994	3.59	9.7	7.34	0.496	3×10^{-4}	Newman, 2001a, 2001b, 2001c
Math. co-authorship	70 975	3.9	9.5	8.2	0.59	5.4×10^{-5}	Barabási *et al.*, 2001
Neurosci. co-authorship	209 293	11.5	6	5.01	0.76	5.5×10^{-5}	Barabási *et al.*, 2001
E. coli, substrate graph	282	7.35	2.9	3.04	0.32	0.026	Wagner and Fell, 2000
E. coli, reaction graph	315	28.3	2.62	1.98	0.59	0.09	Wagner and Fell, 2000
Ythan estuary food web	134	8.7	2.43	2.26	0.22	0.06	Montoya and Solé, 2000
Silwood Park food web	154	4.75	3.40	3.23	0.15	0.03	Montoya and Solé, 2000
Words, co-occurrence	460.902	70.13	2.67	3.03	0.437	0.0001	Ferrer i Cancho and Solé, 2001
Words, synonyms	22 311	13.48	4.5	3.84	0.7	0.0006	Yook *et al.*, 2001b
Power grid	4941	2.67	18.7	12.4	0.08	0.005	Watts and Strogatz, 1998
C. Elegans	282	14	2.65	2.25	0.28	0.05	Watts and Strogatz, 1998

Albert, R. et.al. Rev. Mod. Phy. (2002)

Distance - Random Graphs

low clustering coefficient=>

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



- nr. of first neighbors: $N(u)_1 = < k >$
- nr. of second neighbors: $N(u)_2 = < k >^2$
- nr. of neighbours at distance d: $N(u)_d = < k >^d$

Intuition: At which distance are all nodes reached?

$$n = < k >^d \Rightarrow \log_{<k>} n = d \Rightarrow d = \frac{\log n}{\log < k >}$$

Diameter, avg. distance in $\mathcal{O}(\log n)$

Distance - ER Random Networks

- **Logarithmically short distance among nodes**

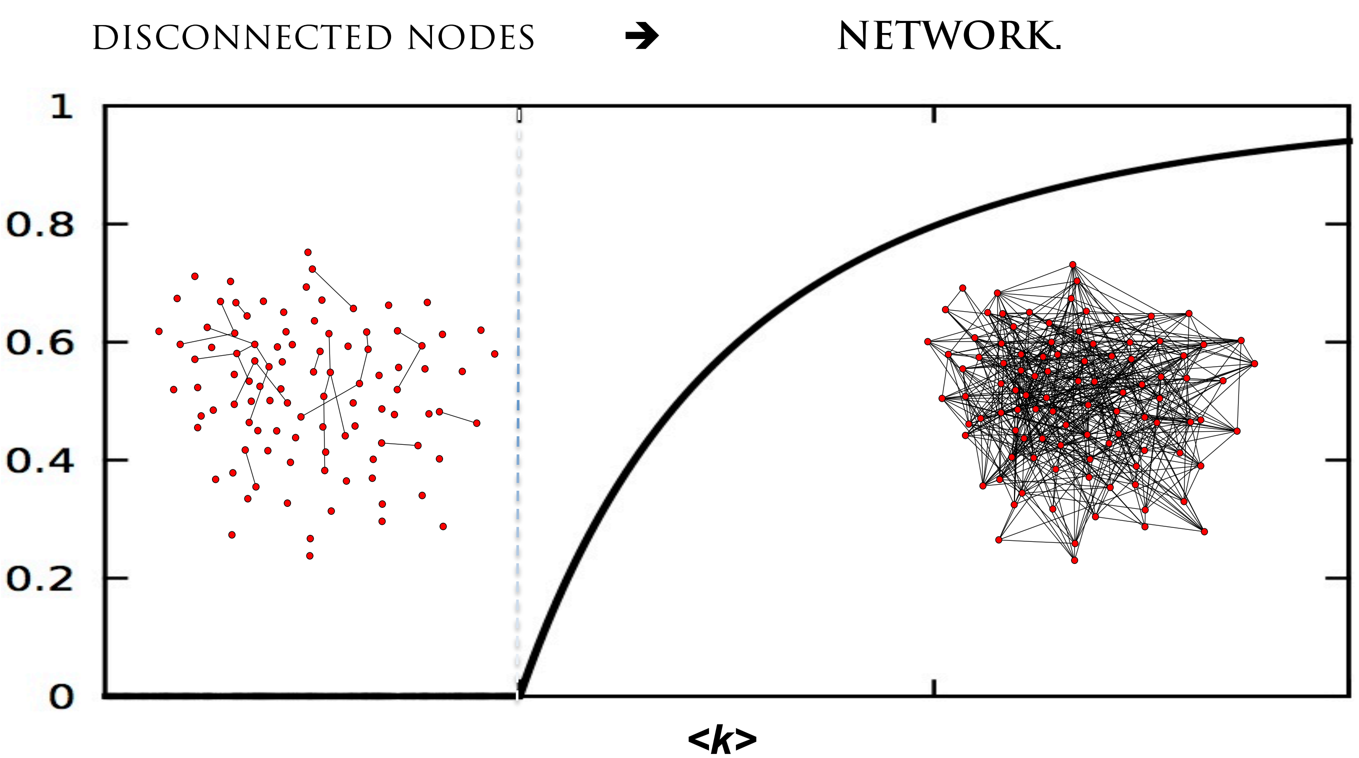
$$d = \frac{\log n}{log\langle k\rangle}$$

