HUDM5124 Topic 15:
network models of social structure

Overview:

- Basic ideas
- overview of methods
- some applications

Some References: Social networks

Burt, R. S. (1980). Models of network structure. *Annual Review of Sociology, 6,* 79-141.

**Scott, J. (1991). *Social Network Analysis*. Newbury Park CA: Sage.

**Ronald Breiger, Kathleen Carley, and Philippa Pattison (2003). *Dynamic Social Network Modeling and Analysis: Workshop summary and papers, Workshop on Dynamic Social Network Modeling and Analysis (2002 : Washington, D.C.)* [electronic resource] Washington, D.C. : National Academies Press, c2003.

***Social Networks* (journal)

Watts, D. J. (1998). Collective dynamics of ‘small-world’ networks. *Nature,* 6(393), 440-442.

Barabási, B. A. L., & Bonabeau, E. (2003). Scale-free networks. *Scientific American*. May 2003, 50-59.

**Watts, D. J. (2004). The “new” science of networks. *Annual Review of Sociology, 30*, 243-270.

**Weng, L., Flammini, A., Vespignani, A., & Menczer, F. (2012). Competition among memes in a world with limited attention. *Scientific Reports, 2* (doi:10.1038/srep00335).

Nobi Hanaki, Alex Peterhansl, Peter S. Dodds, and Duncan J. Watts (2007). Cooperation in evolving social networks. *Management Science, 7*.

Gueorgi Kossinets and Duncan J. Watts (2006). Empirical analysis of evolving social networks. *Science,*

Memon, Nasrullah (2010). *Data Mining for Social Network Data* [electronic resource]. New York : Springer.

Ajith Abraham, Aboul-Ella Hassanien, Vaclav Snasel (2010). *Computational social network analysis: Trends, tools and research advances [electronic resource].* Dordrecht: Springer.

Martin G. Everett, & David Krackhardt (2012). A second look at Krackhardt’s graph theoretical dimensions of informal organizations. *Social Networks, 34,* 159–163.

A little graph theory:

- For social networks, structure is about *connections*, not shared features or (purely) spatial position
- Social networks are naturally represented a graph, with nodes representing actors and lines between nodes (arcs) representing social connections
- A set of *nodes* (vertices) and *arcs* or lines (edges) defined on the power set of V (=VxV) constitute a graph:

  $$G = \langle V, E \rangle$$

- Clique = a completely connected subgraph of G
- If the arcs are directional, the structure is referred to as a *directed graph*
- *Diameter* of a network = max length of shortest path between any two nodes

Communication, social network, association data may be different:

- Structure is about *connections*, not similarity, shared features or (purely) spatial position
- *Connections* are naturally represented as *edges* between nodes (*nodes* represent *actors o*r other entities)

Example 1: asymmetric social nominations data: (from Corter, 1996)

TABLE 3: Summed number of nominations of column departments by respondents in row department.

