Chapter 3

Random Networks

Summary

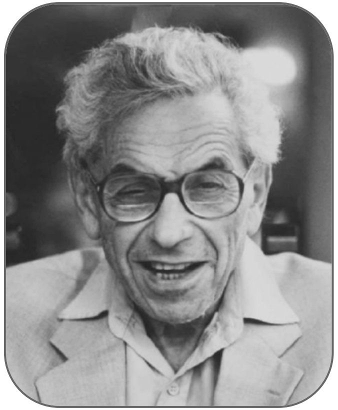
- Random Graphs
- Erdos-Renyi model
- Paths, Connectedness & Density
- Advanced: Configuration Model

Reading

- Chapter 3 of Barabasi's book

Random Graphs

The Erdős-Rényi
Random Graph model (ER)



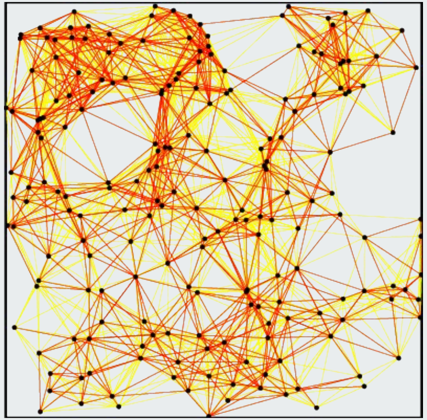
Pál Erdős
(1913-1996)

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)"*

[1] P. Erdős and A. Rényi.
On random graphs, I. Publicationes Mathematicae. 1959.
[2] P. Erdős and A. Rényi.
On the evolution of random graphs. Publ. Math. Inst. Hung. Acad. Sci., 1960.

Why using Random Graph models?



- 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

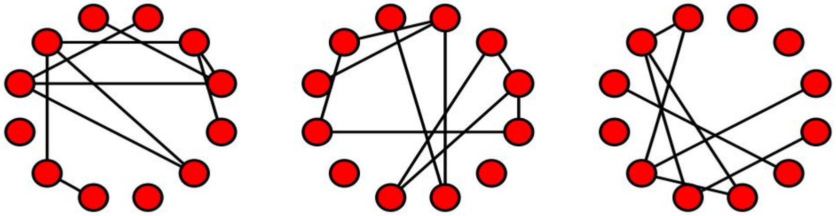
ER model

(General) Definition.
A random graph is a graph of **N** nodes where each pair of nodes is connected by probability **p**.

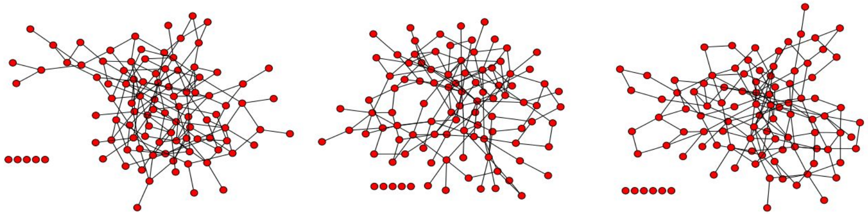
The **G(n,L)** definition.
1. Take **N** disconnected nodes
2. Add **L** edges uniformly at random

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

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



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)

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

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

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

Reminder (Binomial Coefficient)
Number of ways, disregarding order, that k objects can be chosen from among n objects

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

<L>: The average number of links in a random graph
<k>: The average degree (and its variance)

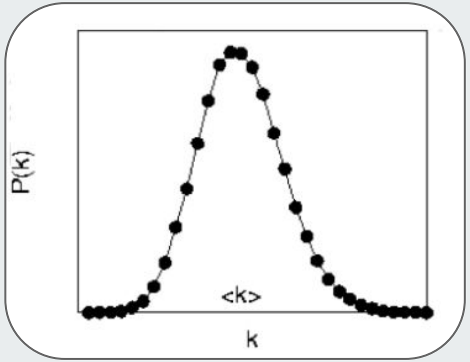
$$< L >= \sum_{l=0}^{\frac{N(N-1)}{2}} LP(L) = p\frac{N(N-1)}{2}$$

$$< k >= \frac{2L}{N} = p(N-1)$$

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

Degree Distribution

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



probability of having k edges

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

Select k nodes from N-1

probability of missing N-1-k edges

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>.

