5 Network Models

Networks: An Introduction,
M. E. J. Newman, Oxford University Press (2010), Chapter 12–15.

5.1 Random graphs

Real world networks have distinctive structural properties.

- Often details are not known, only averages.
- Models can inform about importance of features (for dynamics etc.)

"Scale free" degree distributions

Degree distribution: probability that random node has degree k: $p_k = f(k)$.
Empirically: real world nets often scale free:

$$p_k \sim k^{-\alpha} \quad \text{or cumulatively:} \tag{1}$$

$$P_k = \sum_{k'=k}^{\infty} p'_k \tag{2}$$

$$\sim k^{-(\alpha-1)} \tag{3}$$

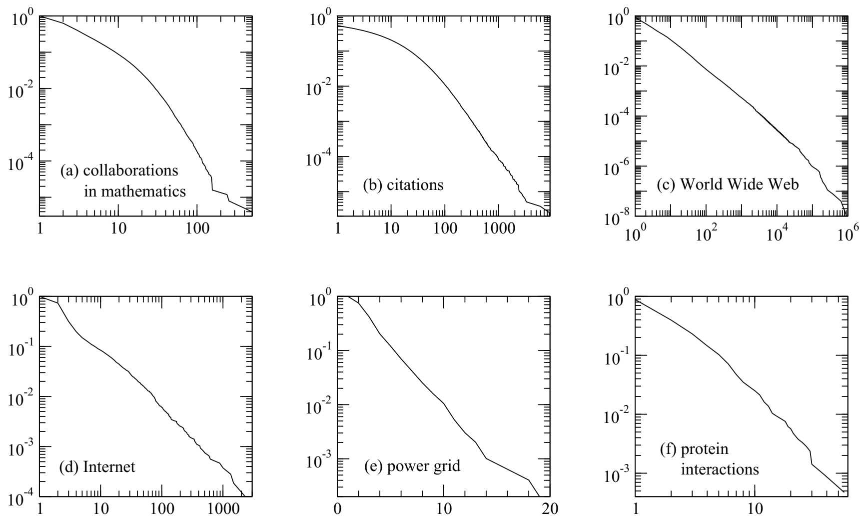


Figure 1: Cumulative degree distributions showing scale-free characteristics.

"Small world" property

Average shortest path-length $\bar{l}$ between node-pairs:

$$\bar{l} = \frac{1}{\frac{1}{2}N(N-1)} \sum_{i \geq j} d_{ij} \tag{4}$$

- For regular D-dimensional lattices: $\bar{l} \propto N^{\frac{1}{D}}$

Data shows a short path usually exists between pairs of nodes:

- for people ($N \approx 7 \times 10^9$) measure (Milgram etc.): $\bar{l} \sim 6$.

High clustering

Real world transitivity ($c = \frac{\#(\text{triangles})}{\#(\text{connected triples})}$) much higher than random chance. Also define average local clustering coefficient:

$$c_i = \frac{\#(\text{triangles on } i)}{\#(\text{connected triples on } i)} \tag{5}$$

Assortativity

Correlation ("de-mixing") occurs between some property of adjacent nodes.

- variable $x_i = x_j$ or degree $k_i \approx k_j$

Can mesure Pearson correlation coefficient:

$$r = \sum_{xy} \frac{xy(e_{xy} - a_x b_y)}{\sigma_x \sigma_y}, \tag{6}$$

where e_{xy} is fraction of edges between x and y types, a_x and b_y: fraction of x, x and y, y edges resp; σ are st. dev. of a and b. For degree distributions this becomes:

$$r = \frac{\sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m} \right) k_i k_j}{\sum_{ij} \left(k_i \delta_{ij} - \frac{k_i k_j}{2m} \right) k_i k_j} \tag{7}$$

- Correlated network: $r \to 1$
- Uncorrelated network: $r = 0$
- Anti-correlated network: $r \to -1$

In real world correlations or anti-correlations occur, rather than random uncorrelated (see table).

