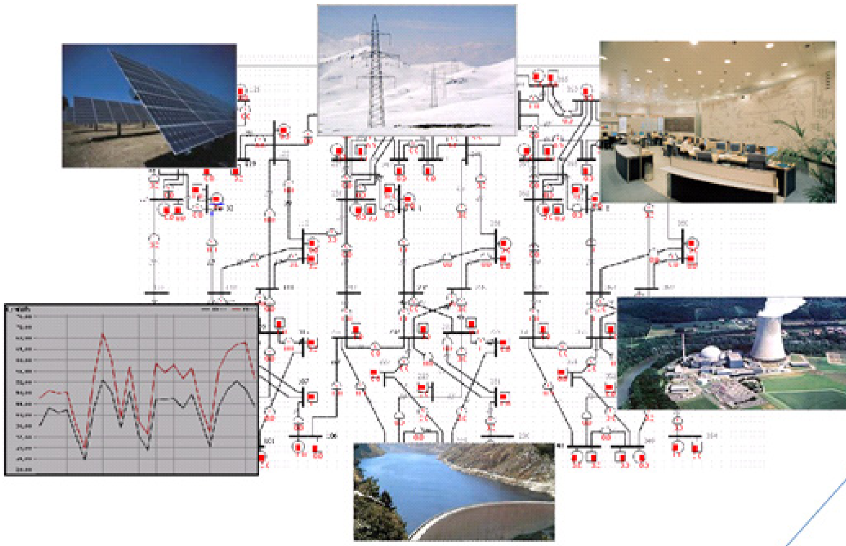
Advanced Methods for Complex Systems' Modeling and Simulation I:

Network theory for the vulnerability analysis of infrastructure systems

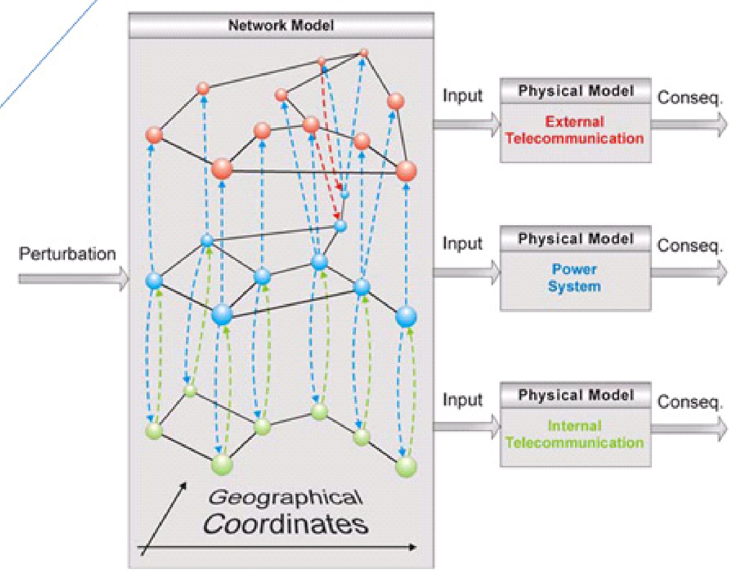
Lecture contents:

1. Introduction and problem description
2. Basic network characteristics
3. Static network vulnerability analysis
   (random failure and attack tolerance)
4. Cascading failures within infrastructure systems

Real Networks



Network Models



Introduction and Problem Description (I)

- Infrastructure systems provide essential goods and services to the industrialised society including transport, water, communication and energy.

- A disruption or malfunction often has a significant economic impact and potentially propagates to other systems due to mutual interdependencies.

- Wide-area breakdowns of such large-scale engineering networks are often caused by technical equipment failures and their coincidence in time which eventually result in a series of fast cascading component outages.

- Illustrative examples are a number of large electric power blackouts and near-misses as has been increasingly experienced in the last few years

Introduction and Problem Description (II)

How can we quantify the reliability of infrastructures and **assess the risk** of such large-area breakdowns?

Basic problem: Infrastructures are **highly complex** and interdependent systems, consisting of an enormous number of technical and non-technical, interacting components; classic reliability analysis methods become limited due to the state space explosion.

