Distributed Algorithms
Complex Networks

Alberto Montresor

Università di Trento

2019/10/09

Acknowledgments: Mark Jelasity, Sebastian Ahnert, Lada Adamic, Albert Diaz Guilera

Table of contents

Introduction

What these structures have in common?

- World Wide Web
- Internet
- Movie actor collaborations
- Science collaborations
- Citations of papers
- Sexual relationships
- Food webs
- Facebook & Linkedin
- Co-occurrence of words
- Your brain
- The power network of USA
- Protein folding

Introduction

What these structures have in common?

- World Wide Web
- Internet
- Movie actor collaborations
- Science collaborations
- Citations of papers
- Sexual relationships
- Food webs
- Facebook & Linkedin
- Co-occurrence of words
- Your brain
- The power network of USA
- Protein folding

- They can be described as graphs

Introduction

What these structures have in common?

- World Wide Web
- Internet
- Movie actor collaborations
- Science collaborations
- Citations of papers
- Sexual relationships
- Food webs
- Facebook & Linkedin
- Co-occurrence of words
- Your brain
- The power network of USA
- Protein folding

- They can be described as graphs
- They all show similar features!

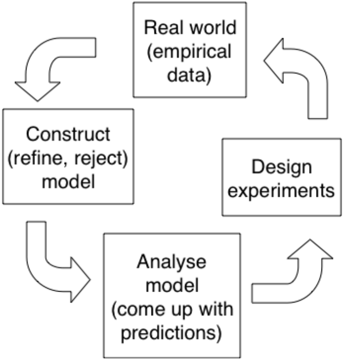
Brief historical overview:

- 1736: Graph theory (Euler)
- 1937: Journal Sociometry founded
- 1959: Random graphs (Erdős-Rényi)
- 1967: Small-world (Milgram)
- 1990s: "Complex networks"

Complex network research

- Rapidly increasing interest over the last decade, since much more network data available now
- Multidisciplinary research
  - Physics
  - Biology
  - Sociology
  - Mathematics
  - Epidemiology
  - ...
- Strong implications for Computer Science
  - Robustness of networks
  - Efficiency: function of networks depends on their structure
  - Design and engineering

Complex network research



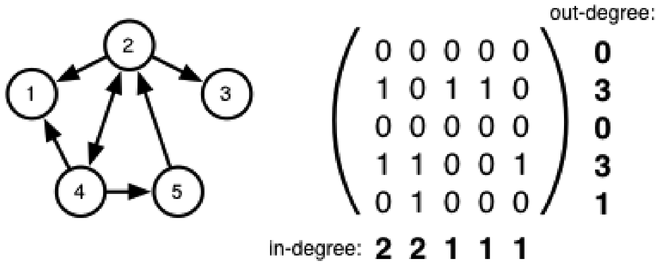
- Complex networks is a branch of physics
- Empirical science: loop of modeling and observation
- Models are models
  - Explain some aspects
  - Don't explain other aspects
- Has a lot to do with condensed matter physics

Table of contents

Degree statistics (1)

Degree

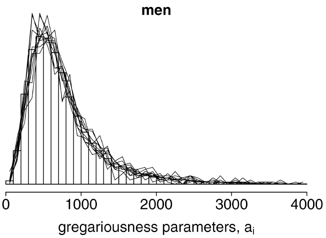
- The degree k_i of node i is the number of edges the node has to other nodes
- In this lecture, we will only consider undirected graphs
- Called local centrality in social network analysis
- Measures how important is a node w.r.t. its neighbors



Degree statistics (2)

Degree distribution

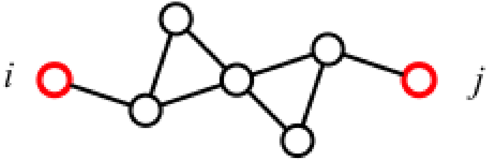
- The average degree of a graph $G = (V, E)$ is $\langle k \rangle = \frac{|E|}{|V|}$
- $P(k)$ is the probability that a random node has degree k
- The probability distribution gives an idea of the spread in the number of links the nodes have

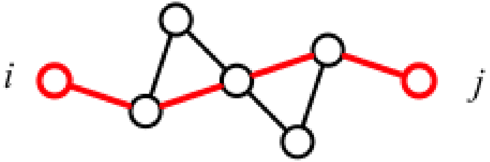


Distance statistics (1)

Path length

The path length $d(i,j)$, or goedesic distance, between two nodes i and j is the length of the shortest path connecting them.