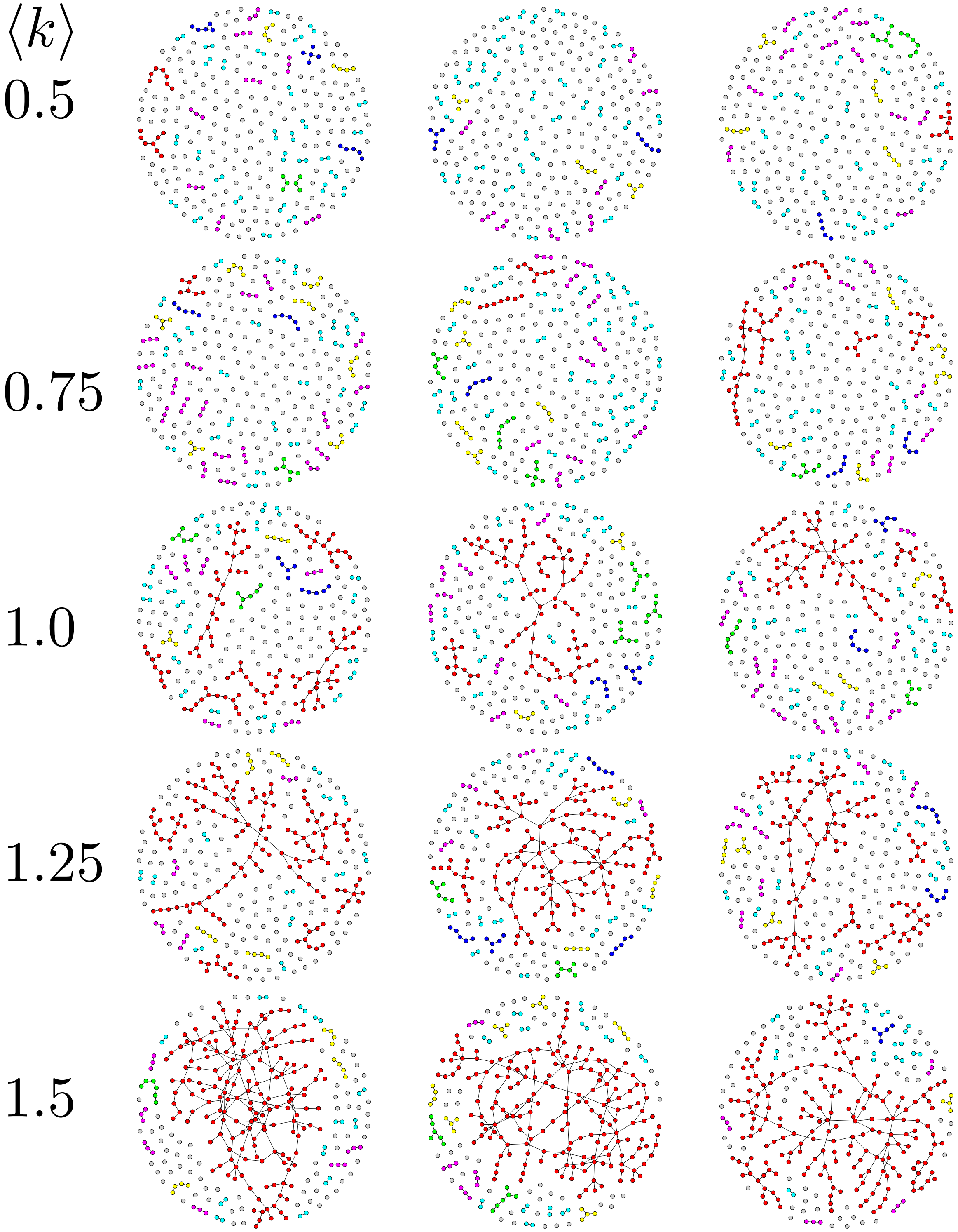
Real-world networks

Network	Size	$\langle k\rangle$	ℓ	ℓ_{rand}	C	C_{rand}	Reference
WWW, site level, undir.	153 127	35.21	3.1	3.35	0.1078	0.00023	Adamic, 1999
Internet, domain level	3015–6209	3.52–4.11	3.7–3.76	6.36–6.18	0.18–0.3	0.001	Yook *et al.*, 2001a, Pastor-Satorras *et al.*, 2001
Movie actors	225 226	61	3.65	2.99	0.79	0.00027	Watts and Strogatz, 1998
LANL co-authorship	52 909	9.7	5.9	4.79	0.43	1.8×10^{-4}	Newman, 2001a, 2001b, 2001c
MEDLINE co-authorship	1 520 251	18.1	4.6	4.91	0.066	1.1×10^{-5}	Newman, 2001a, 2001b, 2001c
SPIRES co-authorship	56 627	173	4.0	2.12	0.726	0.003	Newman, 2001a, 2001b, 2001c
NCSTRL co-authorship	11 994	3.59	9.7	7.34	0.496	3×10^{-4}	Newman, 2001a, 2001b, 2001c
Math. co-authorship	70 975	3.9	9.5	8.2	0.59	5.4×10^{-5}	Barabási *et al.*, 2001
Neurosci. co-authorship	209 293	11.5	6	5.01	0.76	5.5×10^{-5}	Barabási *et al.*, 2001
E. coli, substrate graph	282	7.35	2.9	3.04	0.32	0.026	Wagner and Fell, 2000
E. coli, reaction graph	315	28.3	2.62	1.98	0.59	0.09	Wagner and Fell, 2000
Ythan estuary food web	134	8.7	2.43	2.26	0.22	0.06	Montoya and Solé, 2000
Silwood Park food web	154	4.75	3.40	3.23	0.15	0.03	Montoya and Solé, 2000
Words, co-occurrence	460.902	70.13	2.67	3.03	0.437	0.0001	Ferrer i Cancho and Solé, 2001
Words, synonyms	22 311	13.48	4.5	3.84	0.7	0.0006	Yook *et al.*, 2001b
Power grid	4941	2.67	18.7	12.4	0.08	0.005	Watts and Strogatz, 1998
C. Elegans	282	14	2.65	2.25	0.28	0.05	Watts and Strogatz, 1998

Albert, R. et.al. Rev. Mod. Phy. (2002)

Connected components of Random Graphs



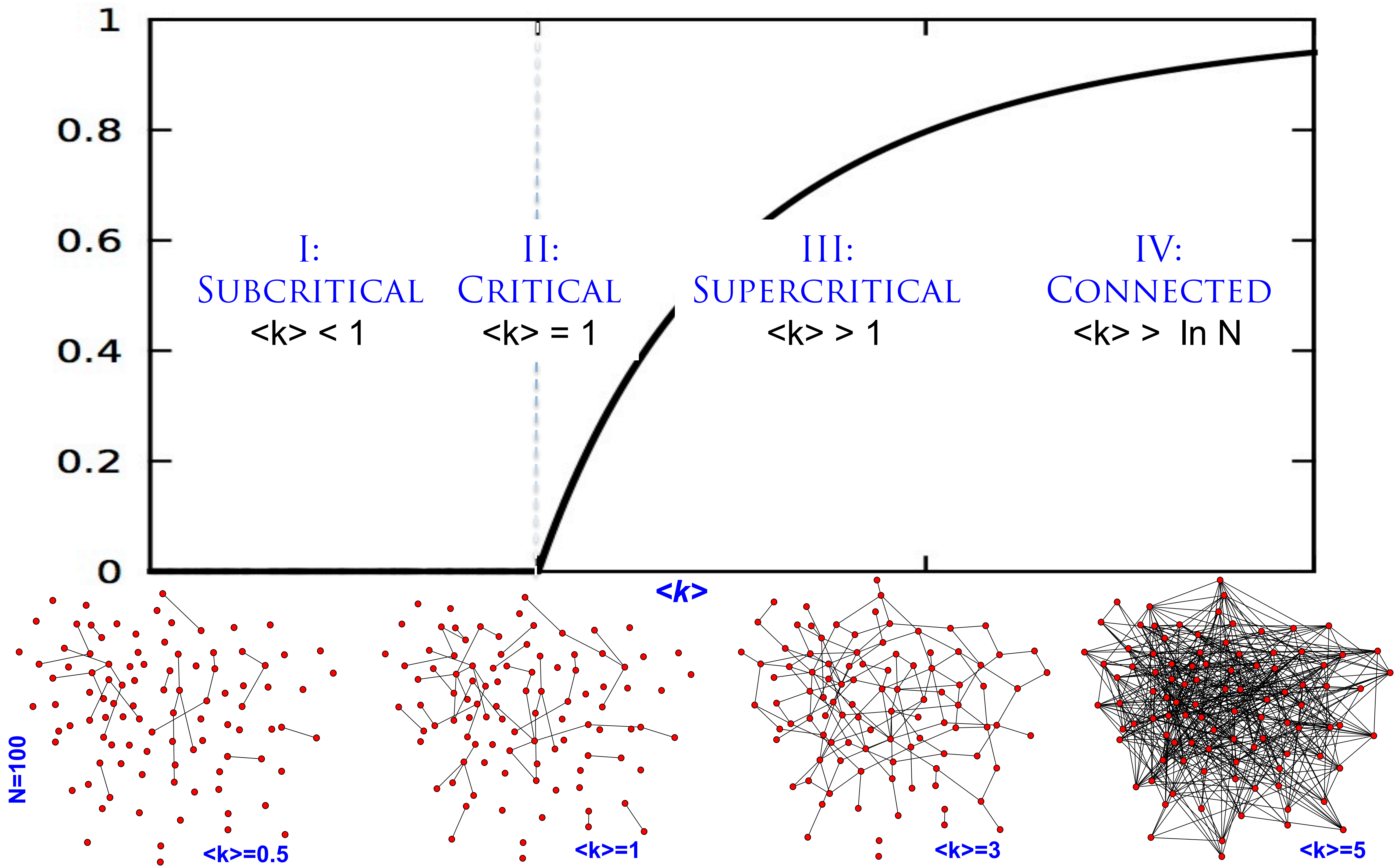


- Network structure goes through a transition
- Question: How and when does this transition happen

Connected components of Random Graphs

https://www.complexity-explorables.org/explorables/the-blob/

ER Random Network - catch up



Structural (percolation) phase transition at <k>=1 (or equivalently when p=1/N)

ER Random Network - catch up

Basic characteristics

- Degree distribution

$$p_k = \binom{n-1}{k} p^k (1-p)^{n-1-k} \xrightarrow[N \to \infty]{} p_k = \frac{\langle k \rangle^k}{k!} e^{-\langle k \rangle}$$

Binomial distribution — Poisson distribution

Degree distribution without tail

- Clustering

$$C_i = \frac{<k>}{n-1} = p$$

Vanishing clustering coefficient for large size

- Path length

$$\mathcal{O}(\log n)$$

Distance with logarithmic relation to nodes

ER Random Network - catch up

Network	Degree distribution	Path length	Clustering coefficient
Real world networks	broad	short	large
Regular lattices	constant	long	large
ER random networks	Poissonian	short	small

It is not capturing the properties of any real system

BUT

it serves as a reference system for any other network model

Configuration model

More details at [http://tuvalu.santafe.edu/~aaronc/courses/5352/fall2013/csci5352_2013_L11.pdf]

Random graphs with specified degrees

Problem

- The ER Random Graph model has a Poisson degree distribution
- Most real-world networks have heavy-tailed degree distributions
- **We need to generate networks which have pre-determined degrees or degree distribution, but they are maximally random otherwise**
- The observed properties (clustering coefficient, etc.) might be due *only* to the difference in degree distribution

Random graphs with specified degrees

Configuration model

How much of some observed pattern is driven by the degrees alone?

Based on an observed network

- Defined as $G(n, \vec{k})$ where $\vec{k} = \{k_i\}$ is a degree sequence on n nodes, with k_i being the degree of node i

Ad hoc degree distribution

- **The degree sequence $\vec{k} = \{k_i\}$ can be sampled from a probability distribution**
  - Delta/Dirac function => Random regular graph
  - Poisson => Similar to ER for proper parameters
  - Scale-free => Power-law random graph
- Only global condition to satisfy is: $\sum_i k_i \textbf{ mod } 2 = 0$

  (even degree sum) i.e. each edge has to have ending nodes

Random graphs with specified degrees

Configuration model

How much of some observed pattern is driven by the degrees alone?