Example: Consider a system of N=20 components with up state and down state. A "state enumeration approach", such as a complete Markovian chain would have to consider $2^N = 2^{20} \sim 10^6$ *system states!*

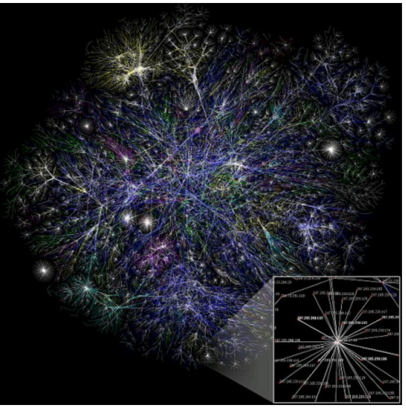
Approach 1: *Simulate the systems realistically by means of extensive modeling methods, including physical laws and operational dynamics.*

Introduction and Problem Description (III)

Approach 2: *Use highly simplified models in order to understand the* ***basic mechanisms*** *leading to infrastructure breakdowns. In this respect* ***network theory*** *allows for gaining valuable qualitative knowledge about the basic functioning of infrastructure systems, being networks in nature.*

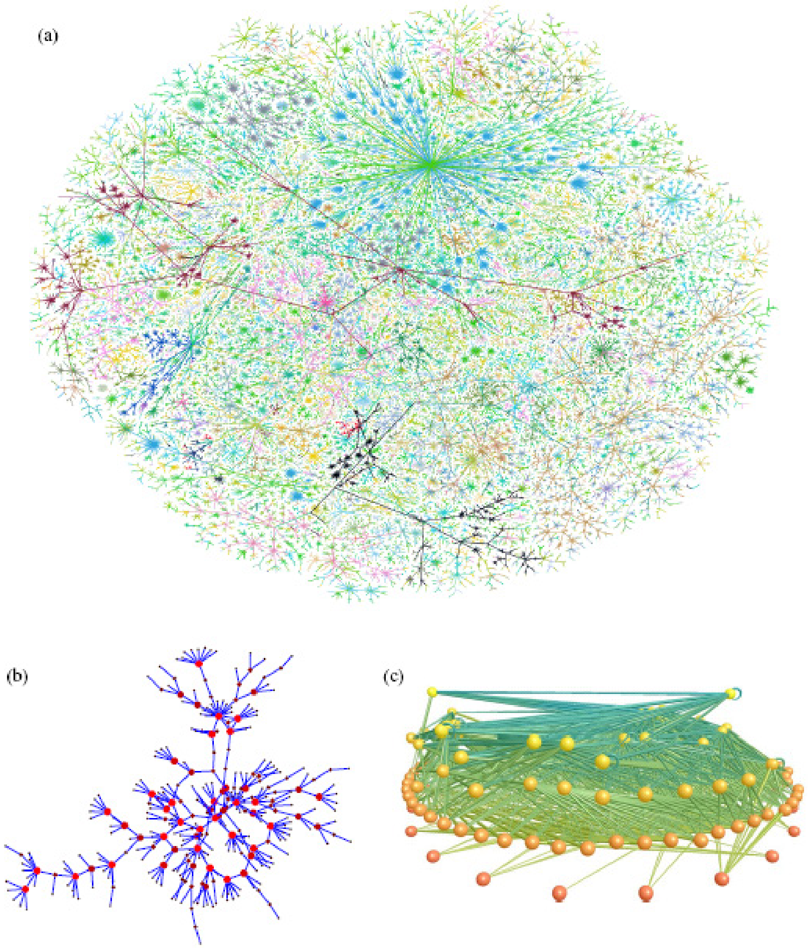
However, due to its highly simplifying approach, network theory cannot replace more detailled reliability analysis methods.

It rather serves as a first „screening analysis“, whereas the findings, e.g. robustness of topology, may serve as an *input* for detailled reliability studies (and vice versa).



What can be represented as networks?

A few examples:



(a) Internet at the level of “autonomous systems”

(b) A social network (e.g. sexual contacts)

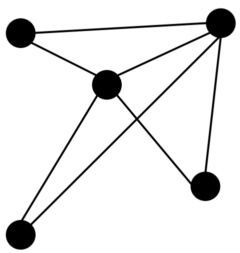
(c) A food web of predator-prey interactions between species.