Distance statistics (2)

Average path length

The average path length of a connected graph $G = (V, E)$ is defined as:

$$\ell(G) = \frac{1}{|V|(|V|-1)} \sum_{i,j \in V, i \neq j} d(i,j)$$

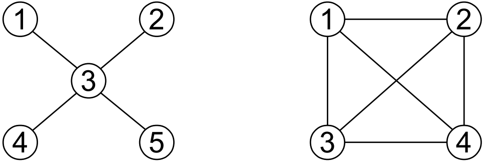


Figure: (a) $\ell(G) = (4 \cdot ((2+2+2+1)/8)+1)/5 = 1.6$. (b) $\ell(G) = 1$

Distance statistics (3)

Diameter

The diameter is the maximal shortest path between any two vertices:

$$diam(G) = \max\{d(i,j) : i, j \in V\}$$

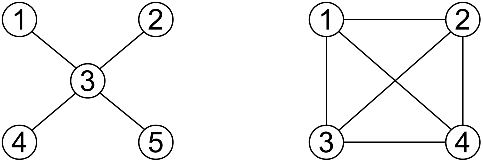


Figure: (a) $diam(G) = 2$. (b) $diam(G) = 1$

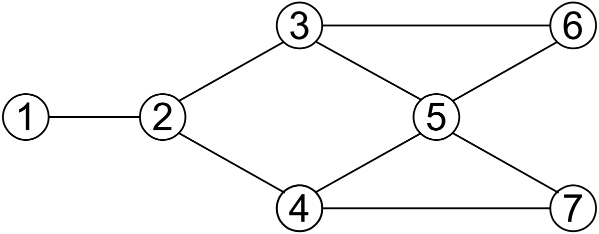
Clustering coefficient (1)

Induced graphs

The subgraph $G_i = (V_i, E_i)$ induced by node i over a graph $G = (V, E)$ is given by the neighbors of i and the edges linking them:

$$V_i = \{j : j \in V \wedge (i,j) \in E\}$$

$$E_i = \{(i,j) : i,j \in V_i \wedge (i,j) \in E\}$$



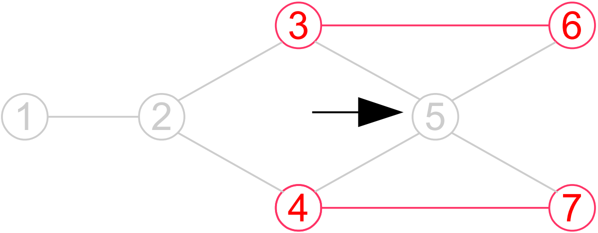
Clustering coefficient (1)

Induced graphs

The subgraph $G_i = (V_i, E_i)$ induced by node i over a graph $G = (V, E)$ is given by the neighbors of i and the edges linking them:

$$V_i = \{j : j \in V \wedge (i, j) \in E\}$$

$$E_i = \{(i, j) : i, j \in V_i \wedge (i, j) \in E\}$$

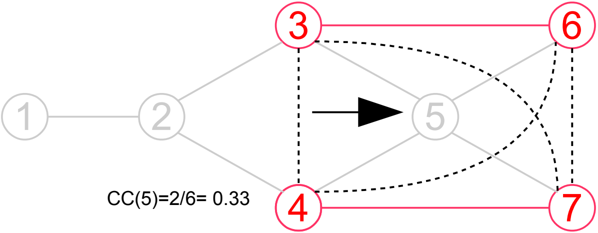


Clustering coefficient (2)

Local clustering coefficient

The local clustering coefficient of node i in graph G is the ratio between the size of $|E_i|$ and the number of all potential edges that link two nodes in V_i:

$$CC(G, i) = \frac{|E_i|}{|V_i|(|V_i| - 1)/2} = \frac{2|E_i|}{|V_i|(|V_i| - 1)}$$

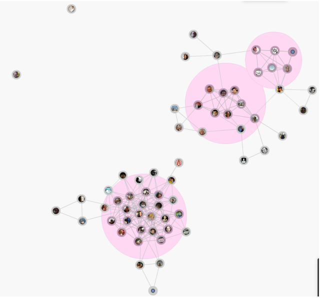


Clustering coefficient (3)

