Network models – random graphs

Properties common to many large-scale networks, independently of their origin and function:

1. The degree and betweenness distribution are decreasing functions, usually power-laws.
2. The distances scale logarithmically with the network size

$$l \approx \frac{\log N}{\log\langle k\rangle}$$

3. The clustering coefficient does not seem to depend on the network size

$$C \propto \langle k\rangle$$

As though all these networks were part of the same family/class.

Random networks

The average distance and clustering coefficient only depend on the number of nodes and edges in the network.

This suggests that general models based only on the number of nodes and edges in the network could be successful in describing the properties of an “expected” (characteristic) network.

Uniformly random network: distributes the edges uniformly among nodes.
Probabilistic interpretation:
There exists a set (ensemble) of networks with given number of nodes and edges. Select a random member of this set.
What are the expected properties of this network? – studied by random graph theory.

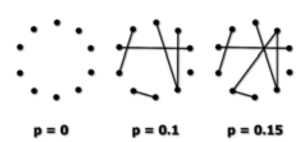
Ex. 1
Start with 10 isolated nodes. For each pair of nodes, throw with a dice, and connect them if the number on the dice is 1. Describe the graph you obtained. How many edges are in the graph? Is it connected or not? What is the average degree and the degree distribution?

Ex. 2
Now connect node pairs if the number on the dice is 1 or 2. How is the graph different from the previous case?

Ex. 2
How many edges do you expect a graph with N nodes would have if each edge is selected with throwing with a dice?

Random graph theory

Erdös-Rényi algorithm - Publ. Math. Debrecen 6, 290 (1959)



- fixed node number N
- connecting pairs of nodes with probability p

Expected number of edges: $E = p\frac{N(N-1)}{2}$

Random graph theory studies the expected properties of graphs with $N \to \infty$

The properties of random graphs depend on *p*

Properties studied:

- is the graph connected?
- does the graph contain a giant connected component?
- what is the diameter of the graph?
- does the graph contain cliques (complete subgraphs)?

Probabilistic formulation: what is the probability that a graph with *N* nodes and connection probability *p* is connected?
Increase *p* from 0 to 1. Some of these properties appear suddenly, at a threshold $p_c(N)$

$$\lim_{N\to\infty} P_{N,p}(Q) = \begin{cases} 0 \text{ if } \dfrac{p(N)}{p_c(N)} \to 0 \\ 1 \text{ if } \dfrac{p(N)}{p_c(N)} \to \infty \end{cases}$$

Subgraphs of a random graph

Consider a subgraph with *n* nodes and *e* edges.
Expected number of of these subgraphs in a graph with N nodes and connection prob. *p*

$$E(X) = C_N^n p^e \frac{n!}{a}$$

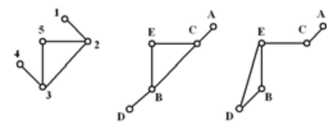
We can permute the *n* nodes in any way we want...

Ways of selecting *n* nodes from *N*

Probability of having *e* edges

but identical (isomorphic) graphs do not count

Isomorphic graphs: there exists a one-to-one mapping of the nodes in such a way that if (and only if) node i and j are connected in one then their images i' and j' are also connected.



Special subgraphs

Consider a subgraph with *n* nodes and *e* edges.
Expected number of subgraphs with *n* nodes and *e* edges in a graph with *N* nodes and connection prob. *p*

$$E(X) = C_N^n p^e \frac{n!}{a} \cong \frac{N^n p^e}{a}$$

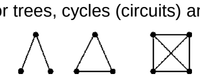
If the connection probability is a function of the number of the nodes, we can find the condition of having a non-vanishing number of subgraphs.

$$\lim_{N\to\infty} p(N) N^{n/e} \neq 0$$

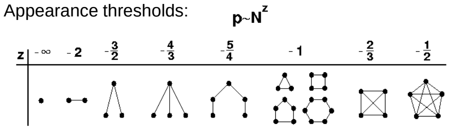
Ex. Find the condition of having a non-vanishing number of trees, cycles and completely connected subgraphs.