Source: Newman, SIAM Rev. 45, 167 (2003).

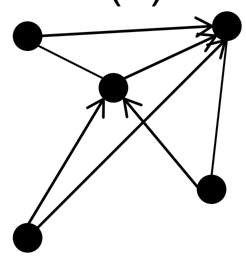
Network Characteristics: Some Basic Notations

- A network (or **graph**) is a set of *N* **nodes** (or **vertices** or **sites**) connected by *L* **links** (or **edges** or **bonds**)
- **G**(*N,L)*: arbitrary graph of **order** *N* and **size** *L*
- Networks with undirected links are called **undirected networks** (a), those with directed links are called **directed networks** (b)

(a)

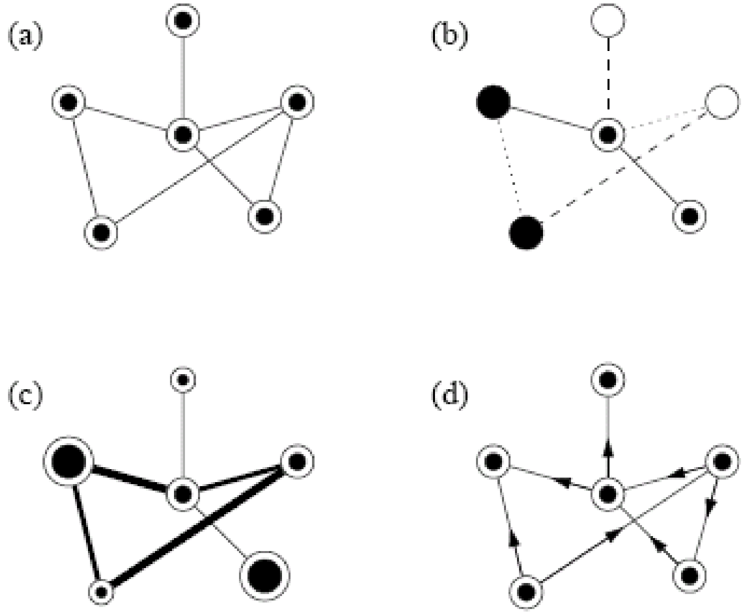


(b)



- The total number of connections of a node to ist nearest neighboring nodes is called its **degree** *k*

Network Characteristics: Vertex Types and Edge Weights to Represent more Diversity



(a) an undirected network with only a single type of vertex and a single type of edge; (b) a network with a number of discrete vertex and edge types; (c) a network with varying vertex and edge weights; (d) a directed network in which each edge has a direction.
Source: Newman, SIAM Rev. 45, 167 (2003).

Network Characteristics: Adjacency Matrix

- The adjacency matrix ***A*** provides a complete description of a network
- Consider a network with N nodes labelled by their index i (i=1,…,N). Then the adjacency matrix is a N x N matrix with elements a_{ij} :

if the network is undirected:

$a_{ij} = a_{ji}$, $a_{ij} = 1$ *if there exists a link between node i and j*

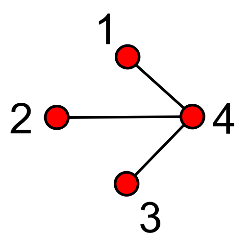
$a_{ij} = 0$ *otherwise*

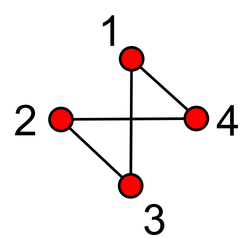
if the network is directed:

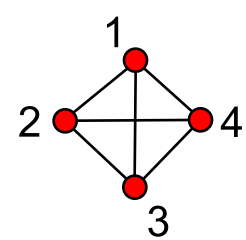
$a_{ij} \neq a_{ji}$, $a_{ij} = 1$ *if there exists a link leaving node i and going to node j*