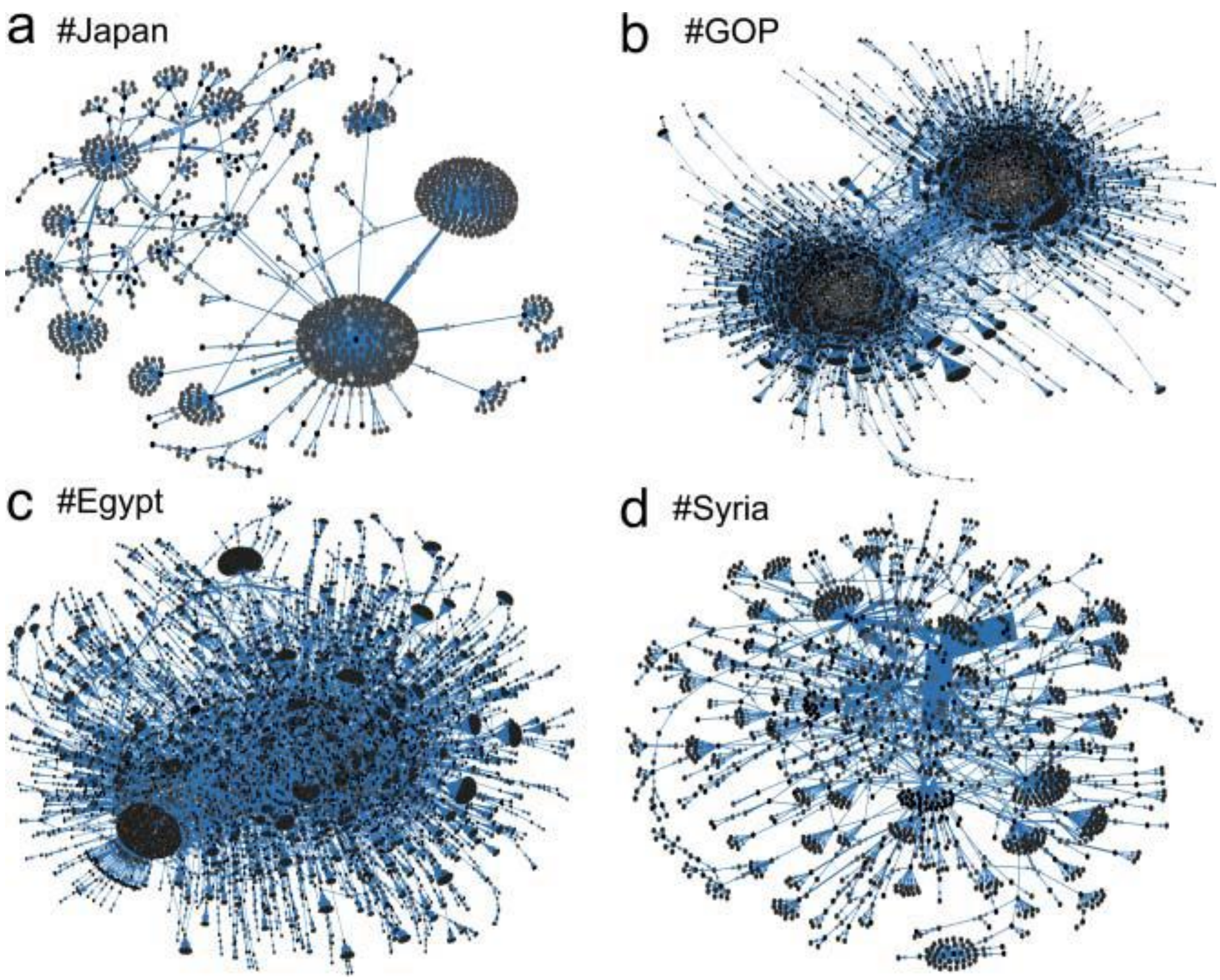
dept:	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	n
01	0	0	1	2	1	1	0	0	2	1	1	1	0	2	0	1	0	2
02	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	2
03	0	0	0	2	1	1	0	0	0	1	1	0	0	1	0	0	0	2
04	3	0	0	0	1	2	0	1	3	3	0	2	0	1	0	3	0	3
05	0	3	3	2	1	0	0	0	1	1	3	1	0	0	4	2	3	5
06	0	0	1	1	0	0	0	4	0	0	1	0	0	3	0	1	0	4
07	0	0	0	0	0	0	0	0	0	1	0	2	1	0	0	0	1	2
08	0	0	0	1	1	1	0	0	0	0	0	0	0	1	1	0	0	2
09	2	0	1	3	1	0	0	2	1	0	0	0	0	2	0	1	1	6
10	2	1	3	1	1	0	2	1	2	0	2	2	1	1	1	1	1	3
11	0	3	0	1	3	1	0	0	1	2	0	0	1	0	3	0	0	4
12	1	0	0	1	0	0	3	0	1	3	0	0	1	0	0	1	1	3
13	0	2	0	1	0	2	2	1	0	0	2	1	0	0	2	1	1	2
14	0	0	0	1	0	1	1	1	1	0	1	0	1	0	1	0	0	4
15	0	6	0	0	6	0	0	1	0	0	4	0	0	0	0	0	0	7
16	0	0	0	3	3	2	0	0	1	1	0	0	0	0	0	0	2	5
17	0	1	0	0	1	0	1	0	0	0	1	1	0	0	1	2	0	2

num	label	div	full name of department
01	arts_edu	IV	The Arts in Education
02	clin_psy	II	Clinical Psychology
03	comput_e	IV	Communication, Computing, and Technology in Education
04	curric_t	III	Curriculum and Teaching
05	devel_ed	II	Developmental and Educational Psychology
06	ed_admin	III	Educational Administration
07	hlth_nut	V	Health and Nutrition Education
08	adult_ed	III	Higher and Adult Education
09	lang_lit	IV	Languages, Literature, and Social Studies in Education
10	math_sci	IV	Mathematics and Science Education
11	measrmnt	II	Measurement, Evaluation, and Applied Statistics
12	move_sci	IV	Movement Sciences and Education
13	nurse_ed	V	Nursing Education
14	phil_soc	I	Philosophy and the Social Sciences
15	soc_cnsl	II	Social, Organizational, and Counseling Psychology
16	specl_ed	III	Special Education
17	speech_p	II	Speech and Language Pathology and Audiology