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

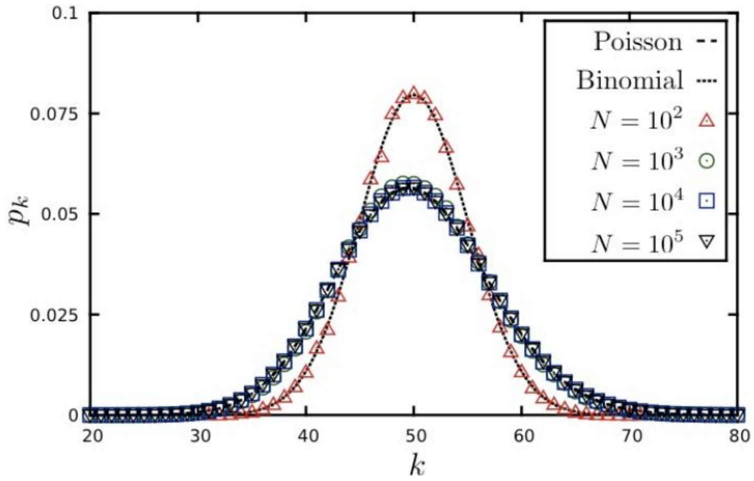
Characteristics:

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

Visual Simulation
http://www.networkpages.nl/CustomMedia/Animations/RandomGraph/ERRG/DegreeDistribution.html

Degree Distribution

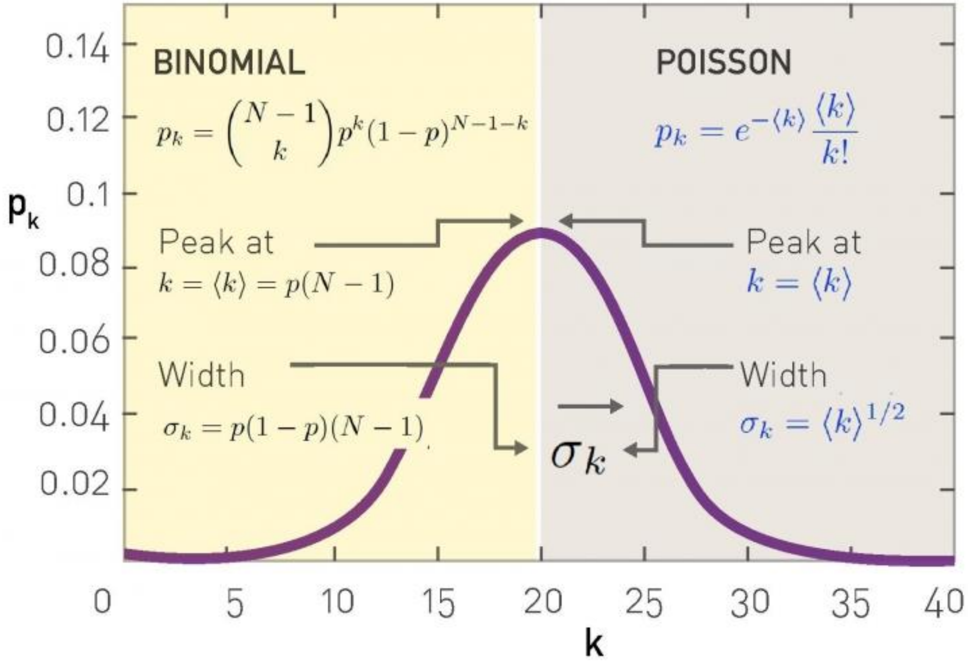
For large **N** and small **k** (p,L), we can approximate the degree distribution using a poisson distribution of parameter (mean) **λ = < k >**