Clustering coefficient

The clustering coefficient of graph G is the average over the local clustering coefficient of all nodes in the graph

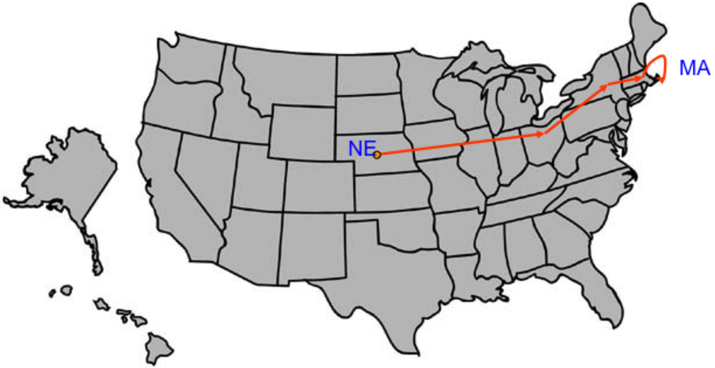
$$CC(G) = \frac{1}{N} \sum_{i \in V} CC(G, i)$$



- Facebook clustering coefficient 0.16

The Milgram Small-World Experiment

Milgram's experiment (1967):
Given a target individual and a particular property, pass the message to a person you correspond with who is "closest" to the target.



The Milgram Small-World Experiment

Some facts

- Target person worked in Boston as a stockbroker.
- 296 senders from Boston and Omaha.
- 20% of senders reached target.
- Average path length = 6.5.

"Six degrees of separation"

- It's a small world after all!
- Kevin-Bacon game
- Erdös number

Six Degrees and Popular Culture

"Everything is Different"

A fascinating game grew out of this discussion. One of us suggested performing the following experiment to prove that the population of the Earth is closer together now than they have ever been before. We should select any person from the 1.5 billion inhabitants of the Earth—anyone, anywhere at all. He bet us that, using no more than five individuals, one of whom is a personal acquaintance, he could contact the selected individual using nothing except the network of personal acquaintances.

Frigyes Karinthy, 1929

Statistics over real networks

Network	Size	$\langle k \rangle$	ℓ	ℓ_{rand}	C	C_{rand}	Reference	Nr.
WWW, site level, undir.	153 127	35.21	3.1	3.35	0.1078	0.00023	Adamic, 1999	1
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	2
Movie actors	225 226	61	3.65	2.99	0.79	0.00027	Watts and Strogatz, 1998	3
LANL co-authorship	52 909	9.7	5.9	4.79	0.43	1.8×10^{-4}	Newman, 2001a, 2001b, 2001c	4
MEDLINE co-authorship	1 520 251	18.1	4.6	4.91	0.066	1.1×10^{-5}	Newman, 2001a, 2001b, 2001c	5
SPIRES co-authorship	56 627	173	4.0	2.12	0.726	0.003	Newman, 2001a, 2001b, 2001c	6
NCSTRL co-authorship	11 994	3.59	9.7	7.34	0.496	3×10^{-4}	Newman, 2001a, 2001b, 2001c	7
Math. co-authorship	70 975	3.9	9.5	8.2	0.59	5.4×10^{-5}	Barabási *et al.*, 2001	8
Neurosci. co-authorship	209 293	11.5	6	5.01	0.76	5.5×10^{-5}	Barabási *et al.*, 2001	9
E. coli, substrate graph	282	7.35	2.9	3.04	0.32	0.026	Wagner and Fell, 2000	10
E. coli, reaction graph	315	28.3	2.62	1.98	0.59	0.09	Wagner and Fell, 2000	11
Ythan estuary food web	134	8.7	2.43	2.26	0.22	0.06	Montoya and Solé, 2000	12
Silwood Park food web	154	4.75	3.40	3.23	0.15	0.03	Montoya and Solé, 2000	13
Words, co-occurrence	460.902	70.13	2.67	3.03	0.437	0.0001	Ferrer i Cancho and Solé, 2001	14
Words, synonyms	22 311	13.48	4.5	3.84	0.7	0.0006	Yook *et al.*, 2001b	15
Power grid	4941	2.67	18.7	12.4	0.08	0.005	Watts and Strogatz, 1998	16
C. Elegans	282	14	2.65	2.25	0.28	0.05	Watts and Strogatz, 1998	17