Evolution of a random graph

Assume that the connection probability is a power-law of N, $p = cN^z$
Assume that z increases from $-\infty$ to 0
Look for trees, cycles (circuits) and cliques in the graph.



Appearance thresholds: $p \sim N^z$

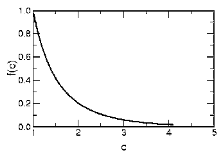


The graph contains cycles of any length if $z \geq -1$

Clusters in a random graph

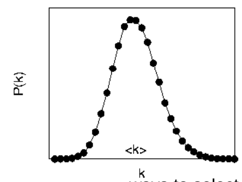
- For $p < N^{-1}$ the graph contains only isolated trees.
- If $p = cN^{-1}$ with $c < 1$ the graph has isolated trees and cycles.
- At $p = cN^{-1}$ with $c = 1$ a giant connected component appears.
- The size of the giant connected component approaches N rapidly as c increases.

$$S = (f(1) - f(c))N$$



- The graph becomes connected if $\lim_{N\to\infty} \frac{p}{\ln N / N} = \infty$

Node degrees in random graphs



- average degree: $\langle k \rangle = \frac{2E}{N} \cong pN$
- degree distribution:

$$P(k) \cong C^{k}_{N-1} p^{k} (1-p)^{N-1-k}$$

ways to select k nodes from $N-1$

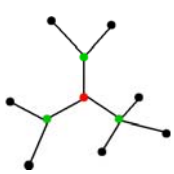
probability of having k edges

probability of missing $N-1-k$ edges

Most of the nodes have approximately the same degree.
The probability of very highly connected nodes is exponentially small.

Distances in random graphs

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



- nr. of first neighbors: $N_1 \cong \langle k \rangle$
- nr. of second neighbors: $N_2 \cong \langle k \rangle^2$
- estimate maximum distance:

$$1 + \sum_{l=1}^{l_{max}} \langle k \rangle^{i} = N \Rightarrow l_{max} = \frac{\log N}{\log \langle k \rangle}$$

This scaling was proven by Chung and Lu, Adv. Appl. Math 26, 257 (2001).

There is no local order in random graphs

Measure of local order: $C_i \equiv \frac{n_i}{k_i(k_i-1)/2}$

Since edges are independent and have the same probability p,

$$n_i \cong p\frac{k_i(k_i-1)}{2} \Rightarrow C \cong p = \frac{<k>}{N}$$

The clustering coefficient of random graphs is small.

Are real networks like random graphs?

As quantitative data about real networks becomes available, we can compare their topology with that of random graphs.
Starting measures: N, <k> for the real network.
Determine *l*, C and P(k) for a random graph with the same N and <k>.

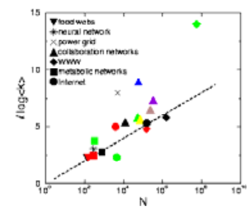
$$l_{rand} \approx \frac{\log N}{\log\langle k\rangle} \qquad C_{rand} = p = \frac{\langle k\rangle}{N}$$

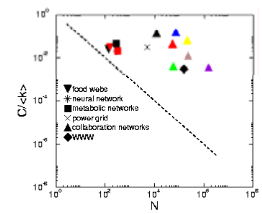
$$P_{rand}(k) \cong C^{k}_{N-1} p^{k} (1-p)^{N-1-k}$$

Measure *l*, C and P(k) for the real network. Compare.

Path length and order in real networks

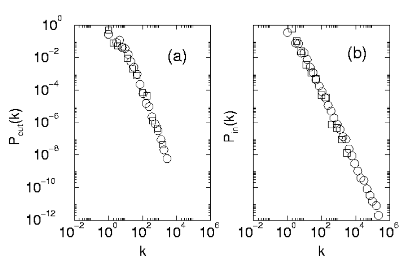
$$l_{rand} = \frac{\log N}{\log\langle k\rangle} \qquad C_{rand} = \frac{\langle k\rangle}{N}$$





Real networks have short distances like random graphs yet show signs of local order.