	network	type	n	m	z	ℓ	α	$C^{(1)}$	$C^{(2)}$	r	Ref(s).
social	film actors	undirected	449 913	25 516 482	113.43	3.48	2.3	0.20	0.78	0.208	20, 416
	company directors	undirected	7 673	55 392	14.44	4.60	–	0.59	0.88	0.276	105, 323
	math coauthorship	undirected	253 339	496 489	3.92	7.57	–	0.15	0.34	0.120	107, 182
	physics coauthorship	undirected	52 909	245 300	9.27	6.19	–	0.45	0.56	0.363	311, 313
	biology coauthorship	undirected	1 520 251	11 803 064	15.53	4.92	–	0.088	0.60	0.127	311, 313
	telephone call graph	undirected	47 000 000	80 000 000	3.16		2.1				8, 9
	email messages	directed	59 912	86 300	1.44	4.95	1.5/2.0		0.16		136
	email address books	directed	16 881	57 029	3.38	5.22	–	0.17	0.13	0.092	321
	student relationships	undirected	573	477	1.66	16.01	–	0.005	0.001	−0.029	45
	sexual contacts	undirected	2 810				3.2				265, 266
information	WWW nd.edu	directed	269 504	1 497 135	5.55	11.27	2.1/2.4	0.11	0.29	−0.067	14, 34
	WWW Altavista	directed	203 549 046	2 130 000 000	10.46	16.18	2.1/2.7				74
	citation network	directed	783 339	6 716 198	8.57		3.0/–				351
	Roget's Thesaurus	directed	1 022	5 103	4.99	4.87	–	0.13	0.15	0.157	244
	word co-occurrence	undirected	460 902	17 000 000	70.13		2.7		0.44		119, 157
technological	Internet	undirected	10 697	31 992	5.98	3.31	2.5	0.035	0.39	−0.189	86, 148
	power grid	undirected	4 941	6 594	2.67	18.99	–	0.10	0.080	−0.003	416
	train routes	undirected	587	19 603	66.79	2.16	–		0.69	−0.033	366
	software packages	directed	1 439	1 723	1.20	2.42	1.6/1.4	0.070	0.082	−0.016	318
	software classes	directed	1 377	2 213	1.61	1.51	–	0.033	0.012	−0.119	395
	electronic circuits	undirected	24 097	53 248	4.34	11.05	3.0	0.010	0.030	−0.154	155
	peer-to-peer network	undirected	880	1 296	1.47	4.28	2.1	0.012	0.011	−0.366	6, 354
biological	metabolic network	undirected	765	3 686	9.64	2.56	2.2	0.090	0.67	−0.240	214
	protein interactions	undirected	2 115	2 240	2.12	6.80	2.4	0.072	0.071	−0.156	212
	marine food web	directed	135	598	4.43	2.05	–	0.16	0.23	−0.263	204
	freshwater food web	directed	92	997	10.84	1.90	–	0.20	0.087	−0.326	272
	neural network	directed	307	2 359	7.68	3.97	–	0.18	0.28	−0.226	416, 421

Figure 2: (TABLE II from Newman) The properties measured are: type of graph, directed or undirected; total number of vertices n (N); total number of edges m; mean degree z ($\bar{k}$); mean vertexvertex distance l ($\bar{l}$); exponent α of degree distribution (if power law); clustering coefficient $C(1)$ (c); clustering coefficient $C(2)$; and degree correlation coefficient r. (Citation(s) from Newman).

5.2 The Erdős Renyí model

Simplest "random graph" model: assign all edges randomly: 2 ways to construct:

1. G^{ER}_{NK}: Create N nodes and randomly allocate K edges.
   - randomly assign $2K$ elements $A_{ij,i>j} = A_{ji,i<j} = 1$.
2. G^{ER}_{Np}: For each pair of nodes i and j, assign edge with prob. p.
   - G^{ER}_{NK} will appear from ensemble with prob. $p^K(1-p)^{N(N-1)/2-k}$

Properties of ER model:

Can solve for many features:

Short path lengths: trivial for random case: each step gain (on average) $\bar{k}$ neighbours, #(nei at dist $l \sim \bar{k}^l$), hence for whole network $N \sim \bar{k}^l$:

$$\bar{l} \sim \frac{\log N}{\log \bar{k}}. \tag{8}$$

- for world population: $\bar{l} \sim \frac{\log(\approx 7\times 10^9)}{\log(\sim 50)} \approx 6$.