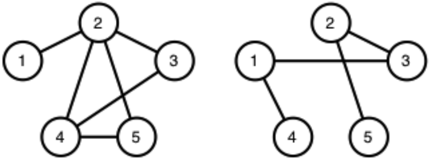
Table of contents

Erdös-Rényi Random (E-R) Networks

Random Graph

Given a network of N nodes, connect each pair (p_i, p_j) of nodes with probability p

Two different realizations for $N = 5$ and $p = 0.5$



Some properties of E-R networks

Given an E-R graph with N nodes and probability p

- Expected number of edges: $|E| = pN(N-1)/2$
- Expected average degree: $\langle k \rangle = 2|E|/N = p(N-1) \approx pN$
- Degree distribution: $P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}$ (binomial)
- In the limit of large N, $diam(G) = \frac{\log N}{\log \langle k \rangle}$
- Average path length is bounded by diameter
- Clustering coefficient: $CC(G) = p$

Random networks vs real ones

Question

Are random networks a good model for real networks?

Statistics over real networks

Network	Size	$\langle k \rangle$	ℓ	ℓ_{rand}	C	C_{rand}	Reference	Nr.
WWW, site level, undir.	153 127	35.21	3.1	3.35	0.1078	0.00023	Adamic, 1999	1
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	2
Movie actors	225 226	61	3.65	2.99	0.79	0.00027	Watts and Strogatz, 1998	3
LANL co-authorship	52 909	9.7	5.9	4.79	0.43	1.8×10^{-4}	Newman, 2001a, 2001b, 2001c	4
MEDLINE co-authorship	1 520 251	18.1	4.6	4.91	0.066	1.1×10^{-5}	Newman, 2001a, 2001b, 2001c	5
SPIRES co-authorship	56 627	173	4.0	2.12	0.726	0.003	Newman, 2001a, 2001b, 2001c	6
NCSTRL co-authorship	11 994	3.59	9.7	7.34	0.496	3×10^{-4}	Newman, 2001a, 2001b, 2001c	7
Math. co-authorship	70 975	3.9	9.5	8.2	0.59	5.4×10^{-5}	Barabási *et al.*, 2001	8
Neurosci. co-authorship	209 293	11.5	6	5.01	0.76	5.5×10^{-5}	Barabási *et al.*, 2001	9
E. coli, substrate graph	282	7.35	2.9	3.04	0.32	0.026	Wagner and Fell, 2000	10
E. coli, reaction graph	315	28.3	2.62	1.98	0.59	0.09	Wagner and Fell, 2000	11
Ythan estuary food web	134	8.7	2.43	2.26	0.22	0.06	Montoya and Solé, 2000	12
Silwood Park food web	154	4.75	3.40	3.23	0.15	0.03	Montoya and Solé, 2000	13
Words, co-occurrence	460.902	70.13	2.67	3.03	0.437	0.0001	Ferrer i Cancho and Solé, 2001	14
Words, synonyms	22 311	13.48	4.5	3.84	0.7	0.0006	Yook *et al.*, 2001b	15
Power grid	4941	2.67	18.7	12.4	0.08	0.005	Watts and Strogatz, 1998	16
C. Elegans	282	14	2.65	2.25	0.28	0.05	Watts and Strogatz, 1998	17

Small-World Networks

Answer – No!

- Average path length of real networks is similar to the one of random networks...
- ...but their clustering coefficient is many orders of magnitude larger

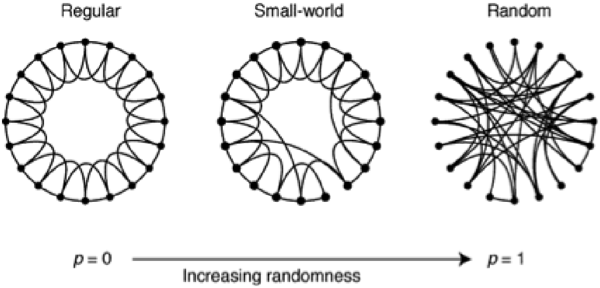
Problem – Modeling small-world networks

How we can build a network with small degree, small average path length and large clustering coefficient?