Example 2: Davis, Gardner & Gardner (1941) -- Southern women's club association data: Attendance of 18 women at 14 social events

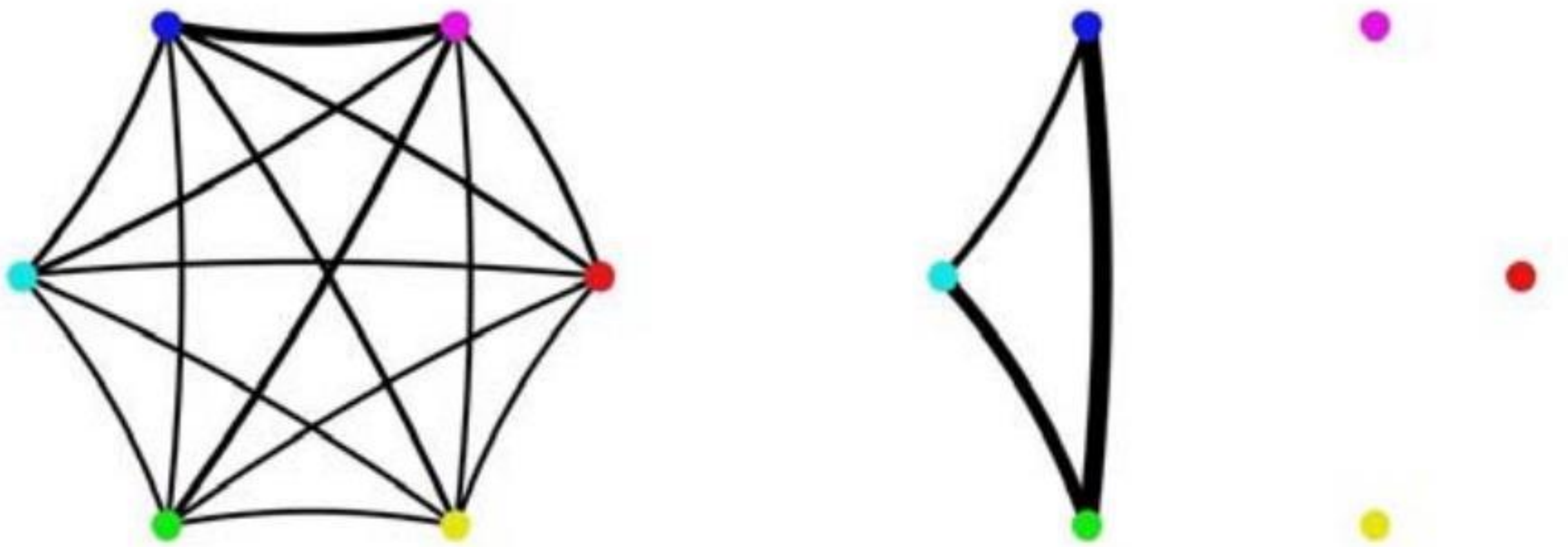
	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13	E14
EVELYN	1	1	1	1	1	1	0	1	1	0	0	0	0	0
LAURA	1	1	1	0	1	1	1	1	0	0	0	0	0	0
THERESA	0	1	1	1	1	1	1	1	1	0	0	0	0	0
BRENDA	1	0	1	1	1	1	1	1	0	0	0	0	0	0
CHARLOTTE	0	0	1	1	1	0	1	0	0	0	0	0	0	0
FRANCES	0	0	1	0	1	1	0	1	0	0	0	0	0	0
ELEANOR	0	0	0	0	1	1	1	1	0	0	0	0	0	0
PEARL	0	0	0	0	0	1	0	1	1	0	0	0	0	0
RUTH	0	0	0	0	1	0	1	1	1	0	0	0	0	0
VERNE	0	0	0	0	0	0	1	1	1	0	0	1	0	0
MYRNA	0	0	0	0	0	0	0	1	1	1	0	1	0	0
KATHERINE	0	0	0	0	0	0	0	1	1	1	0	1	1	1
SYLVIA	0	0	0	0	0	0	1	1	1	1	0	1	1	1
NORA	0	0	0	0	0	1	1	0	1	1	1	1	1	1
HELEN	0	0	0	0	0	0	1	1	0	1	1	1	1	1
DOROTHY	0	0	0	0	0	0	0	1	1	1	0	1	0	0
OLIVIA	0	0	0	0	0	0	0	0	1	0	1	0	0	0
FLORA	0	0	0	0	0	0	0	0	1	0	1	0	0	0