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

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

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



Exact Result - Binomial distribution

Large N limit - Poisson distribution

Real Networks are not Poisson

Maximum & Minimum Degree

Let’s assume <k> =1,000, N=10^9.
We can derive the max and min degrees as follows:

$$N[1 - P(k_{\max})] \approx 1$$

$$\mathrm{I} - P(k_{\max}) = 1 - e^{-<k>} \sum_{k=0}^{k_{\max}} \frac{\langle k\rangle^k}{k!} = e^{-<k>} \sum_{k=k_{\max}+1}^{\infty} \frac{\langle k\rangle^k}{k!} \approx e^{-<k>} \frac{\langle k\rangle \langle k\rangle^{k_{\max}}}{(k_{\max}+1)!}$$

k_{max} = 1,185

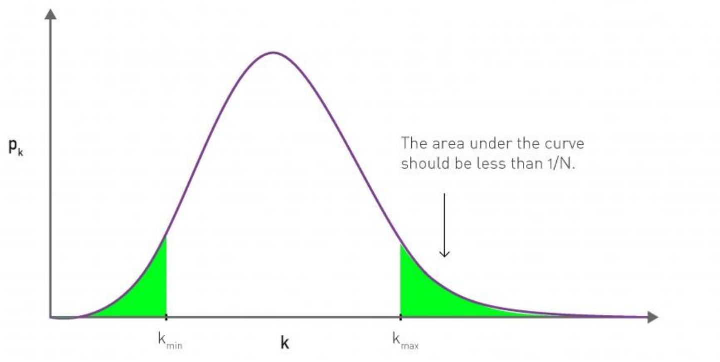
$$NP(k_{\min}) \approx 1$$

$$P(k_{\min}) = e^{-\langle k\rangle} \sum_{k=0}^{k_{\min}} \frac{\langle k\rangle^k}{k!}$$

k_{min} = 816

$$< \mathrm{K} > \pm \sigma_k \quad \sigma_k = < k >^{1/2}$$

$$\sigma_k = 31.62$$



No Outliers in a Random Society!

The most connected individual has degree k_{max}~1,185

The least connected individual has degree k_{min} ~ 816

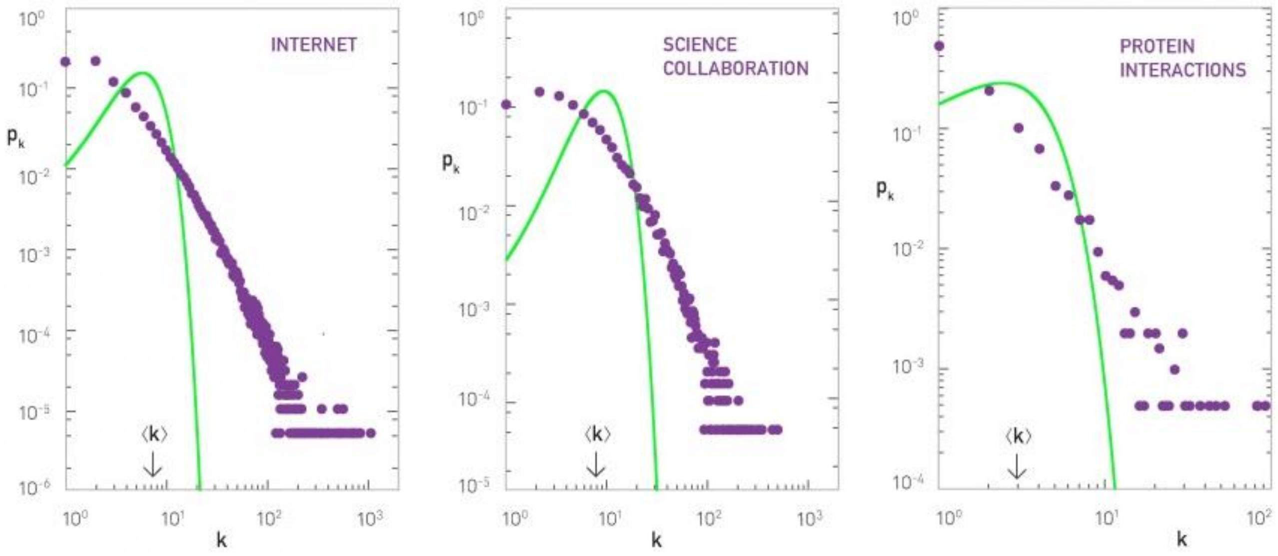
The probability to find an individual with degree k>2,000 is 10^{-27}.