Small-World Networks

Watts-Strogatz Model

Watts and Strogatz (1998) observed that by taking a locally connected network and randomly rewiring a small number of edges, the average distance between two nodes falls drastically

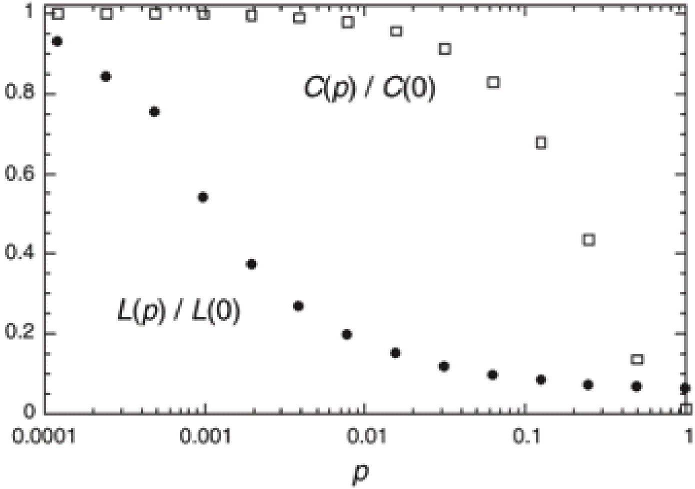


Formal definition

Watts-Strogatz Model

1. Construct a regular ring lattice, a graph with N nodes each connected to k neighbors, $k/2$ on each side
   - Label nodes $0, 1, \ldots, N-1$
   - Add an edge (i, j) if and only if $0 < |i - j| \bmod N \leq k/2$
2. For every node $i \in [0, N-1]$, take each each every edge (i, j) with $i < j$ and rewire it with probability p
   - Replace (i, j) with (i, k) such that $i \neq k$ and $(i, k) \notin E$

The small-world region



Looking back at degree distribution

Question

Are Watts-Strogatz networks a good model for real networks?

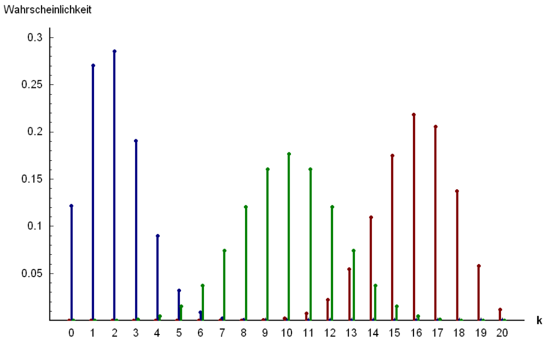


Figure: Binomial distribution for $n = 20$ $p = 0.1$ (blue), $p = 0.5$ (green) and $p = 0.8$ (red). `http://en.wikipedia.org/wiki/File:Binomial_Distribution.PNG`

Looking back at degree distribution

Question

Are Watts-Strogatz networks a good model for real networks?

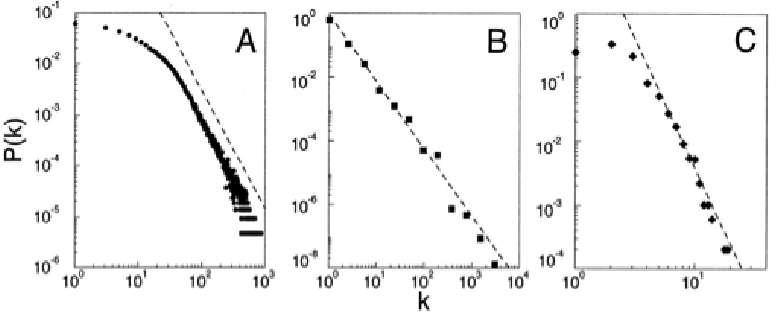


Figure: Real network degree distributions: from left to right, Actors, WWW, Power grid

Scale-Free Networks

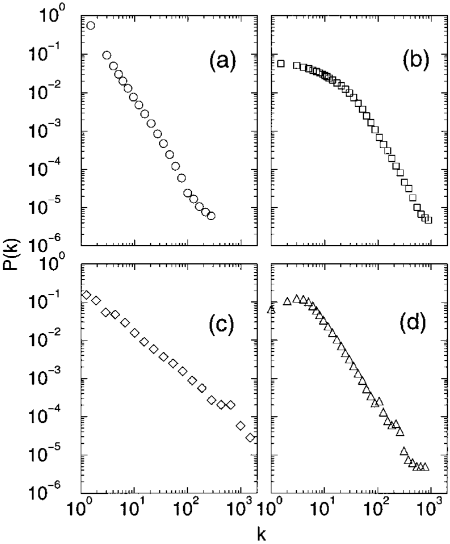
- Many nodes have few connections and a few nodes have many connections
- This observation holds on the local and global scale of the network
- In other words, there is no inherent scale