Concepts on social media (Twitter): Meme diffusion for different topics (weng et al., 2012)



Team activity in two student work groups

(Kay, Maisonneuve, Yacef & Reimann, 2006)



Bipartite graphs (connections between two sets of entities = two-mode association data)

Davis Gardner & Gardner (1941) -- Southern women's social event association data

Attendance of 18 women at 14 social events

	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13	E14
EVELYN	1	1	1	1	1	1	0	1	1	0	0	0	0	0
LAURA	1	1	1	0	1	1	1	1	0	0	0	0	0	0
THERESA	0	1	1	1	1	1	1	1	1	0	0	0	0	0
BRENDA	1	0	1	1	1	1	1	1	0	0	0	0	0	0
CHARLOTTE	0	0	1	1	1	0	1	0	0	0	0	0	0	0
FRANCES	0	0	1	0	1	1	0	1	0	0	0	0	0	0
ELEANOR	0	0	0	0	1	1	1	1	0	0	0	0	0	0
PEARL	0	0	0	0	0	1	0	1	1	0	0	0	0	0
RUTH	0	0	0	0	1	0	1	1	1	0	0	0	0	0
VERNE	0	0	0	0	0	0	1	1	1	0	0	1	0	0
MYRNA	0	0	0	0	0	0	0	1	1	1	0	1	0	0
KATHERINE	0	0	0	0	0	0	0	1	1	1	0	1	1	1
SYLVIA	0	0	0	0	0	0	1	1	1	1	0	1	1	1
NORA	0	0	0	0	0	1	1	0	1	1	1	1	1	1
HELEN	0	0	0	0	0	0	1	1	0	1	1	1	1	1
DOROTHY	0	0	0	0	0	0	0	1	1	1	0	1	0	0
OLIVIA	0	0	0	0	0	0	0	0	1	0	1	0	0	0
FLORA	0	0	0	0	0	0	0	0	1	0	1	0	0	0

Software for fitting and displaying graph and network models

R tools:

igraph package (see provided documentation)