Exact or approximate degree distribution

- The model can preserve the **expected** degree sequence, or the **exact** degree sequence
    - Chung-lu (appoximate)
    - Molloy-reed (Exact)

Random graphs with specified degrees

Chung-Lu model for configuration networks = Approximate degree distribution

- Probabilistic model which produce a network with degrees approximating (on average) the original degree
- It is a *"coin-flipping"* process as ER model but **the probability that two nodes i and j are connected depends on the degree k_i and k_j of the ending nodes**
- From the point of node i with degree k_i, the probability that one of its edges will connect to j with k_j:

$$k_j/2m$$

- This can happen via k_i links, thus the probability that they are connected:

$$p_{ij} = \frac{k_i k_j}{2m}$$

assuming that: $[\max(k_i)]^2 < 2m$

(/!\ inconsistent probability, it is rather expected number of edges)

- Chung-Lu model takes each pairs of nodes and connects them with this probability

$$\forall_{i>j} \qquad A_{ij} = A_{ji} = \begin{cases} 1 & \text{with probability } p_{ij} \\ 0 & \text{otherwise} \end{cases}$$

Random graphs with specified degrees

Chung-Lu model for configuration networks = Approximate degree distribution

$$\forall_{i>j} \qquad A_{ij} = A_{ji} = \begin{cases} 1 & \text{with probability } p_{ij} \\ 0 & \text{otherwise} \end{cases} \qquad \text{where} \qquad p_{ij} = \frac{k_i k_j}{2m}$$

- Each pairs of nodes are considered once, thus it produces a **simple graph** (without self-loops and multi edges)

- Degree of a node equals only in "expectation" to the originally assigned degree

- It is easy to generalise for directed graphs
- Inconsistency for large degrees in small networks $[\max(k_i)]^2 < 2m$

Complexity:

- *O(n²)*: We need $n(n\text{-}1)$ flips to test all node pairs

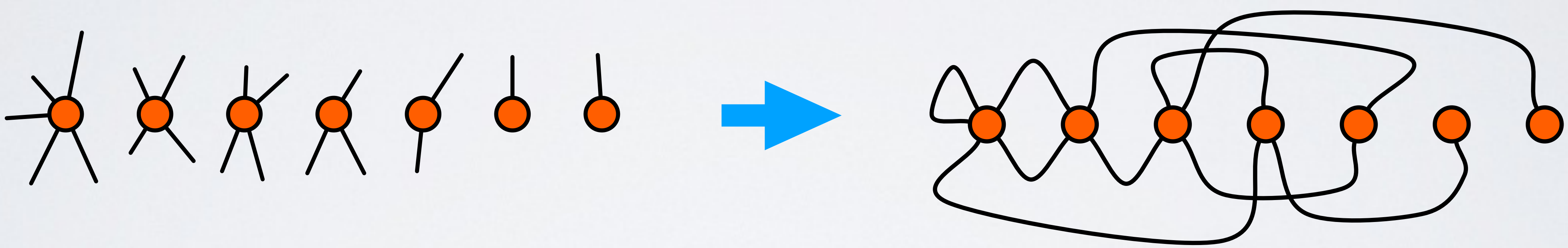
EXPENSIVE!

Random graphs with specified degrees

Molloy-Reed model for configuration networks = exact degree preservation

Original idea:

1. Given a degree sequence $\vec{k} = \{k_1, k_2, \ldots, k_n\}$
2. Assign each node $i \in V$ with k_i number of stubs
3. Select random pairs of unmatched stubs and connect them
4. Repeat 3 while there are unmatched stubs



- This process will produce a configuration model with exact degree sequence
- Possible to select multiple times stubs of the same pair of nodes ➜ **Multilinks**
- Possible to select the stubs of the same node to connect ➜ **Self-links**

The obtained graph is not simple…but the density of multi and self-links ➜ 0 as N ➜ ∞

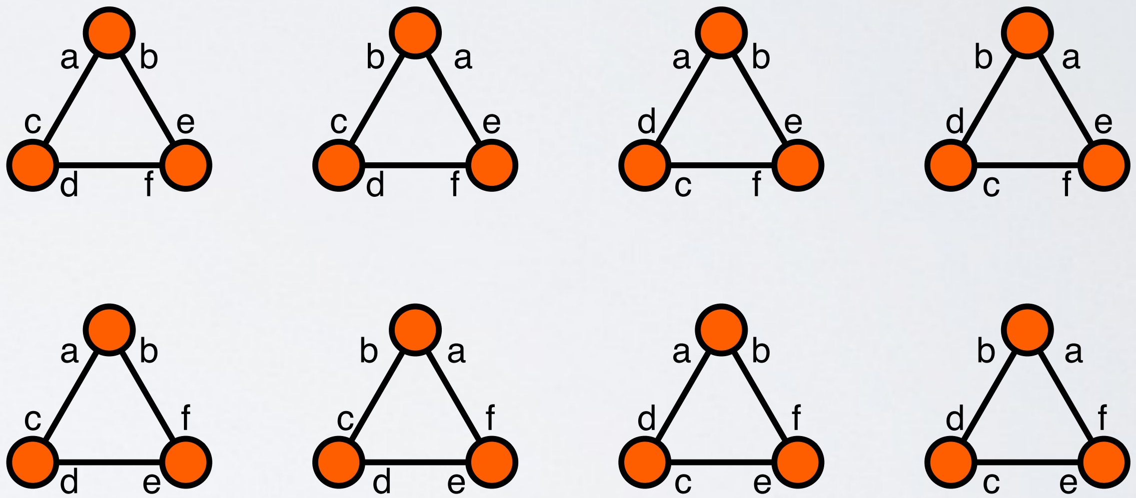
Random graphs with specified degrees

Molloy-Reed model for configuration networks = exact degree preservation

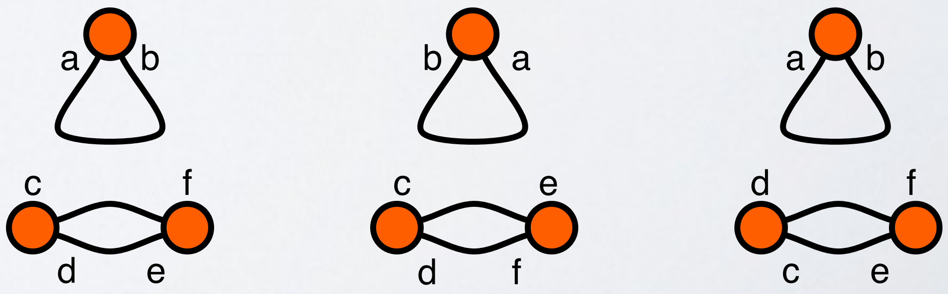
Non-unique problem

- Matching of stubs appears with equal probability
- **BUT networks with the same $\{k_i\}$ do not appear with equal probability**
- More than one matching can correspond to the same network (topologically)

Different matchings yield same graphs



Some graphs produced by less combinations =>less likely to appear

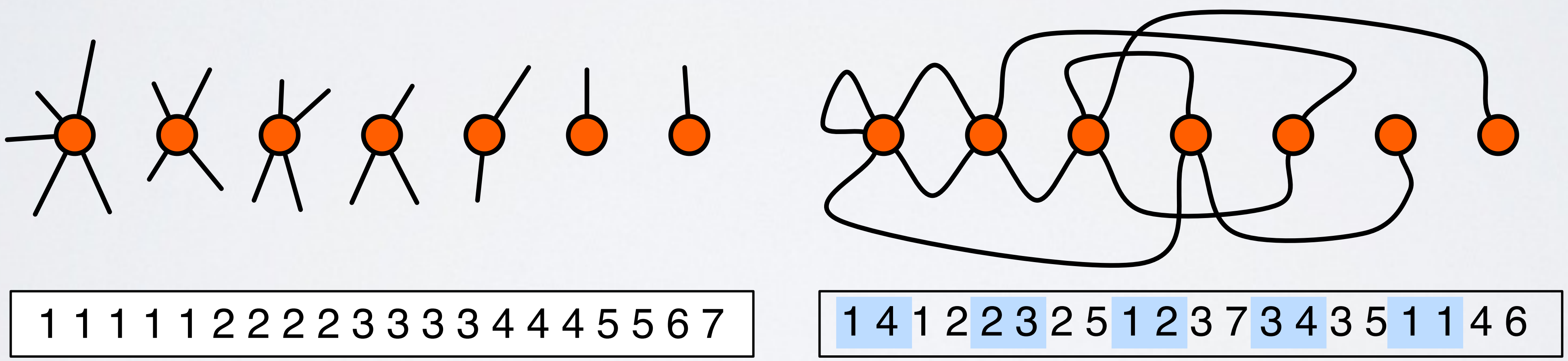


Random graphs with specified degrees

Molloy-Reed model for configuration networks = exact degree preservation

An effective algorithm:

1. Take an array $\vec{v}$ with length $2m$ and fill it with exactly k_i indices of each node $i \in V$
2. Make a random permutation of the array $\vec{v}$
3. Read the content of the array in an order and in pairs
4. Pairs of consecutive node indices will assign links in the configuration network



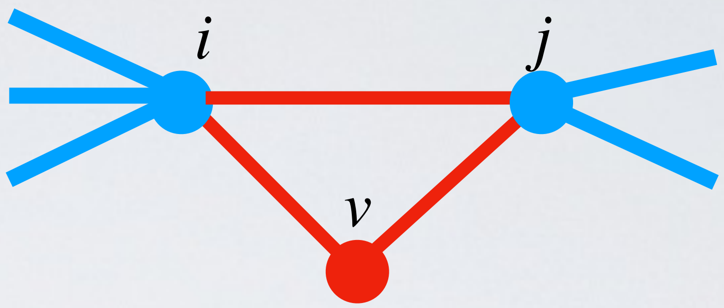
Complexity:

- *O(m)*: Random permutation of an array
- *O(m log m)*: assigning uniformly random variables to indices and quick-sort them

CHEAP!