$a_{ij} = 0$ *otherwise*

- Examples for undirected graphs:







$$\begin{array}{c c} & \begin{matrix} \mathbf{a_1} & \mathbf{a_2} & \mathbf{a_3} & \mathbf{a_4} \end{matrix} \\ \begin{matrix} \mathbf{a_1} \\ \mathbf{a_2} \\ \mathbf{a_3} \\ \mathbf{a_4} \end{matrix} & \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \end{array} \qquad \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix} \qquad \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

$$k_i = \sum_j a_{ij}$$

Network Characteristics: Degree distribution

The **degree distribution** *P(k)* gives the probability that any randomly chosen vertex has degree *k.*

Poisson

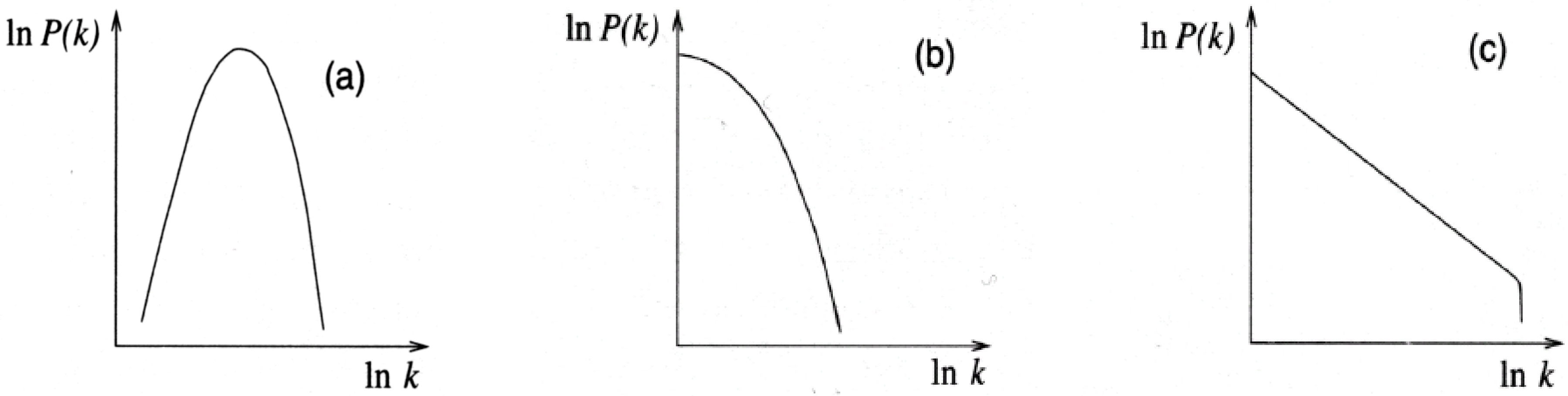
$$P(k) = \frac{e^{-\alpha}\alpha^{k}}{k!}, \text{ where } \alpha = \overline{k}$$

Exponential

$$P(k) \propto e^{-k/\alpha}, \text{ where } \alpha = \overline{k}$$

Power law

$$P(k) \propto k^{-\gamma}, \; k \neq 0$$



Source: Dorogovtsev, S. N. and Mendes (2003)