Hence the chance of finding an individual with 2,000 acquaintances is so tiny that such nodes are virtually nonexistent in a random society.

- A random society would consist of mainly average individuals, with everyone with roughly the same number of friends.
- It would lack outliers, individuals that are either highly popular or recluse.

Facing Reality:

Degree distribution of real networks



Clustering & Distance

Clustering in Random Graphs

For fixed average degree, **C** is decreasing as **N** goes large

- Low clustering coefficient
- It is vanishing with the system size

Reminder (clustering coeff.)

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

where $\mathbf{n_i}$ is the number of links between the neighbours of node **i**

We know that $p = \dfrac{<k>}{n-1}$

thus,

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

Clustering

ER Graphs vs Real-World

ER

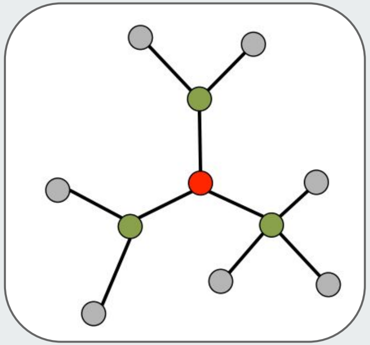
Expected Small Clustering Coefficient

$$C_i \equiv \frac{1}{N} < k >= p$$

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

Distance in 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 is O(log n)

Distance

ER Graphs vs Real-World

ER

Logarithmically short distance among nodes

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

Real-World Networks

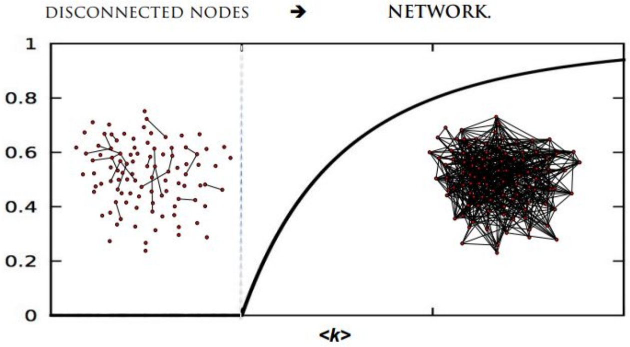
Network	Size	$\langle k \rangle$	ℓ	ℓ_{rand}	C	C_{rand}	Reference
WWW, site level, undir.	153 127	35.21	3.1	3.35	0.1078	0.00023	Adamic, 1999
Internet, domain level	3015–6209	3.52–4.11	3.7–3.76	6.36–6.18	0.18–0.3	0.001	Yook *et al.*, 2001a, Pastor-Satorras *et al.*, 2001
Movie actors	225 226	61	3.65	2.99	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