Configuration model - mathematical properties

Expected clustering coefficient



It is the average probability that two neighbours of a vertex are neighbours

- Start at some vertex v (with degree $k \geq 2$)
- Choose a random pair of its neighbours i and j
- The probability that i and j are themselves connected is $k_i k_j / 2m$

Clustering coefficient

independent of network size

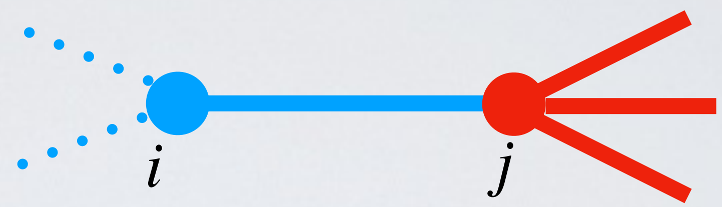
$$C = \ldots = \frac{1}{n} \boxed{\frac{[\langle k \rangle^2 - \langle k \rangle]^2}{\langle k \rangle^3}}$$

- It is a vanishing quantity $O(1/n)$ as long as the second moment is finite (not power law)

For details, see: http://tuvalu.santafe.edu/~aaronc/courses/5352/fall2013/csci5352_2013_L12.pdf

Configuration model - mathematical properties

Neighbors's degrees



What is the degree distribution of neighbors of a randomly chosen vertex?

- Let p_k be the fraction of vertices in the network with degree k
- There are np_k vertices of degree k in the network.
- The end point of every edge in the network has the same probability $\frac{k}{2m}$ of connecting to a vertex of degree *k*
- **Degree distribution of a randomly picked neighbor (of any node)**

$$p_{neighb,k} = \frac{k}{2m} np_k = \frac{kp_k}{\langle k \rangle}$$

Configuration model - mathematical properties

- Degree distribution of a randomly picked neighbor (of any node)

$$p_{neighb,k} = \frac{k}{2m} n p_k = \frac{k p_k}{\langle k \rangle}$$

- Average degree of a randomly picked neighbor

$$\langle k_{neighb} \rangle = \sum_k k p_{neighb,k} = \frac{\langle k^2 \rangle}{\langle k \rangle}$$

- Larger than $\langle k \rangle$ as soon as degrees are heterogeneous ➡ **Friendship paradox**

1 node with degree 10, 10 nodes with degree 1:

$$\langle k \rangle = \frac{10 + 1 * 10}{11} = 1.81..$$

$$\langle k^2 \rangle = \frac{10^2 + 1^2 * 10}{11} = 10$$

$$\frac{\langle k^2 \rangle}{\langle k \rangle} = \frac{10}{1.82} = 5.5$$

ER Random Network - catch up

Network	Degree distribution	Path length	Clustering coefficient
Real world networks	broad	short	large
Regular lattices	constant	long	large
ER random networks	Poissonian	short	small
Configuration Model	Custom, can be broad	short	small

Watts-Strogatz model of small-world networks

Small-world networks

- On of the first paper of Network Science…

  D.J. Watts and S. Strogatz,

  "Collective dynamics of 'small-world' networks", Nature 393, 440–442, 1998

- Observation in real world networks:

letters to nature

typically slower than ~1 km s^{-1}) might differ significantly from what is assumed by current modelling efforts[27]. The expected equation-of-state differences among small bodies (ice versus rock, for instance) presents another dimension of study; having recently adapted our code for massively parallel architectures (K. M. Olson and E.A, manuscript in preparation), we are now ready to perform a more comprehensive analysis.

The exploratory simulations presented here suggest that when a young, non-porous asteroid (if such exist) suffers extensive impact damage, the resulting fracture pattern largely defines the asteroid's response to future impacts. The stochastic nature of collisions implies that small asteroid interiors may be as diverse as their shapes and spin states. Detailed numerical simulations of impacts, using accurate shape models and rheologies, could shed light on how asteroid collisional response depends on internal configuration and shape, and hence on how planetesimals evolve. Detailed simulations are also required before one can predict the quantitative effects of nuclear explosions on Earth-crossing comets and asteroids, either for hazard mitigation[28] through disruption and deflection, or for resource exploitation[29]. Such predictions would require detailed reconnaissance concerning the composition and internal structure of the targeted object. □

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

Networks of coupled dynamical systems have been used to model biological oscillators[1–4], Josephson junction arrays[5,6], excitable media[7], neural networks[8–10], spatial games[11], genetic control networks[12] and many other self-organizing systems. Ordinarily, the connection topology is assumed to be either completely regular or completely random. But many biological, technological and social networks lie somewhere between these two extremes. Here we explore simple models of networks that can be tuned through this middle ground: regular networks 'rewired' to introduce increasing amounts of disorder. We find that these systems can be highly clustered, like regular lattices, yet have small characteristic path lengths, like random graphs. We call them 'small-world' networks, by analogy with the small-world

Table 1 Empirical examples of small-world networks

	L_{actual}	L_{random}	C_{actual}	C_{random}	N
Film actors	3.65	2.99	0.79	0.00027	22500
Power grid	18.7	12.4	0.080	0.005	4941
C. elegans	2.65	2.25	0.28	0.05	282

Contradiction: Real-world networks have

High clustering coefficient

AND

Short distances

Clustering vs. Interconnectedness

Random networks

- **Logarithmically short distance among nodes** $d = \frac{logN}{log\langle k\rangle}$ ✔
- Vanishing clustering coefficient for large size $C_i \equiv \frac{1}{N}\langle k\rangle = p$

Real-world networks

Network	Size	$\langle k\rangle$	ℓ	ℓ_{rand}	C	C_{rand}	Reference
WWW, site level, undir.	153 127	35.21	3.1	3.35	0.1078	0.00023	Adamic, 1999
Internet, domain level	3015–6209	3.52–4.11	3.7–3.76	6.36–6.18	0.18–0.3	0.001	Yook *et al.*, 2001a, Pastor-Satorras *et al.*, 2001
Movie actors	225 226	61	3.65	2.99	0.79	0.00027	Watts and Strogatz, 1998
LANL co-authorship	52 909	9.7	5.9	4.79	0.43	1.8×10^{-4}	Newman, 2001a, 2001b, 2001c
MEDLINE co-authorship	1 520 251	18.1	4.6	4.91	0.066	1.1×10^{-5}	Newman, 2001a, 2001b, 2001c
SPIRES co-authorship	56 627	173	4.0	2.12	0.726	0.003	Newman, 2001a, 2001b, 2001c
NCSTRL co-authorship	11 994	3.59	9.7	7.34	0.496	3×10^{-4}	Newman, 2001a, 2001b, 2001c
Math. co-authorship	70 975	3.9	9.5	8.2	0.59	5.4×10^{-5}	Barabási *et al.*, 2001
Neurosci. co-authorship	209 293	11.5	6	5.01	0.76	5.5×10^{-5}	Barabási *et al.*, 2001
E. coli, substrate graph	282	7.35	2.9	3.04	0.32	0.026	Wagner and Fell, 2000
E. coli, reaction graph	315	28.3	2.62	1.98	0.59	0.09	Wagner and Fell, 2000
Ythan estuary food web	134	8.7	2.43	2.26	0.22	0.06	Montoya and Solé, 2000
Silwood Park food web	154	4.75	3.40	3.23	0.15	0.03	Montoya and Solé, 2000
Words, co-occurrence	460.902	70.13	2.67	3.03	0.437	0.0001	Ferrer i Cancho and Solé, 2001
Words, synonyms	22 311	13.48	4.5	3.84	0.7	0.0006	Yook *et al.*, 2001b
Power grid	4941	2.67	18.7	12.4	0.08	0.005	Watts and Strogatz, 1998
C. Elegans	282	14	2.65	2.25	0.28	0.05	Watts and Strogatz, 1998

Albert, R. et.al. Rev. Mod. Phy. (2002)

Clustering vs. Interconnectedness

Random networks

- Logarithmically short distance among nodes $d = \dfrac{logN}{log\langle k\rangle}$ ✓
- **Vanishing clustering coefficient for large size** $C_i \equiv \dfrac{1}{N}\langle k\rangle = p$ ✗