Clustering: c is random, independent of other edges:

$$c = p = \frac{\bar{k}}{n-1}, \quad \to 0 \text{ for } N \to \infty \tag{9}$$

For world pop. $c \approx 7 \times 10^{-7}$, but actual measured: 0.01–0.5

Degree distribution : Prob. vertex having degree k (holding $\bar{k}$ const.):

$$p(k) = \binom{N}{k} p^k (1-p)^{N-k} \approx \frac{\bar{k}^k \exp(-\bar{k})}{k!}, \tag{10}$$

exact for $N \to \infty$

- **Not** scale-free.
  (also uncorrelated)

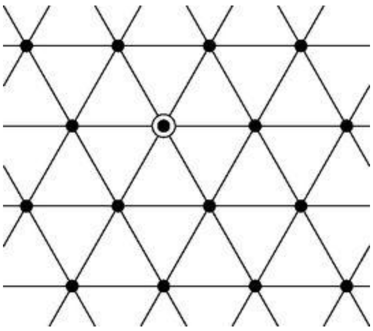
5.3 The configuration model

- Define desired degree distribution p_k,
- Assign m edges to N nodes by either:
  1. Assign k_i stubs (drawn from p_k) to each node i,
     - randomly join stubs
  2. assign k_i "desired degree" then:
     - $m = \frac{1}{2}\sum_h k_h$ edges placed with prob. $p \propto k_i k_j$.

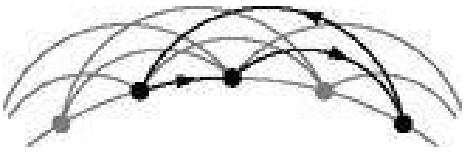
Features: can choose degree distribution, path lengths short as in ER, can derive (Newman) $c = \frac{1}{N}\frac{[<k^2>-<k>]^2}{<k>^3}$ which can be large in special cases, but not general.

5.4 The Watts Strogatz model

Regular lattices can have high clustering:



$c = 0.4$: comparable to real world
Can vary c by changing $\tilde{k}$ on a 1D lattice:



$$c = \frac{3(\tilde{k} - 2)}{4(\tilde{k} - 1)} \tag{11}$$

But: path-lengths are long: $\bar{l} \sim \frac{N}{2\tilde{k}}$.
Make model "small world" by introducing short-cuts with prob p:

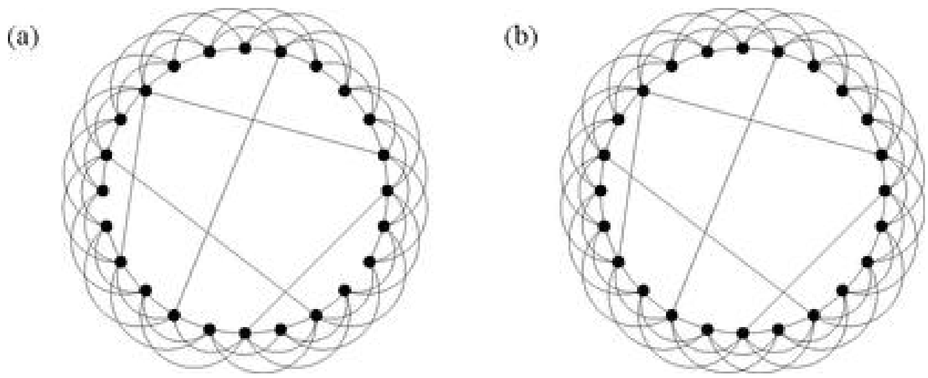


Figure 3: (a) Watts Strogatz, (b) Newman Watts Strogatz

Can either replace existing links or just add links.
Mixture of regular and random networks - contains features of both.
Difficult to treat analytically (2nd model is easier), but numerically observe range of p with clustering and short $\bar{l}$.
However, the degree distribution isn't scale free, e.g. for NWS:

$$p_k \sim e^{-\tilde{k}p} \frac{(\tilde{k}p)^{k-\tilde{k}}}{(k - \tilde{k})!} \tag{12}$$