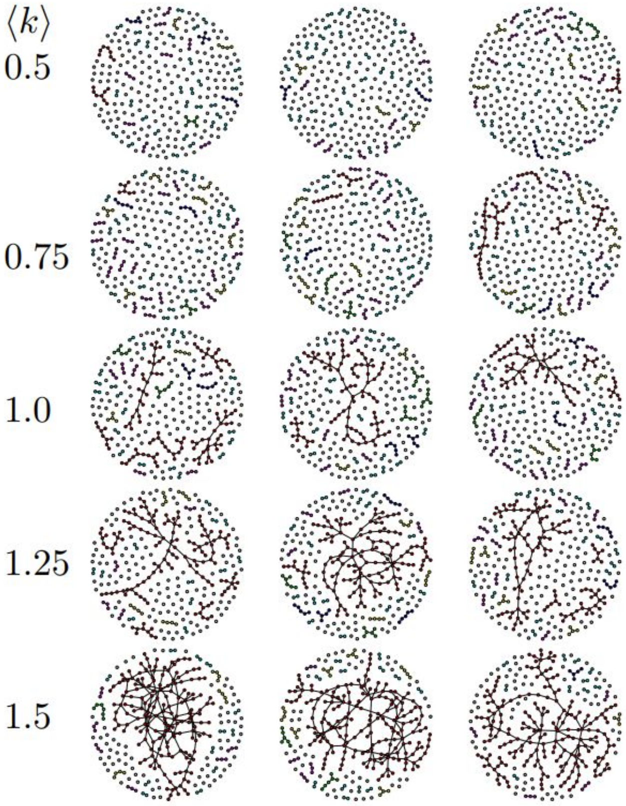
Connected Components

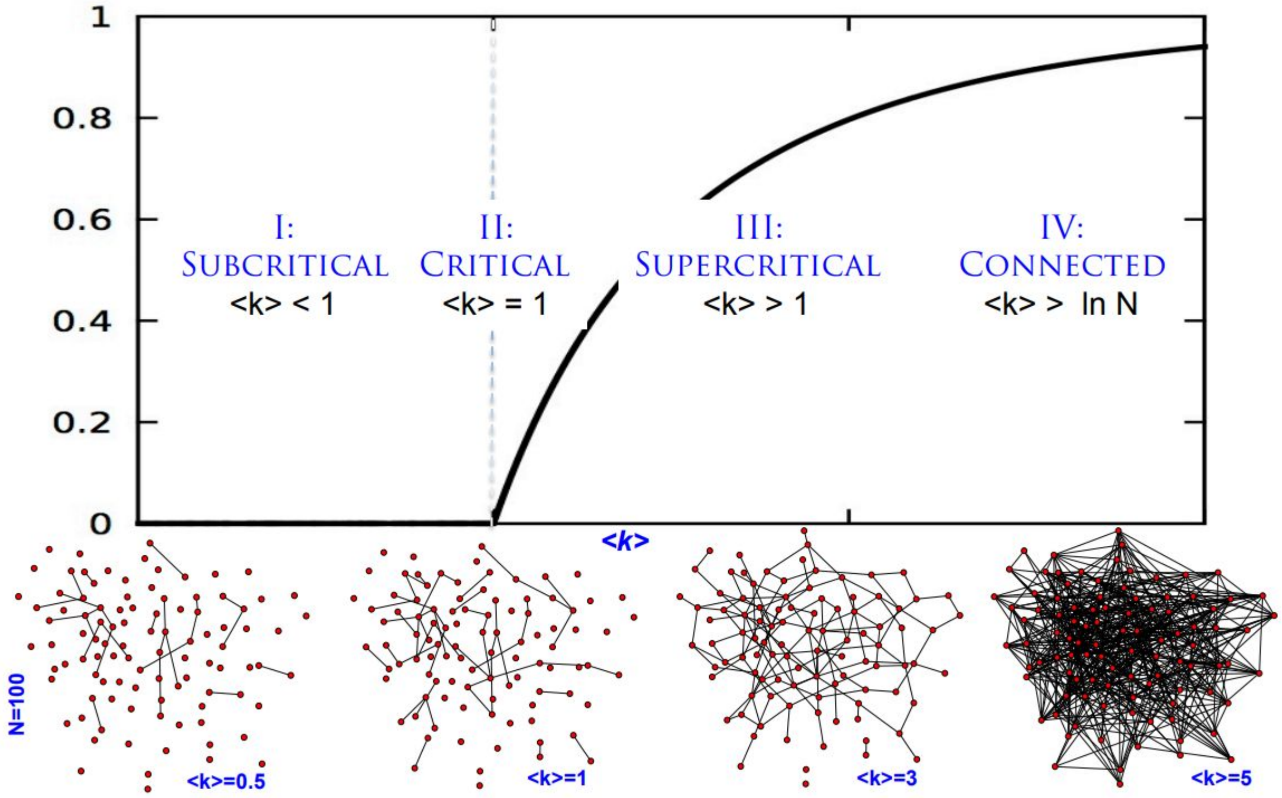
Random Graphs

Connected Components

Network structure goes through a transition.
How and when does this transition happen?







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

Network Regimes

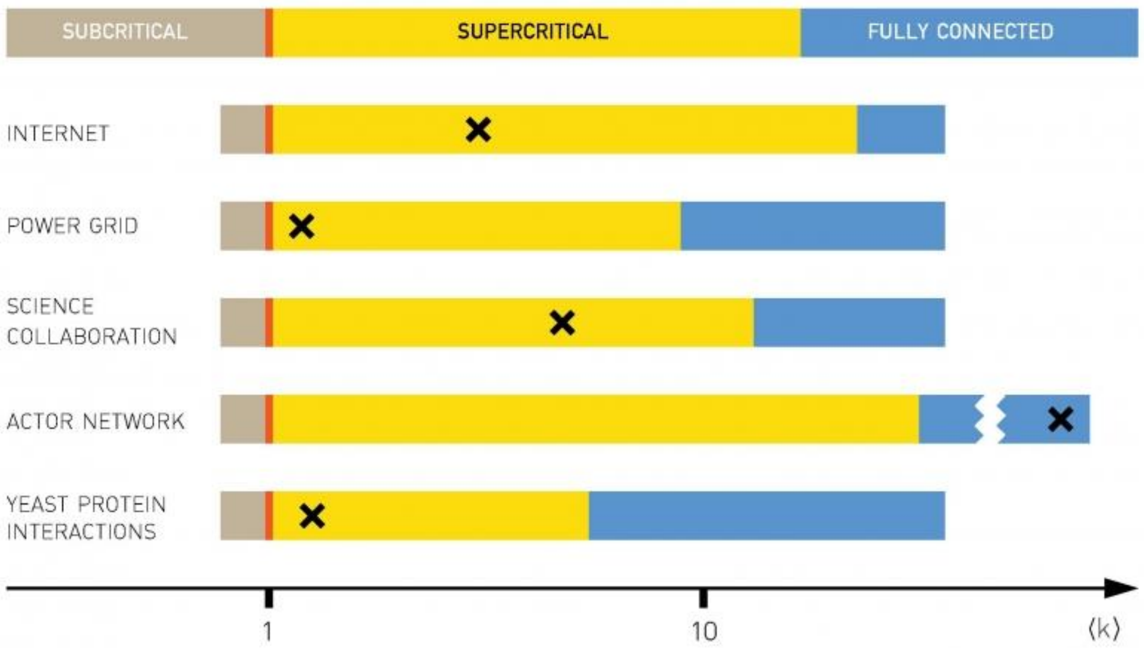
Subcritical ($\langle k \rangle < 1$, $p < p_c = 1/N$)
No giant component;
N-L isolated clusters, cluster size distribution is exponential;
The largest cluster is a tree, its size ~ ln N.

Supercritical ($\langle k \rangle > 1$, $p > p_c = 1/N$)
Unique giant component: $N_G \sim (p-p_c)N$;
GC has loops;
Cluster size distribution: exponential.