Examples of typical degree distributions

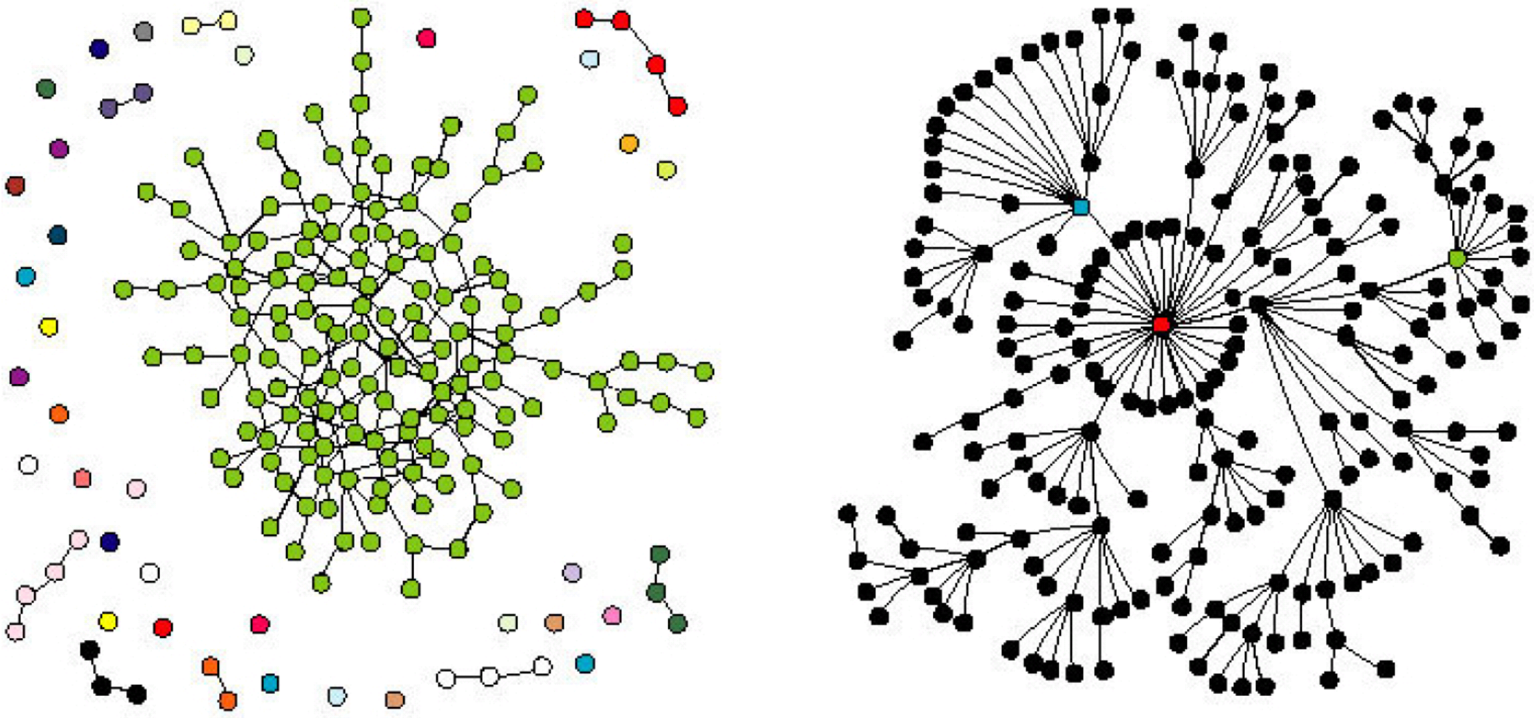


Illustration of network architectures. Left: random graph (Poisson), right: scale-free network (power law). Source: Strogatz, S. H., Nature 410, 268-276 (2001)