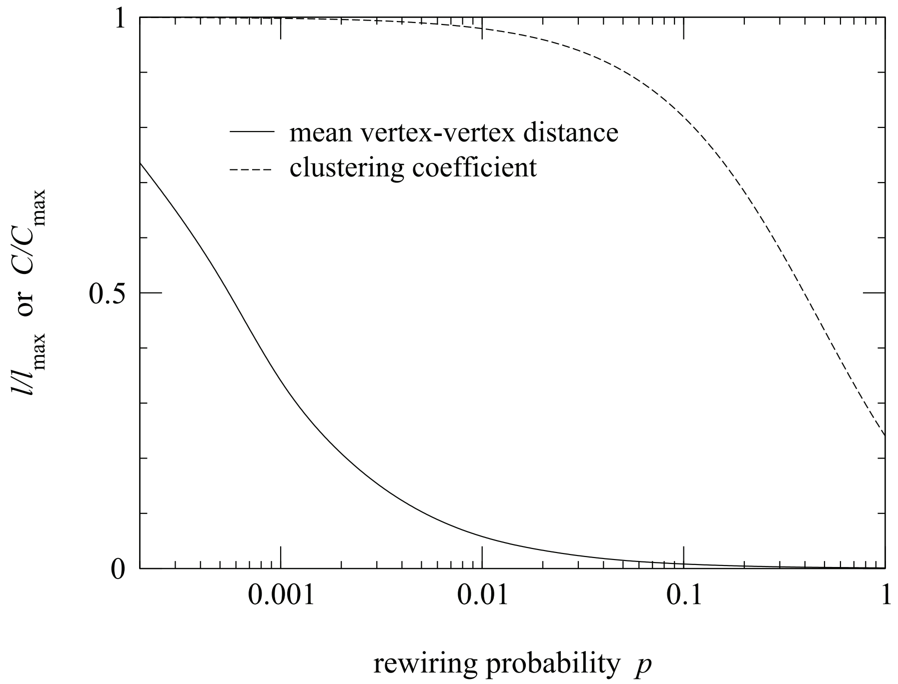


Figure 4: Mean path-length and clustering for different rewiring in WS model.

5.5 The Barabási–Albert preferential attachment model

Models can try to explain features, e.g. scale-free degree distribution.

Barabási and Albert proposed a rich-get-richer model.

- start with $\tilde{k}$ connected seed nodes
- at each generation step a new node is added with edges to $\tilde{k}$ others.
- Probability of i connecting to j $p \propto k_j$

Degree distribution: Find empirically:

$$p_k \sim k^{-3} \tag{13}$$

Non-linear BA model

- Probability of i connecting to j $p \propto k_j^{\gamma}$

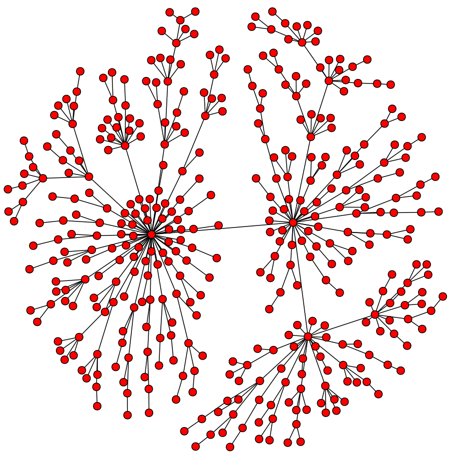


Figure 5: A Barabasí-Albert *preferential attachment* graph which has a power-law degree distribution.