Critical ($\langle k \rangle = 1$, $p = p_c = 1/N$)
Unique giant component: $N_G \sim N^{2/3}$
Contains a vanishing fraction of all nodes, $N_G/N \sim N^{-1/3}$
Small components are trees, GC has loops.

Connected ($\langle k \rangle > \ln N$, $p > (\ln N)/N$)
Only one cluster: $N_G = N$;
GC is dense;
Cluster size distribution: None.

Visual Simulation
http://www.networkpages.nl/CustomMedia/Animations/RandomGraph/ERRG/AddoneEdgepATime.html



Network	N	L	‹k›	‹d›	d_{max}	lnN/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

Real Network are Supercritical

Summarizing...

Random Networks
in a Nutshell

Degree Distribution
(Poisson for large N)

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

Clustering
(vanishing for large size)

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

Path length
(distance with logarithmic relation to nodes)

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

More on distances in **Chapter 4**!

Network	Degree Distribution	Path Length	Clustering Coefficient
Real-world networks	Broad	Short	Large
ER graphs	Poissonian	Short	Small

ER model is not capturing the properties of any real system but it serves as a reference system for any other network model

Advanced Topic:

Configuration Model

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

Configuration Model

How many observed patterns are driven by the degrees alone?

Random Graphs with specified degrees:

Based on an observed network

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

Sampled from an ad-hoc degree distributions

Delta/Dirac function, Poisson, Scale-free

Global condition to satisfy (even degree sum):

$$\sum_i k_i \bmod 2 = 0$$

each edge has to have ending nodes

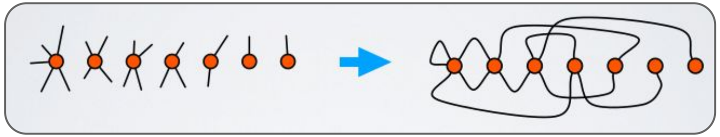
Configuration model

Molloy-Reed

Theory

Original idea:

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



Such process produces a configuration model that preserves the input degree sequence, allowing:

- multi-links,
- self-links

Configuration model

Molloy-Reed

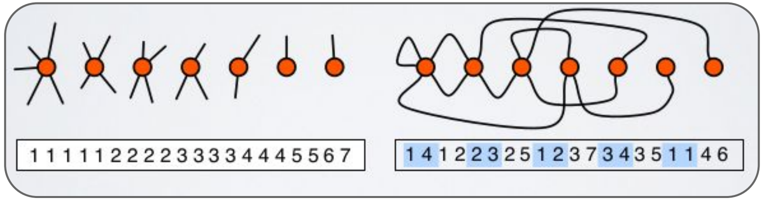
Practice

An effective algorithm

1. Take an array $\vec{v}$ with length 2m and fill it with ki indices of each node $i \in V$

2. Make a random permutation of the array $\vec{v}$

3. Read the content of the array as ordered pairs



Visual Simulation
http://www.networkpages.nl/CustomMedia/Animations/RandomGraph/CM/CmCreation.html

Configuration model

Properties

Clustering coefficient
(independent from network size)

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

Degree distribution
(of a randomly selected node's neighbor)

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

Average Degree
(of a randomly selected node's neighbor)

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

Network	Degree Distribution	Path Length	Clustering Coefficient
Real-world networks	Broad	Short	Large
ER graphs	Poissonian	Short	Small
Configuration model	Custom, can be broad	Short	Small

Chapter 3

Conclusion

Take Away Messages

1. ER model generates random graphs
2. In ER different values of *p* reflects different network regimes
3. Configuration models allow the generation of random graphs having heterogeneous degree distributions

Suggested Readings

- Chapter 3 of Barabasi's book

What's Next

Chapter 4:
It's a Small World!

Notebook

Chapter 3: Random Networks
https://github.com/sna-unipi/SNA_lectures_notebooks