Power-law degree distribution

Formally this translates into a power-law degree distribution:

$$P(k) \propto k^{-\lambda}$$

Statistics over real networks



(a) Internet at the router level
(b) movie actor collaboration network
(c) co-authorship network of high-energy physicists
(d) co-authorship network of neuroscientists

Statistics over real networks

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

Barabási-Albert Model

Preferential attachment

- Start with a small clique of size $m_0 \geq 3$
- Repeat adding a new node with $m \leq m_0$ edges
- New node is connected with node i with a probability proportional to the number of links that i already has. Formally,

$$P(i) = \frac{k_i}{\sum_j k_j}$$

Mathematical properties

- $m_0 + t$ nodes, mt edges
- $P(k) \propto k^{-3}$
- $\ell(G) = \frac{\log N}{\log \log N}$
  (somewhat smaller than random)
- $CC(G) \approx N^{-0.75}$

Summary

- Models degree distribution
- But doesn't model clustering

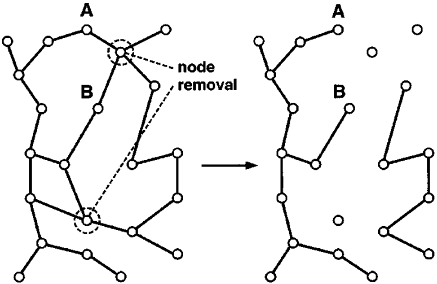
Table of contents

Assortativity

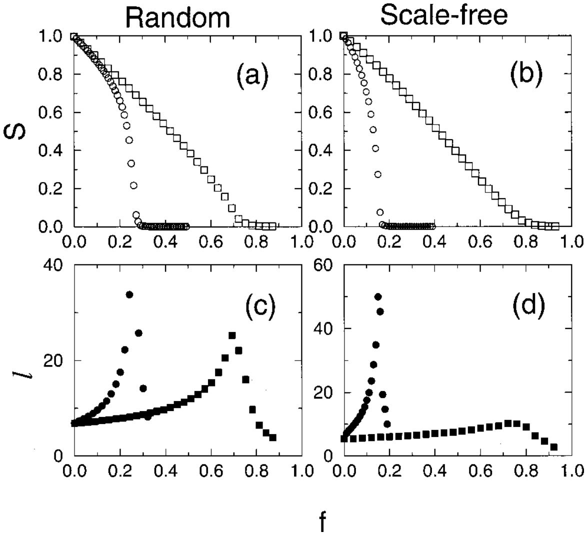
- Assortativity describes the correlation between the degree of a node and the degree of its neighbors.
- Networks in which highly connected nodes are linked to other nodes with a high degree are termed assortative. Such networks include social networks
- Networks in which highly connected nodes are only linked to nodes with a low degree are termed disassortative. Such networks include the World Wide Web and biological networks.

Robustness

- We need to understand how vulnerable existing systems are
- We need to design self-healing and self-protecting systems
- Node removal models:
  - Random: a random node is removed along with all the links
  - Preferential: the most connected (highest degree) nodes are removed



Robustness



(a,c) random network, $N = 10000$, $\langle k \rangle = 4$

(b,d) BA network, $N = 10000$, $\langle k \rangle = 4$

□ random removal

○ preferential removal

The story is not over...

- k-core decomposition
- Betweenness centrality
- Closeness centrality
- Eigenvector centrality
- Cohesive subgroups
- ...

Conclusions

As computer scientists, what we can learn from complex networks?

- Technological networks are resembling more and more to "real-life" networks
- It opens a new topic of research: empirical computer science
- Measurement studies on:
  - BitTorrent,
  - Internet,
  - WWW, …
- Mathematical modeling is a useful tool, but it is not enough

Reading Material

- R. Albert and A. L. Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1):47–97, Jan. 2002. http://www.disi.unitn.it/~montreso/ds/papers/ComplexNetworks.pdf

Further reading

- L. da F. Costa, F. A. Rodrigues, G. Travieso, and F. A. Villas Boas.
  Characterization of complex networks: A survey of measurements.
  Technical Report arXiv:cond-mat/0505185v5, arXiv, 2006.
  https://arxiv.org/abs/cond-mat/0505185v5