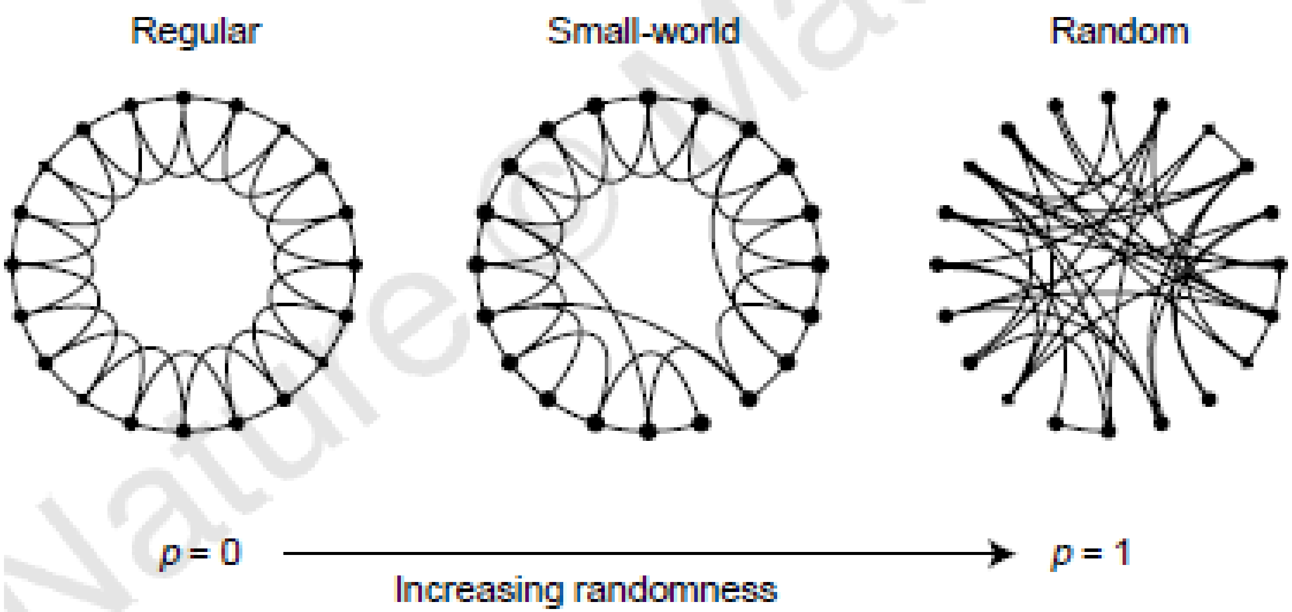
tnet package: social network statistics

Mathematica also has many good routines for computing with graphs and drawing networks

Watts & Strogatz (1998) Collective dynamics of small-world networks

Some networks are neither completely <u>regular</u> nor completely <u>random</u>. We quantify the structural properties of these graphs by their characteristic path length $L(p)$ and clustering coefficient $C(p)$, as defined in Fig. 2 legend. Here $L(p)$ measures the typical separation between two vertices in the graph (a global property), whereas $C(p)$ measures the cliquishness of a typical neighborhood (a local property).

These graphs, termed <u>small-world networks</u>, have many vertices with sparse connections, but not so sparse that the graph is in danger of becoming disconnected.

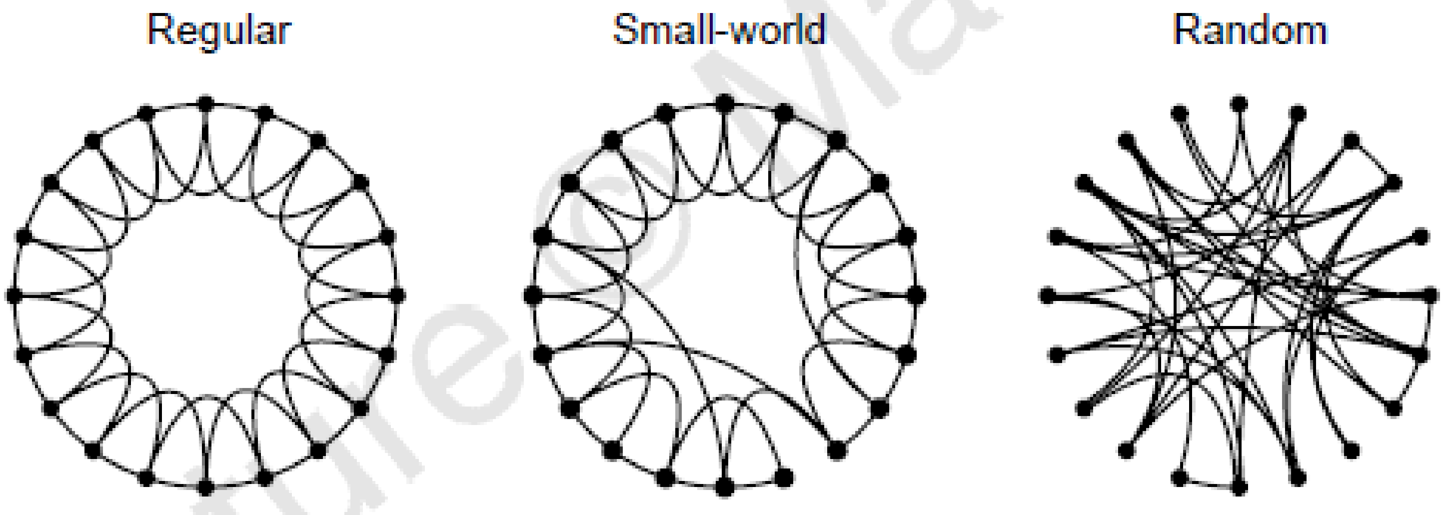


Two network statistics useful to identify small-world networks, as defined by Watts & Strogatz (1998):

- Characteristic path length $L(N)$: the number of edges in the shortest path between two vertices, averaged over all pairs of vertices.
- Clustering coefficient $C(N)$: The clustering coefficient $C(p)$ is defined as follows. Suppose that a vertex v has k_v neighbours; then at most $k_v(k_v-1)/2$ edges can exist between them. Let C_v denote the fraction of these allowable edges that actually exist. Define C as the average of C_v over all v.