Real-world networks

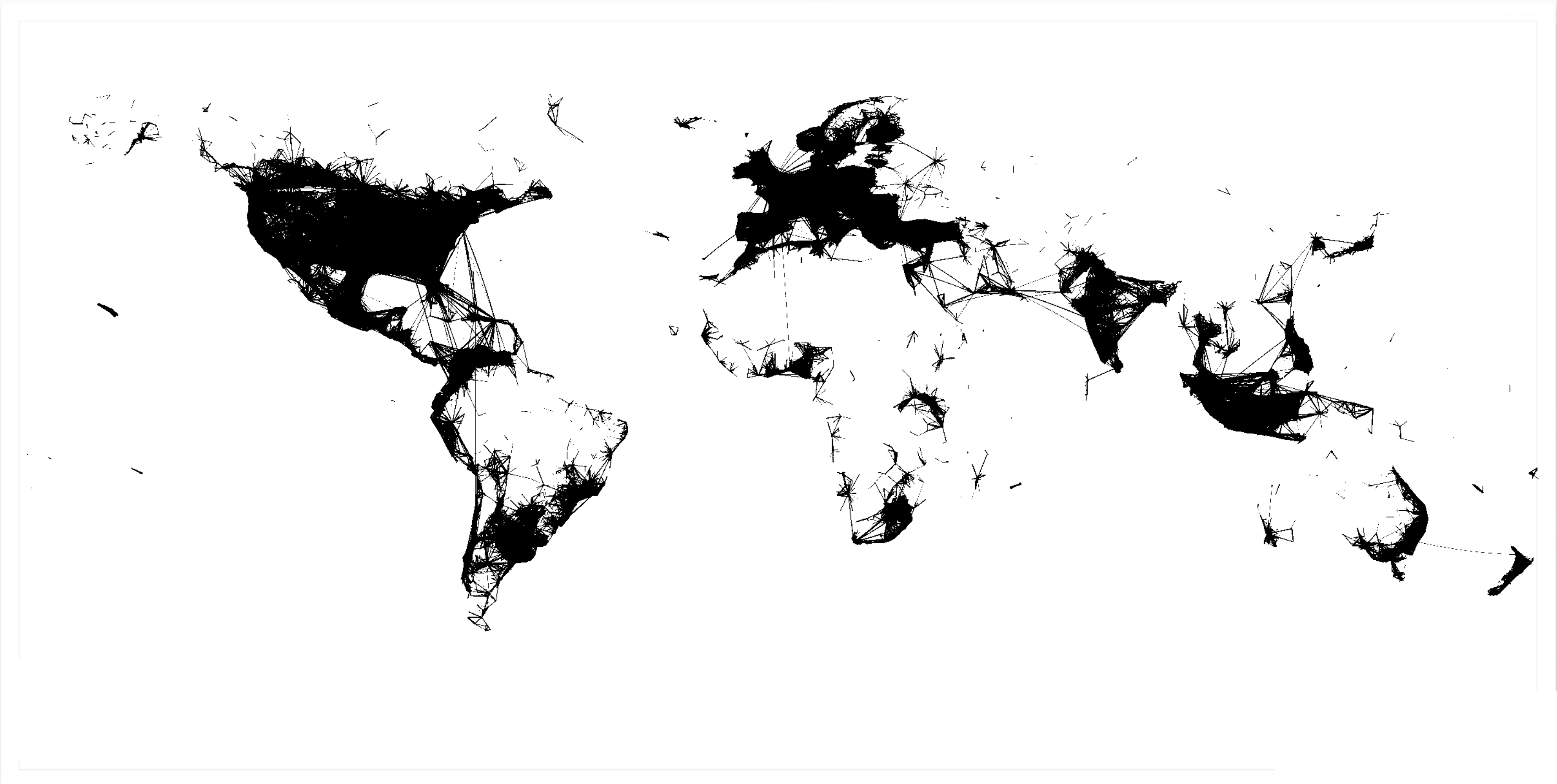
Network	Size	$\langle k\rangle$	ℓ	ℓ_{rand}	C	C_{rand}	Reference
WWW, site level, undir.	153 127	35.21	3.1	3.35	0.1078	0.00023	Adamic, 1999
Internet, domain level	3015–6209	3.52–4.11	3.7–3.76	6.36–6.18	0.18–0.3	0.001	Yook *et al.*, 2001a, Pastor-Satorras *et al.*, 2001
Movie actors	225 226	61	3.65	2.99	0.79	0.00027	Watts and Strogatz, 1998
LANL co-authorship	52 909	9.7	5.9	4.79	0.43	1.8×10^{-4}	Newman, 2001a, 2001b, 2001c
MEDLINE co-authorship	1 520 251	18.1	4.6	4.91	0.066	1.1×10^{-5}	Newman, 2001a, 2001b, 2001c
SPIRES co-authorship	56 627	173	4.0	2.12	0.726	0.003	Newman, 2001a, 2001b, 2001c
NCSTRL co-authorship	11 994	3.59	9.7	7.34	0.496	3×10^{-4}	Newman, 2001a, 2001b, 2001c
Math. co-authorship	70 975	3.9	9.5	8.2	0.59	5.4×10^{-5}	Barabási *et al.*, 2001
Neurosci. co-authorship	209 293	11.5	6	5.01	0.76	5.5×10^{-5}	Barabási *et al.*, 2001
E. coli, substrate graph	282	7.35	2.9	3.04	0.32	0.026	Wagner and Fell, 2000
E. coli, reaction graph	315	28.3	2.62	1.98	0.59	0.09	Wagner and Fell, 2000
Ythan estuary food web	134	8.7	2.43	2.26	0.22	0.06	Montoya and Solé, 2000
Silwood Park food web	154	4.75	3.40	3.23	0.15	0.03	Montoya and Solé, 2000
Words, co-occurrence	460.902	70.13	2.67	3.03	0.437	0.0001	Ferrer i Cancho and Solé, 2001
Words, synonyms	22 311	13.48	4.5	3.84	0.7	0.0006	Yook *et al.*, 2001b
Power grid	4941	2.67	18.7	12.4	0.08	0.005	Watts and Strogatz, 1998
C. Elegans	282	14	2.65	2.25	0.28	0.05	Watts and Strogatz, 1998

Albert R. et.al. Rev. Mod. Phy. (2002)