Degree distributions – WWW

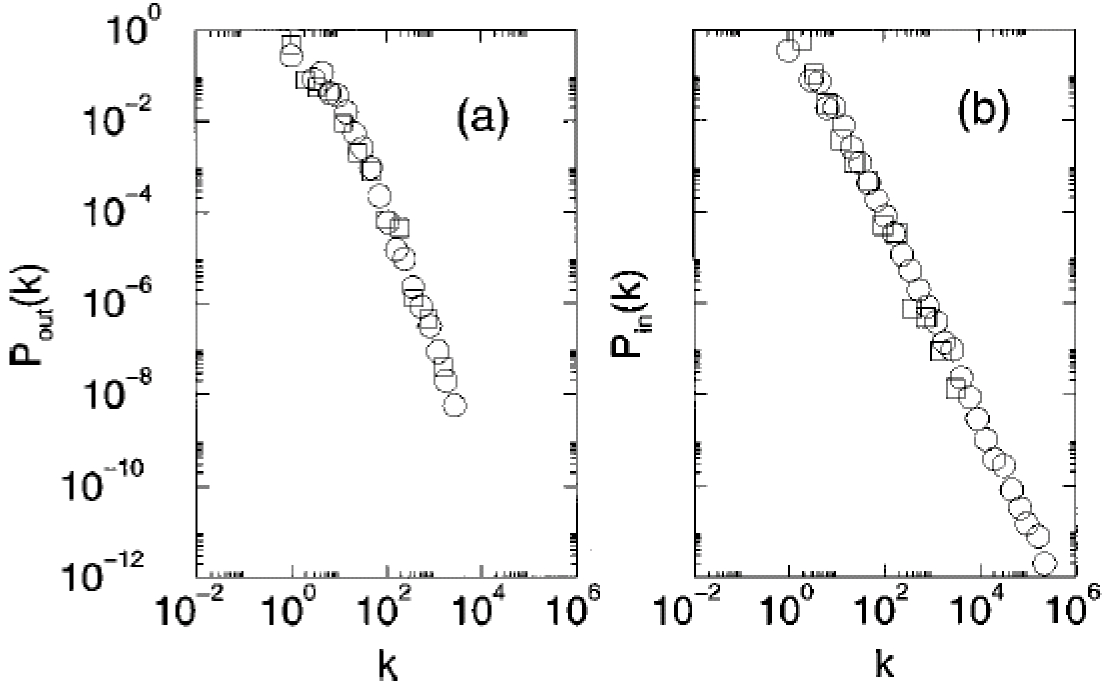
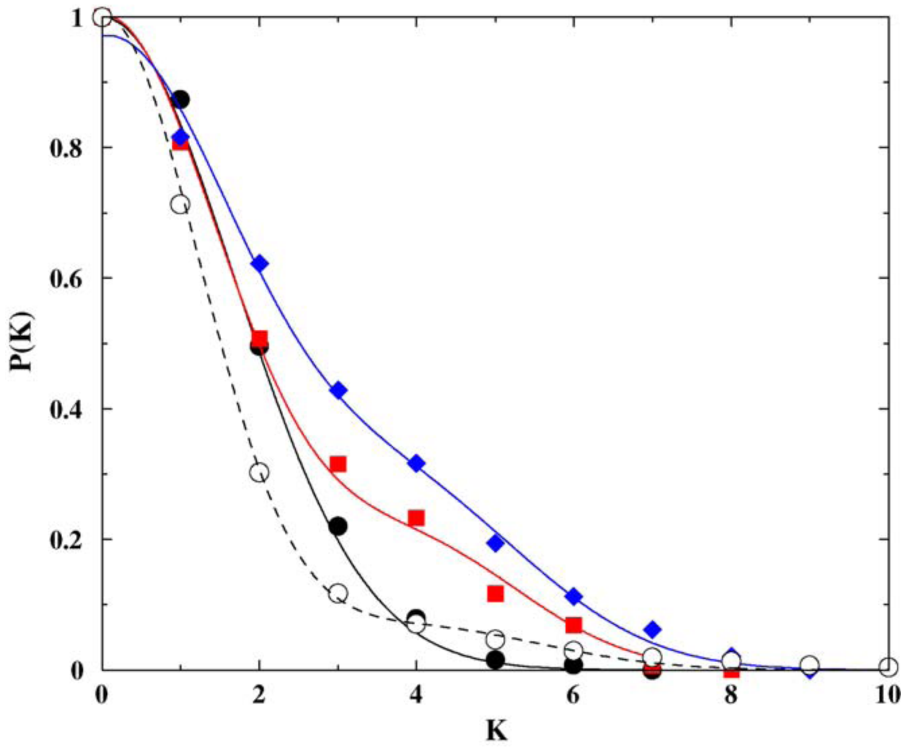


FIG. 2. Degree distribution of the World Wide Web from two different measurements: □, the 325 729-node sample of Albert *et al.* (1999); ○, the measurements of over 200 million pages by Broder *et al.* (2000); (a) degree distribution of the outgoing edges; (b) degree distribution of the incoming edges. The data have been binned logarithmically to reduce noise.

Source: Albert, R., Barabàsi, A.: Statistical Mechanics of Complex Networks, Rev. Mod. Phys., Vol. 74 (2002)