The degree distribution of the WWW is a power-law



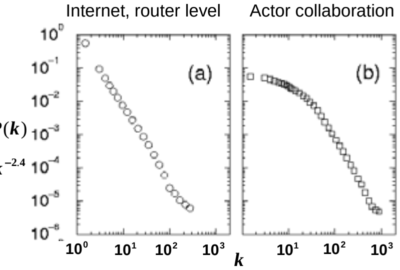
$$P_{out}(k) \approx k^{-2.45}$$
$$P_{in}(k) \approx k^{-2.1}$$

R. Albert, H. Jeong, A.-L. Barabási, Nature 401, 130 (1999)

A. Broder *et al.*, Comput. Netw. 33, 309 (1999)

Power-law degree distributions were found in diverse networks

Internet, router level   Actor collaboration

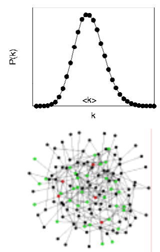


$P(k) \approx k^{-2.4}$ $P(k) \approx k^{-2.3}$

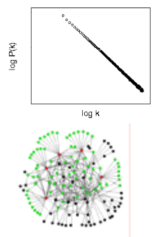
R. Govindan, H. Tangmunarunkit, IEEE Infocom (2000)

A.-L. Barabási, R. Albert, Science 286, 509 (1999)

The power-law degree distribution indicates a heterogeneous topology



The average degree gives
the characteristic scale (value)
of the degree.



Large variability,
the average degree not informative,
no characteristic scale for the degree
Scale-free

Idea: generate random graphs with a power-law degree distribution

Fixed $N, \quad P(k) = Ak^{-\gamma}, \; k < K$

Network assembly - random edges, but enforcing the right $P(k)$

$$\sum_{k=1}^{K} P(k) = 1, \;\Rightarrow\; A = \frac{1}{\sum_1^K k^{-\gamma}}$$

$$<k> = \sum_{k=1}^{K} kP(k), \;\Rightarrow\; <k> = \frac{\sum_1^K k^{-\gamma+1}}{\sum_1^K k^{-\gamma}}$$

The number of edges increases as γ decreases.

Constructing graphs with a given degree distribution

Configuration model:
- choose a degree sequence N(k)=N P(k)
- give the nodes k "stubs" according to N(k)
- connect stubs randomly

M. E. J. Newman, S. H. Strogatz, and D. J. Watts,
Phys. Rev. E 64, 026118 (2001)

Ex. Construct a graph with 10 nodes and degree sequence
N(1)=4, N(2)=3, N(3)=2, N(4)=1.
What is a necessary condition for the graph construction?

Theory of general random graphs

Looks at a characteristic member of the ensemble of graphs with given degree distribution.

Seeks the answers to the same questions as random graph theory
- is the graph connected?
- does the graph contain a giant component?
- what is the diameter of the graph?
- what is the clustering coefficient of the graph?

The theoretical concept needed for the analysis is the generating function.

One important result: A giant connected component exists if the graph is sufficiently heterogeneous. $\langle k^2 \rangle / \langle k \rangle \geq 2$

Connectivity of scale-free random graphs

Given: $N, \quad P(k) \approx k^{-\gamma}$ for $k \leq \kappa$

Graph properties depend on the degree exponent γ

- giant cluster: $\gamma \leq 3.47$
- connected: $\gamma \leq 2$

W. Aiello, F. Chung, L. Lu, Proc. 32th ACM Theor. Comp., 171 (2000)

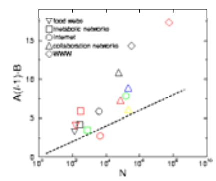
M. E. J. Newman, S. H. Strogatz, D. J. Watts, Phys. Rev. E 64, 026118 (2001)

Average path length of scale-free random graphs

Network: $N, \quad P(k) \approx k^{-\gamma}$ for $k \leq \kappa$

Prediction: $l_{sf} = \dfrac{\ln N + B}{A} + 1 \qquad A, B = f(\gamma, \kappa)$

M. E. J. Newman, S. H. Strogatz, D. J. Watts, Phys. Rev. E 64, 026118 (2001)



- qualitative agreement
- worse than a random graph

Clustering coefficient of scale-free random graphs

$$C = \frac{<k>}{N}\left(\frac{<k^2> - <k>^2}{<k>^2}\right)^2$$

The second term depends on the variance of the degree distribution.