Clustering vs. Interconnectedness

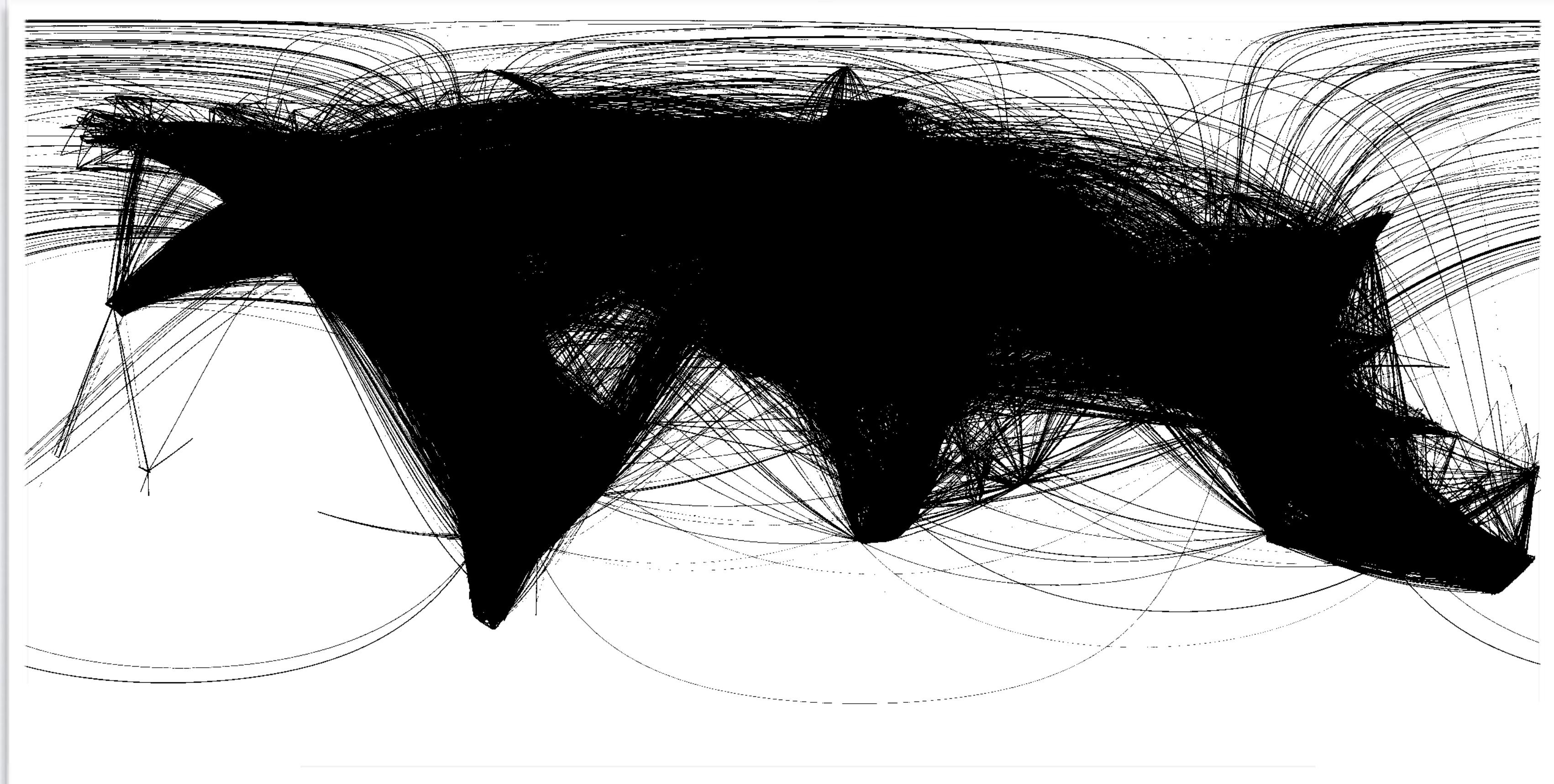
High clustering

- Locally structured
- No connections between nodes apart



Random

- Globally interconnected
- Low clustering



Clustering vs. Interconnectedness

Real networks have high clustering and short distances

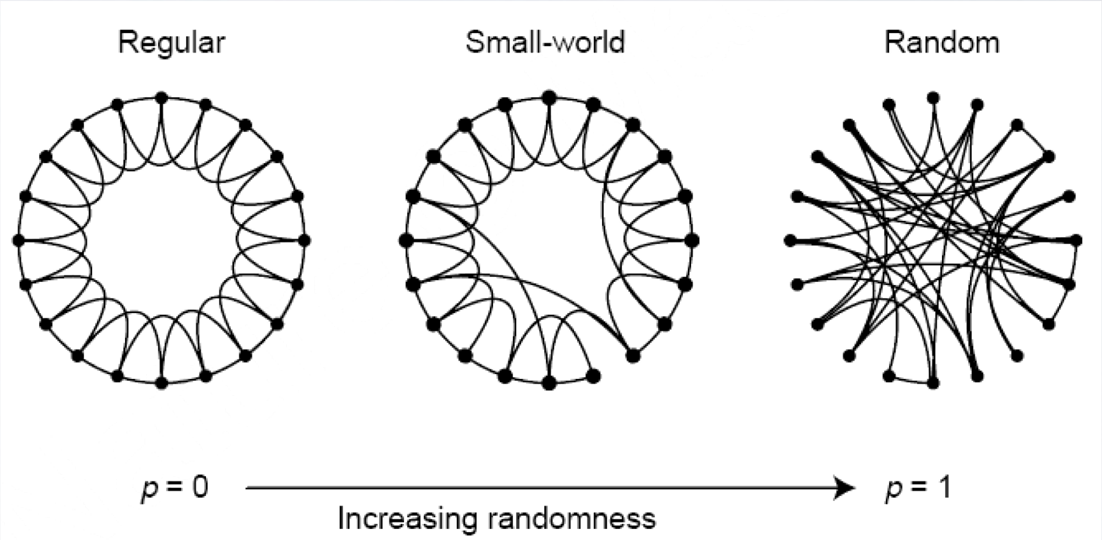
Strength of weak ties

- Mark granovetter, (sociologist)
- The Strength of weak ties (1973)
  - Very influential paper
- Weak ties (distant friendship, relatives) are very important for information flow (marketing, politics, job seeking, etc.)
  - Strong ties connect similar people, and lead to high redundancy
  - Weak ties, connect distant people, allow to leave the "community"

The Watts-Strogatz model

A model to capture large clustering coefficient and short distances observed in real networks

- It interpolates between an ordered finite lattice and a random graph
- Fixed parameters:
  - n - system size
  - K - initial coordination number
- Variable parameters:
  - p - rewiring probability



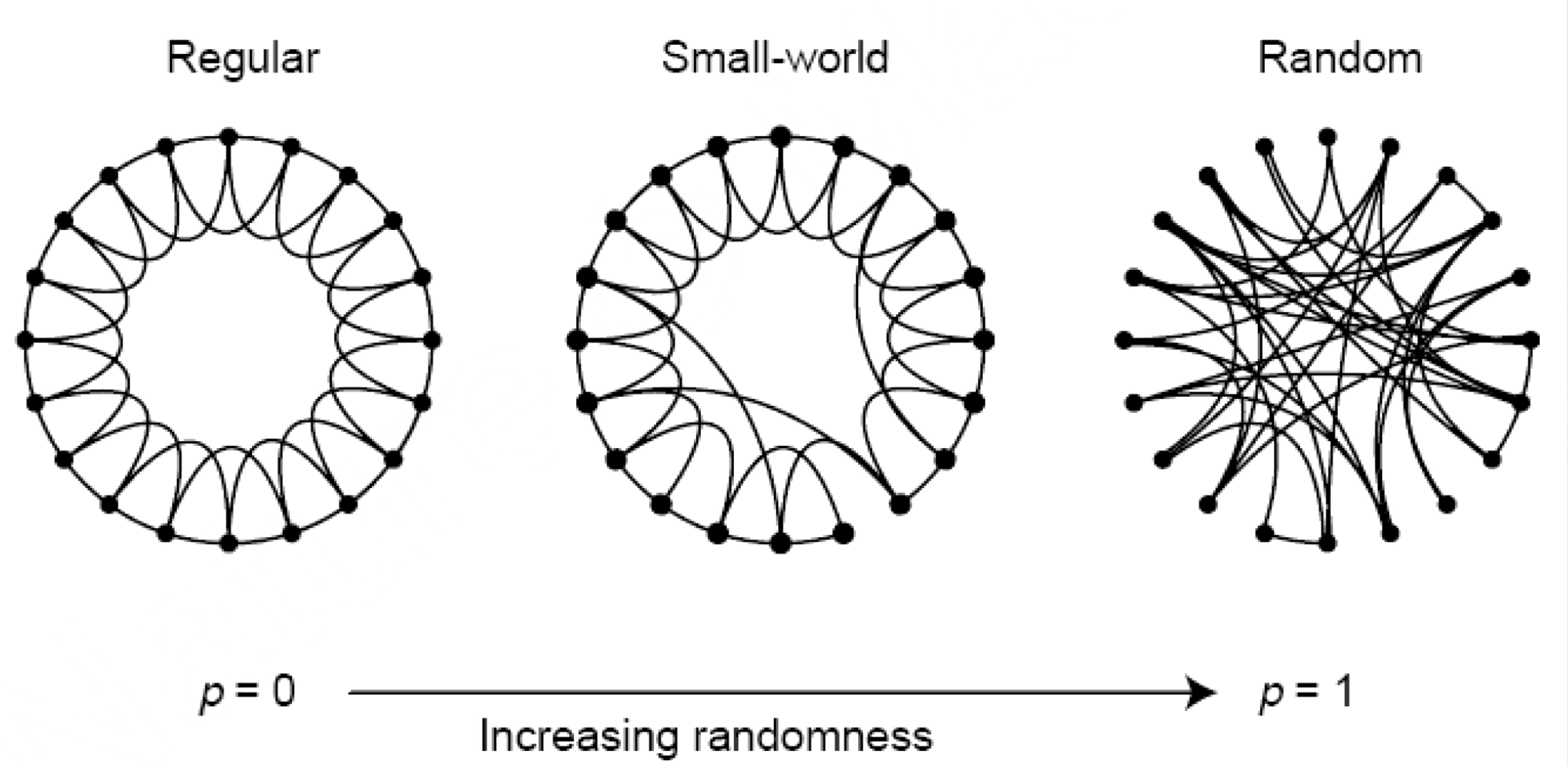
D.J. Watts and S. Strogatz, Nature (1998)

- Algorithm:
  1. Start with a ring lattice with n nodes in which every node is connected to its first K neighbours ($K/2$ on either side).
  2. Randomly rewire each edge of the lattice with probability p such that self-connections and duplicate edges are excluded.

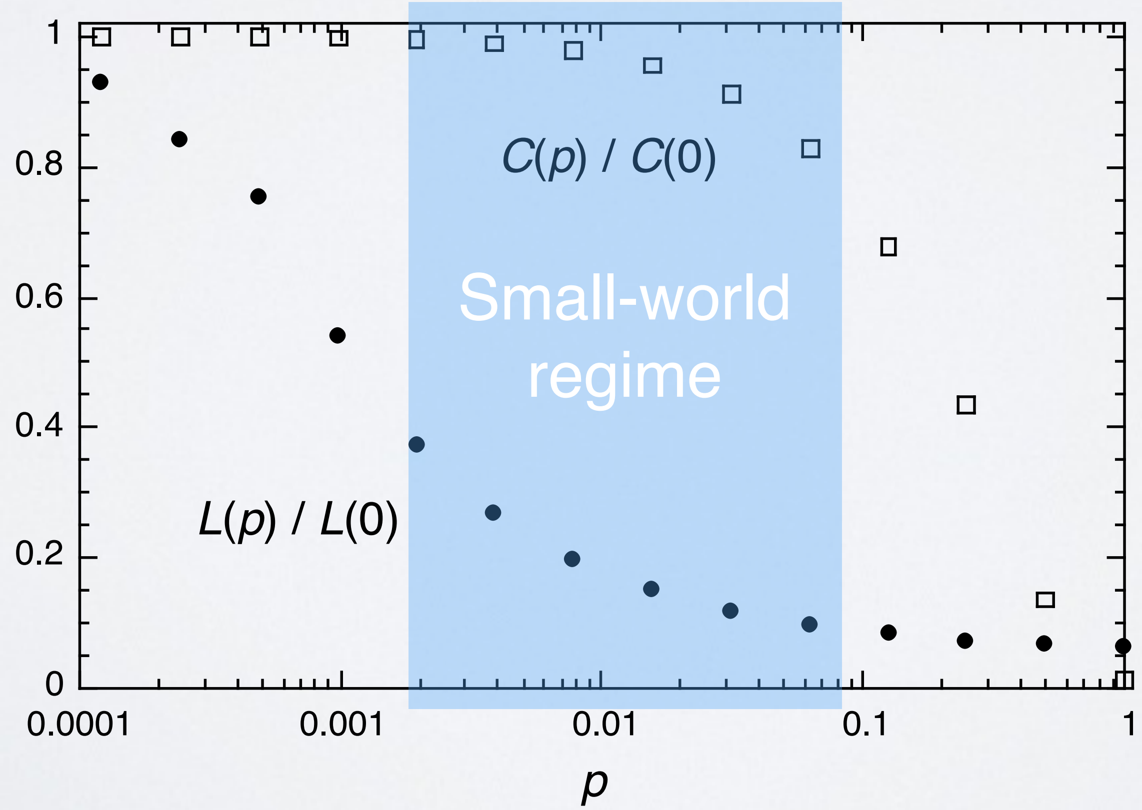
By varying p the network can be transformed from a completely ordered ($p=0$) to a completely random ($p=1$) structure