For friendship networks, these statistics have intuitive meanings: L is the average number of friendships in the shortest chain connecting two people; Cv reflects the extent to which friends of v are also friends of each other; and thus C measures the cliquishness of a typical friendship circle.

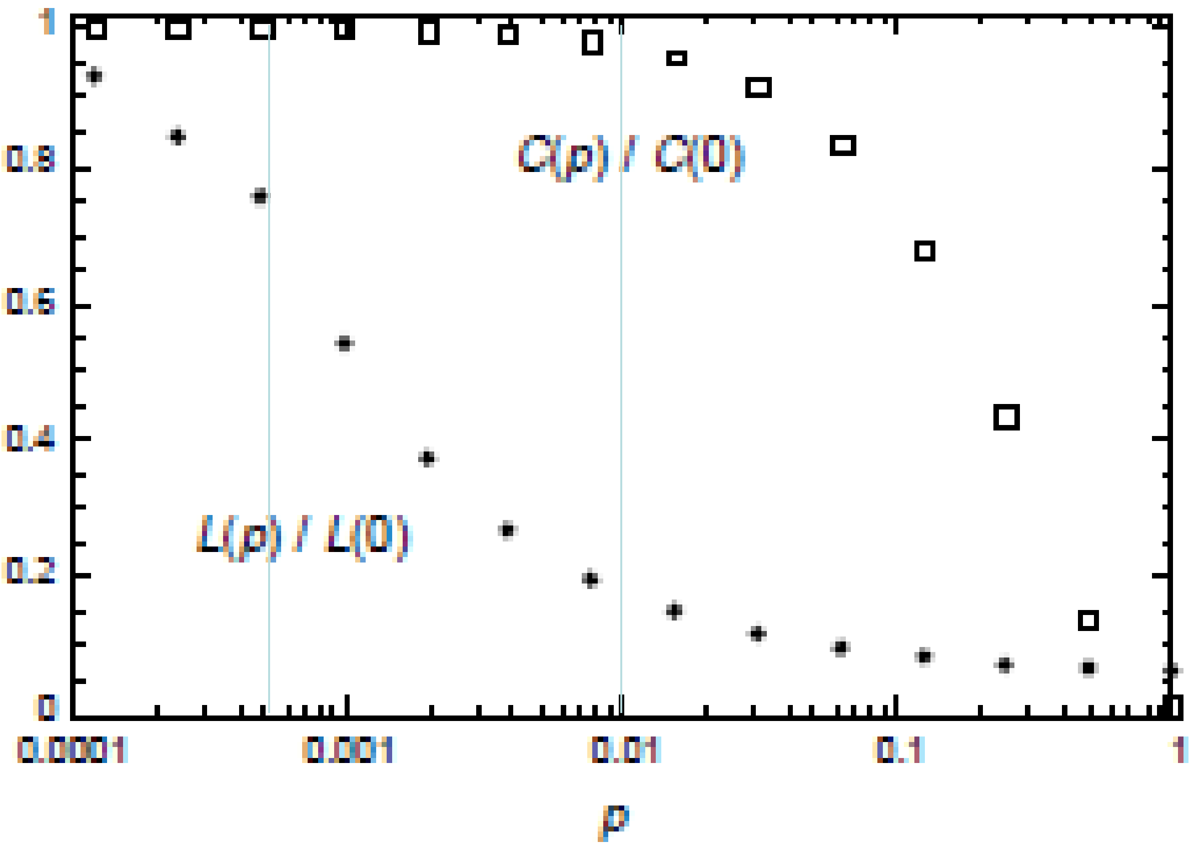
Figure 1 (Watts & Strogatz, 1998)



METHOD: Random rewiring procedure for interpolating between a regular ring lattice and a random network, without altering the number of vertices or edges in the graph.

We start with a ring of *n* vertices, each connected to its *k* nearest neighbors by undirected edges. (For clarity, $n = 20$ and $k = 4$ in the schematic examples shown here, but much larger *n* and *k* are used in the rest of this Letter.) We choose a vertex and the edge that connects it to its nearest neighbor in a clockwise sense. With probability *p*, we reconnect this edge to a vertex chosen uniformly at random over the entire ring, with duplicate edges forbidden; otherwise we leave the edge in place. We repeat this process by moving clockwise around the ring, considering each vertex in turn until one lap is completed. Next, we consider the edges that connect vertices to their second-nearest neighbors clockwise. As before, we randomly rewire each of these edges with probability *p*, and continue this process, circulating around the ring and proceeding outward to more distant neighbors after each lap, until each edge in the original lattice has been considered once.