$P(k) \approx k^{-\gamma} \qquad C \approx N^{-(3\gamma-7)/(\gamma-1)}$

For $\gamma < 7/3$ C increases with N.

M. E. J. Newman, SIAM Review 45, 167 (2003)

Network	Size	$\langle k \rangle$	κ	γ_{out}	γ_{in}	ℓ_{real}	ℓ_{rand}	ℓ_{pow}	Reference	Nr.
WWW	325 729	4.51	900	2.45	2.1	11.2	8.32	4.77	Albert, Jeong, and Barabási 1999	1
WWW	4×10^7	7		2.38	2.1				Kumar *et al.*, 1999	2
WWW	2×10^8	7.5	4000	2.72	2.1	16	8.85	7.61	Broder *et al.*, 2000	3
WWW, site	260 000				1.94				Huberman and Adamic, 2000	4
Internet, domain*	3015–4389	3.42–3.76	30–40	2.1–2.2	2.1–2.2	4	6.3	5.2	Faloutsos, 1999	5
Internet, router*	3888	2.57	30	2.48	2.48	12.15	8.75	7.67	Faloutsos, 1999	6
Internet, router*	150 000	2.66	60	2.4	2.4	11	12.8	7.47	Govindan, 2000	7
Movie actors*	212 250	28.78	900	2.3	2.3	4.54	3.65	4.01	Barabási and Albert, 1999	8
Co-authors, SPIRES*	56 627	173	1100	1.2	1.2	4	2.12	1.95	Newman, 2001b	9
Co-authors, neuro.*	209 293	11.54	400	2.1	2.1	6	5.01	3.86	Barabási *et al.*, 2001	10
Co-authors, math.*	70 975	3.9	120	2.5	2.5	9.5	8.2	6.53	Barabási *et al.*, 2001	11
Sexual contacts*	2810			3.4	3.4				Liljeros *et al.*, 2001	12
Metabolic, *E. coli*	778	7.4	110	2.2	2.2	3.2	3.32	2.89	Jeong *et al.*, 2000	13
Protein, *S. cerev.**	1870	2.39		2.4	2.4				Jeong, Mason, *et al.*, 2001	14
Ythan estuary*	134	8.7	35	1.05	1.05	2.43	2.26	1.71	Montoya and Solé, 2000	14
Silwood Park*	154	4.75	27	1.13	1.13	3.4	3.23	2	Montoya and Solé, 2000	16
Citation	783 339	8.57			3				Redner, 1998	17
Phone call	53×10^6	3.16		2.1	2.1				Aiello *et al.*, 2000	18
Words, co-occurrence*	460 902	70.13		2.7	2.7				Ferrer i Cancho and Solé, 2001	19
Words, synonyms*	22 311	13.48		2.8	2.8				Yook *et al.*, 2001b	20

Expectations:

$\langle k \rangle \geq 1$ giant connected component, $\langle k \rangle \geq \ln N$ connected

$\gamma \leq 3.47$ giant connected component, $\gamma \leq 2$ connected

Exponential random graphs

“Exponential” does not refer to the degree distribution but to the model construction!
This is a statistical method for generating a of graph with N nodes by specifying a distribution function over all graphs with N nodes.

1. Select a set of informative network measures (e.g. number of edges, number of triangles, degree distribution)
2. Then select a network from the ensemble of all graphs using the probability

$$P(G) \sim exp\left(-\sum_i \beta_i \varepsilon_i\right)$$

β_i – parameters, ε_i – network measures

3. Estimate the coefficients such that an observed (real) network corresponds to the most likely graph in that ensemble – maximum likelihood estimation

Markov graphs: edges that do not share a node are independent
Further reading: Frank & Strauss 1986, David Hunter’s webpage