The Watts-Strogatz model

- n and K are chosen $n \gg K \gg ln(n) \gg 1$ thus the random graph remains connected ($K \gg ln(n)$)

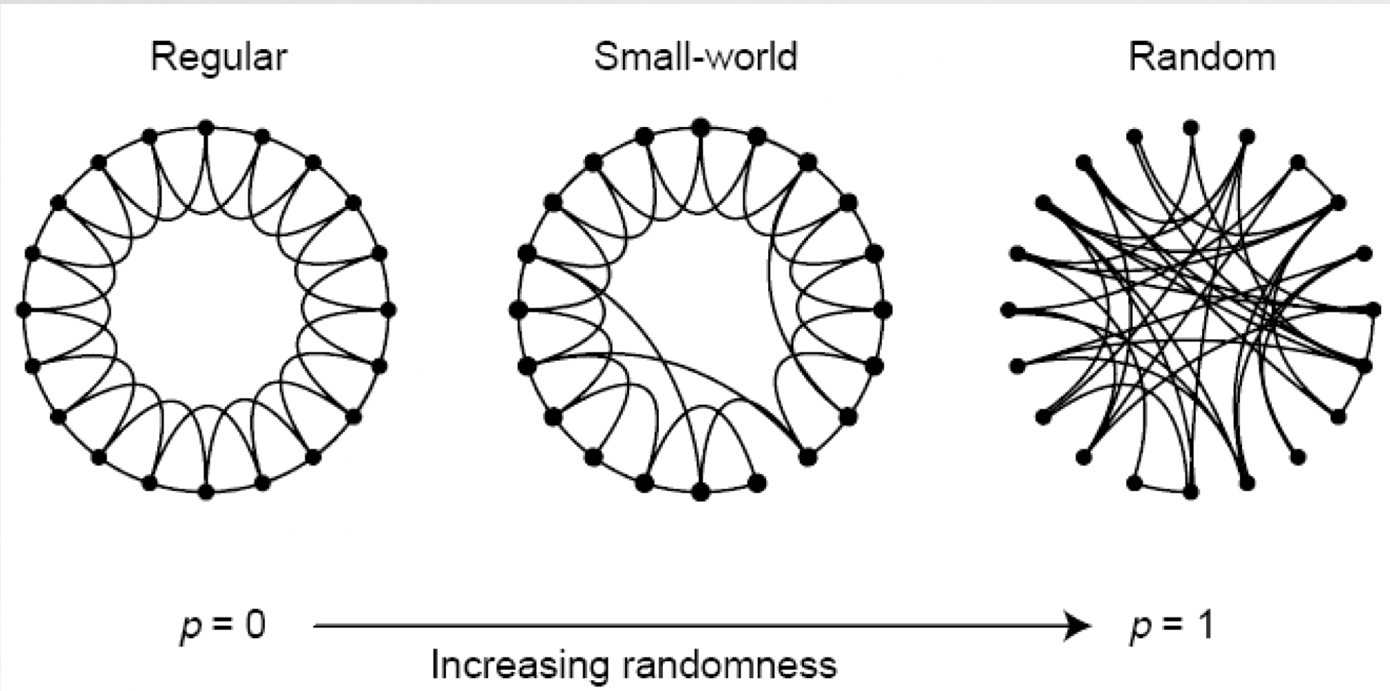


High clustering,
Long distances



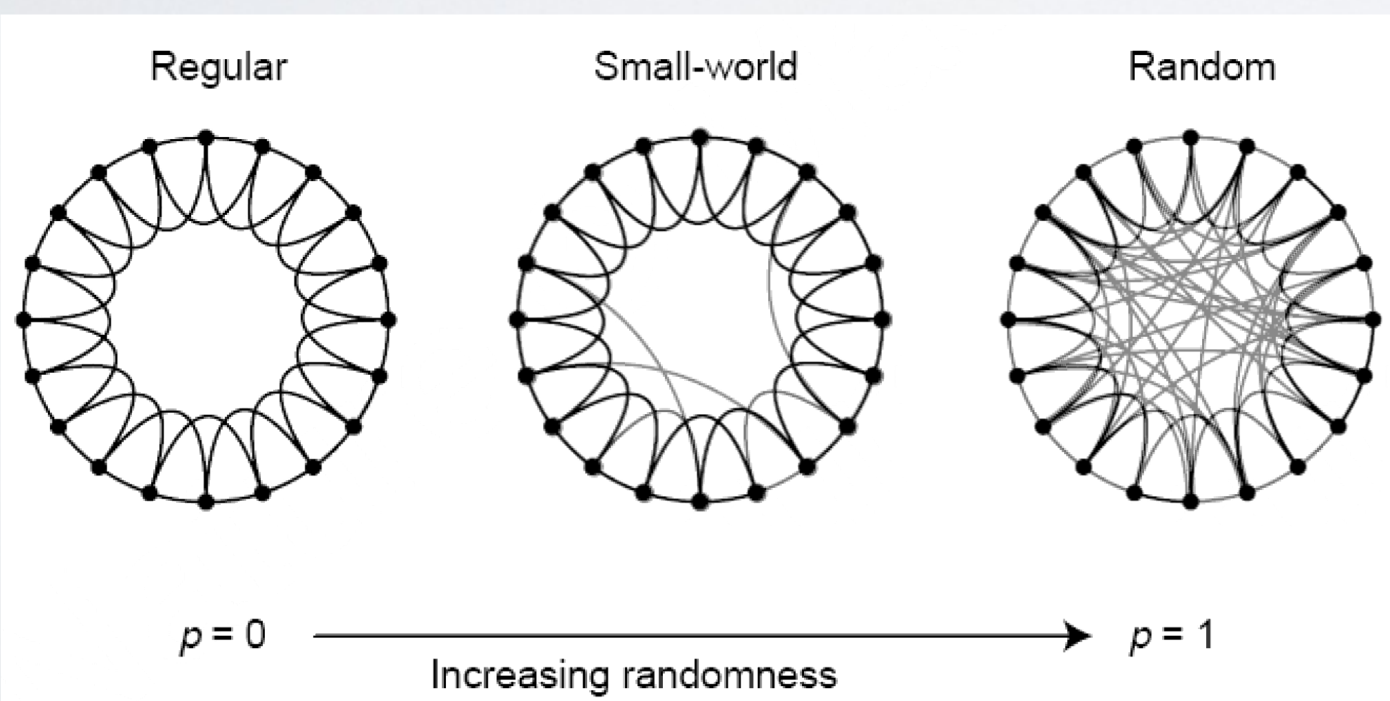
Low clustering,
Short distances

The Watts-Strogatz model



- Definition 1:

    1. Start with a ring lattice with N nodes in which every node is connected to its first K neighbours ($K/2$ on either side).
    2. Randomly rewire each edge of the lattice with probability p such that self-connections and duplicate edges are excluded.

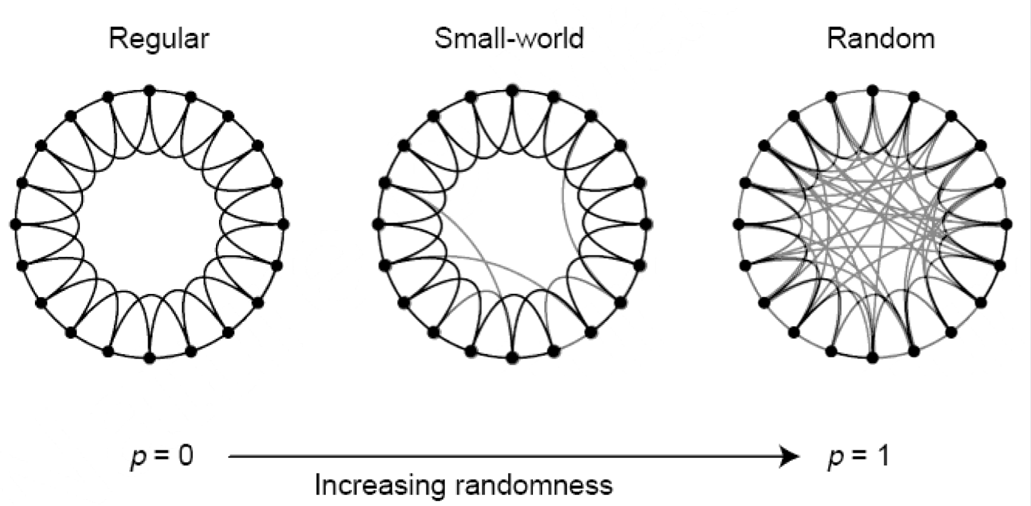


Definition 2:

1. Start with a ring lattice with N nodes in which every node is connected to its first K neighbours ($K/2$ on either side).
2. For every edge in the network we add an additional edge with independent probability p, connected two nodes selected uniformly at random

The Watts-Strogatz model

(Global) Clustering coefficient (Definition 2)



- *p=0* - regular ring with constant clustering: $C = \dfrac{3(K-2)}{4(K-1)}$
  - *0 ≤ C ≤ 3/4*
  - Independent of *n*

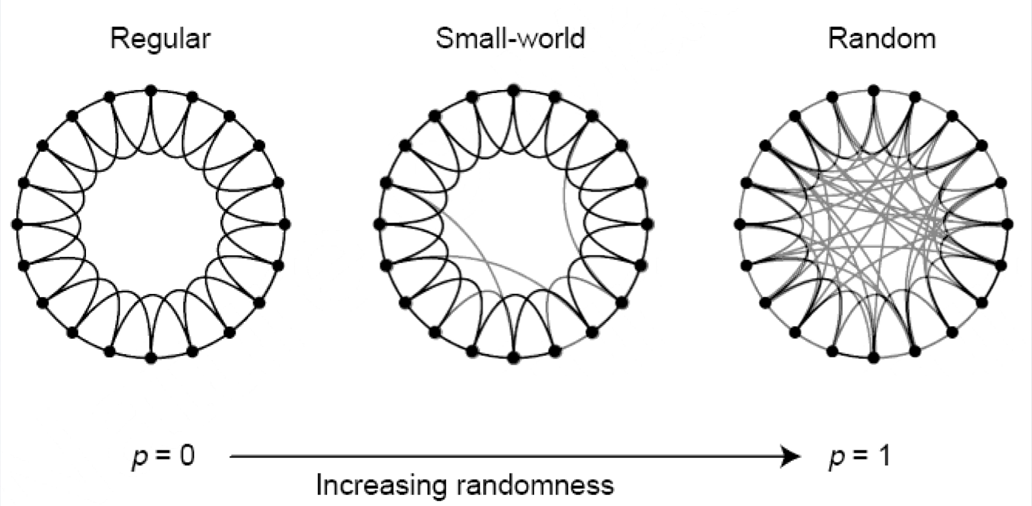
- *p>0* - we can count triangles and tuples

Global clustering coefficient

$$C = \frac{\frac{1}{4}NK(\frac{1}{2}K-1)\times 3}{\frac{1}{2}NK(K-1)+NK^2p+\frac{1}{2}NK^2p^2} = \frac{3(K-2)}{4(K-1)+8Kp+4Kp^2}$$

- Independent of n
- if p→0 it recovers the ring value
- if p→1 it well approximates 1

The Watts-Strogatz model

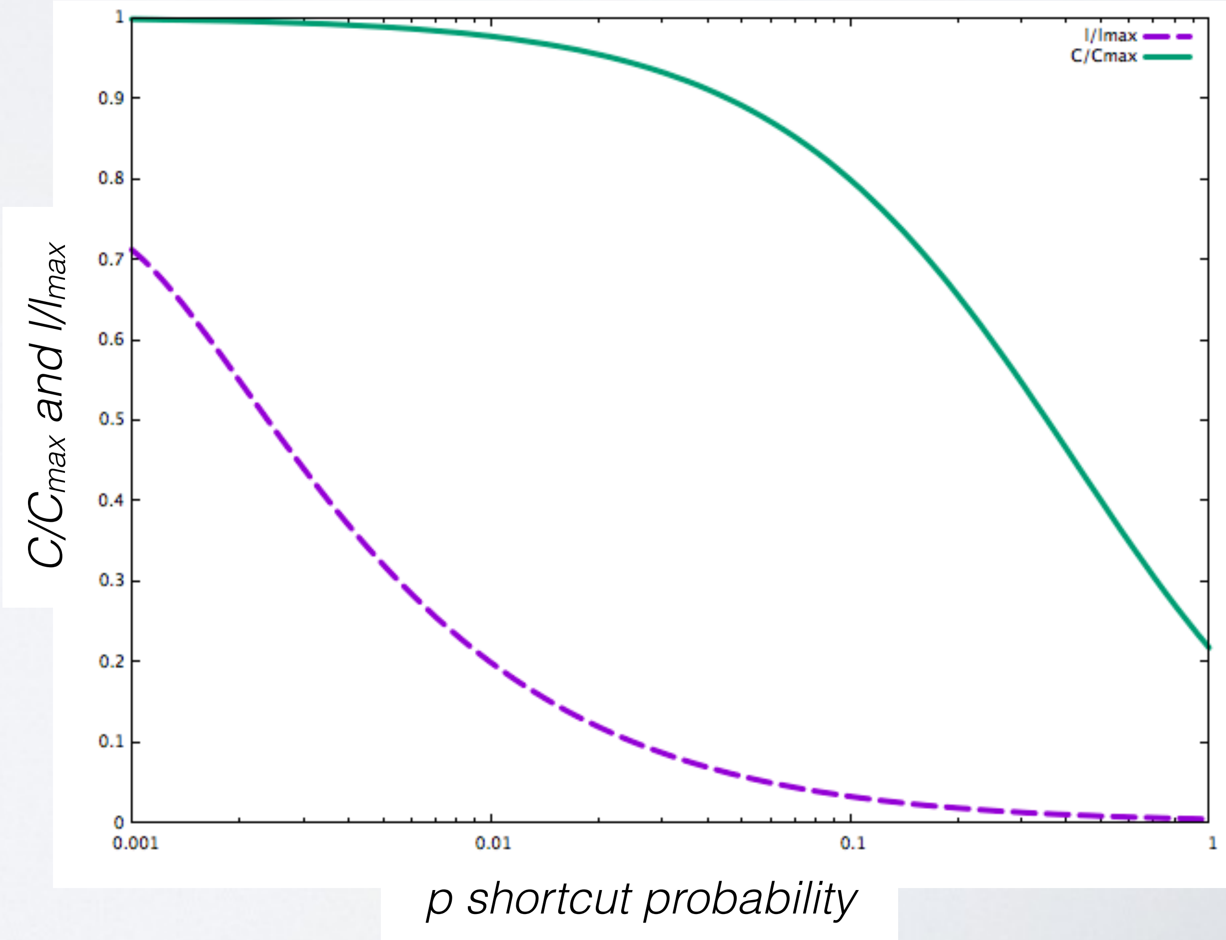


Average path length (Definition 2)

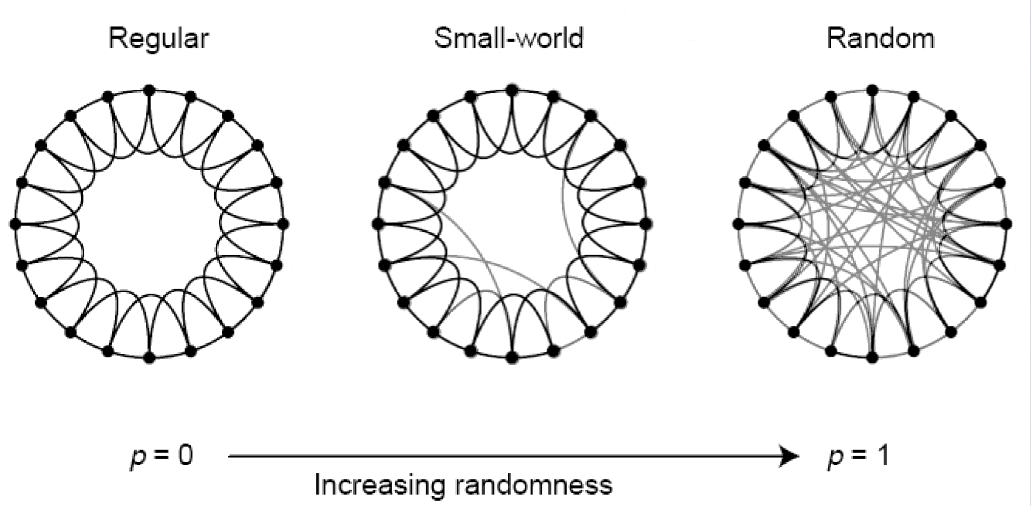
- No closed form solution

- From numerical simulations:

  - average path length $$\boxed{l = \frac{\ln(nKp)}{K^2 p}}$$



The Watts-Strogatz model



Degree distribution (Definition 2)

- p=0 - each node has the same degree K (Dirac delta
- p>0 - each node has degree K + shortcut links
  - Number of shortcut edges: $s = \frac{1}{2} NK \times p$
  - Each node will have on average *Kp* number of shortcuts

- The degree distribution is

$$\boxed{P(k) = e^{-Kp} \frac{(Kp)^{(k-K)}}{(k-K)!}}$$

if $k \geq K$ and $P(k)=0$ if $k<K$

- **p>0 - approximates a Poisson distribution just like a random network**

ER Random Network - catch up

Network	Degree distribution	Path length	Clustering coefficient
Real world networks	broad	short	large
Regular lattices	constant	long	large
ER random networks	Poissonian	short	small
Configuration Model	Custom, can be broad	short	small
Watts & Strogatz (in SW regime)	Poissonian	short	large