FIGURE 2 (Watts & Strogatz 1998). ..The data shown in the figure are averages over 20 random realizations of the rewiring process described in Fig.1, and have been normalized by the values *L*(0), *C*(0) for a regular lattice. All the graphs have *n*=1000 vertices and an average degree of *k*=10 edges per vertex. We note that a logarithmic horizontal scale has been used to resolve **the rapid drop in *L*(*p*), corresponding to the onset of the small-world phenomenon.** During this drop, *C*(*p*) remains almost constant at its value for the regular lattice, indicating that the transition to a small world is almost undetectable at the local level.



Small-world networks: empirical examples

(Watts & Strogatz 1998)

	L_{actual}	L_{random}	C_{actual}	C_{random}
Film actors	3.65	2.99	0.79	0.00027
Power grid	18.7	12.4	0.08	0.005
C. elegans	2.65	2.25	0.28	0.05

Table 1: characteristic path length *L* and clustering coefficient *C* for three real networks, compared to random graphs with the same number of vertices (*n*) and average number of edges per vertex (*k*). (Actors: n=225,226; *k*=61. Power grid: *n*=4941; *k*=2.67. *C. Elegans*: *n*=28; *k*=14) .. We treat all edges as undirected and unweighted, and all vertices as identical.. All three networks show the small-world phenomenon: $L_{actual} \geq L_{random}$ but $C_{actual} \gg C_{random}$

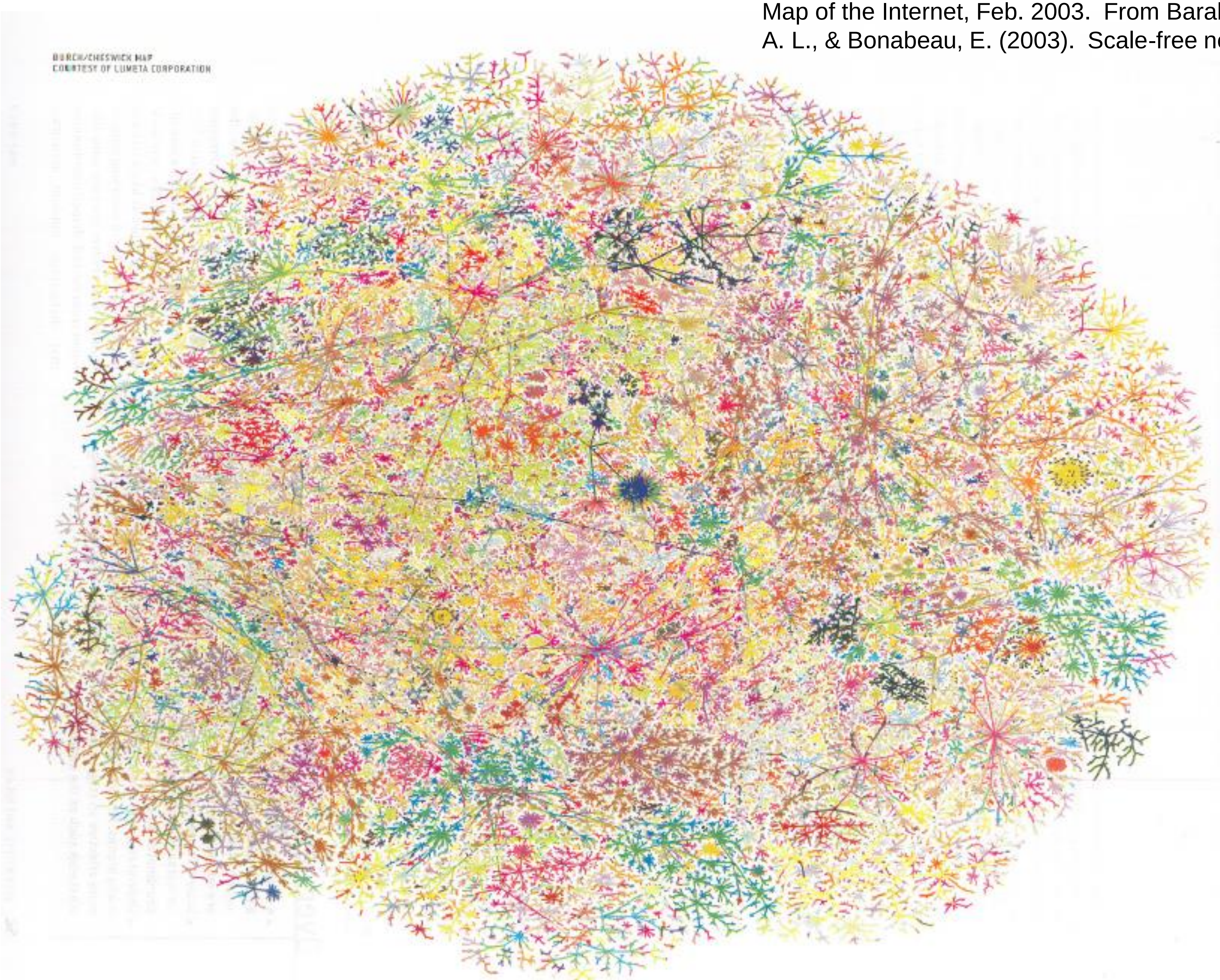
(The graphs are defined as follows. Two actors are joined by an edge if they have acted in a film together. We restrict attention to the giant connected component of this graph, which includes 90% of all actors listed in the Internet Movie Database (available at http://us.imdb.com), as of April 1997. For the power grid, vertices represent generators, transformers and substations, and edges represent high-voltage transmission lines between them. For *C. elegans*, an edge joins two neurons if they are connected by either a synapse or a gap junction.)

Barabasi & Boneau (2003). Scale-free networks. *Scientific American.*

Overview/Scale-Free Networks

- A variety of complex systems share an important property: some nodes have a tremendous number of connections to other nodes, whereas most nodes have just a handful. The popular nodes, called hubs, can have hundreds, thousands or even millions of links. In this sense, the network appears to have no scale.
- Scale-free networks have certain important characteristics. They are, for instance, robust against accidental failures but vulnerable to coordinated attacks.
- Understanding of such characteristics could lead to new applications in many arenas. For example, computer scientists might be able to devise more effective strategies for preventing computer viruses from crippling a network such as the Internet.

Map of the Internet, Feb. 2003. From Barabási, B. A. L., & Bonabeau, E. (2003). Scale-free networks.



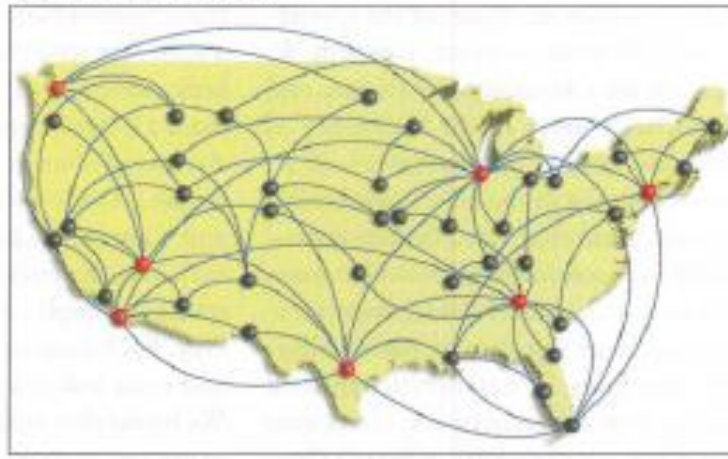
RANDOM VERSUS SCALE-FREE NETWORKS

RANDOM NETWORKS, which resemble the U.S. highway system [*simplified in left map*], consist of nodes with randomly placed connections. In such systems, a plot of the distribution of node linkages will follow a bell-shaped curve [*left graph*], with most nodes having approximately the same number of links.

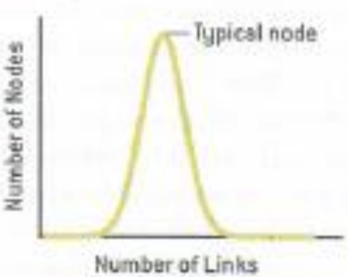
In contrast, scale-free networks, which resemble the U.S. airline system [*simplified in right map*], contain hubs [*red*]—nodes with a very high number of links. In such networks, the distribution of node linkages follows a power law [*center graph*] in that most nodes have just a few connections and some have a tremendous number of links. In that sense, the system has no "scale." The defining characteristic of such networks is that the distribution of links, if plotted on a double-logarithmic scale [*right graph*], results in a straight line.

Random Network

Scale-Free Network



Bell Curve Distribution of Node Linkages



Power Law Distribution of Node Linkages