Degree distributions – Electric Power Systems

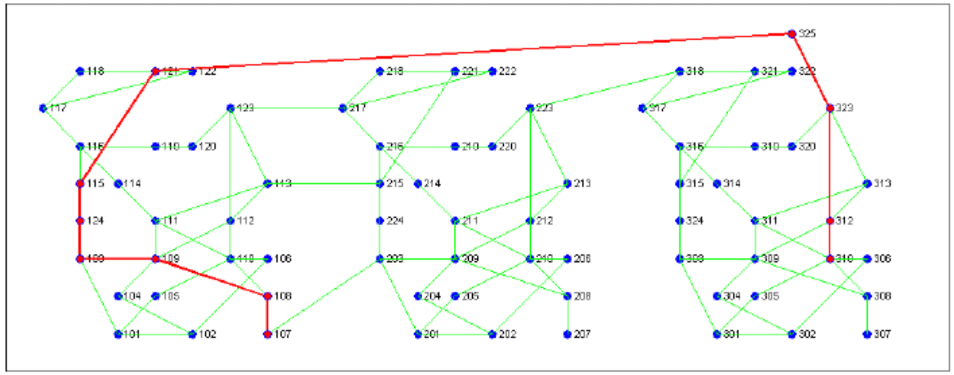


Cumulative distribution of the node degrees for the high-voltage transmission networks in Italy (full circles), Spain (diamonds) and France (squares). The empty circles represent the Italian „fine-grain“ network (from 380kV down to the distribution level).

Source: V. Rosato*, S. Bologna, F. Tiriticco*: Topological properties of high-voltage electrical transmission networks, Electric Power Systems Research, Vol. 77, 2007

Network Characteristics: Shortest Path and Diameter

Shortest Path between node 107 and 310



Different *algorithms* are used to find the shortest path ℓ_{ij} between two nodes *i* and *j*, e.g. Dijkstra‘s algorithm

Average path length: average of all shortest paths in the network:

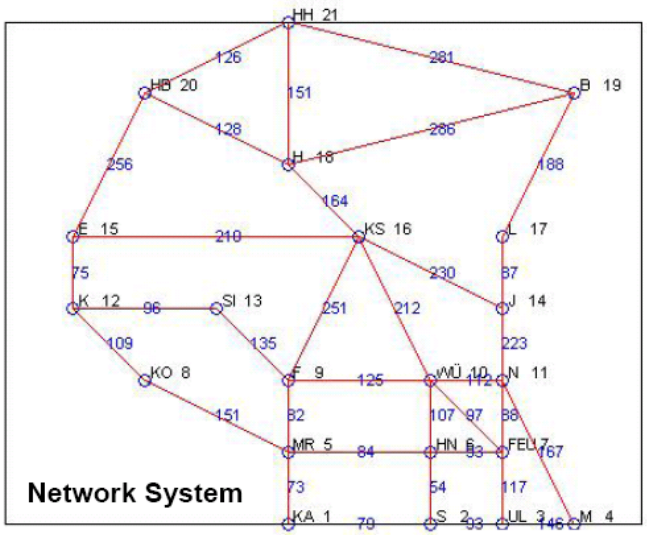
$$\langle \ell \rangle = \frac{1}{N(N-1)} \sum_{ij} \ell_{ij}.$$

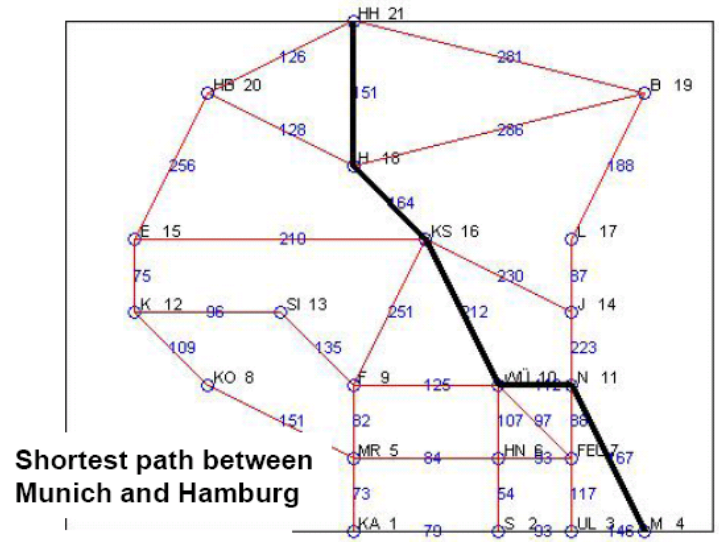
Its value becomes infinity in case of a network splitting, due to e.g. disruption.

Network diameter: $d_G = \max_{i,j} \ell_{ij}$

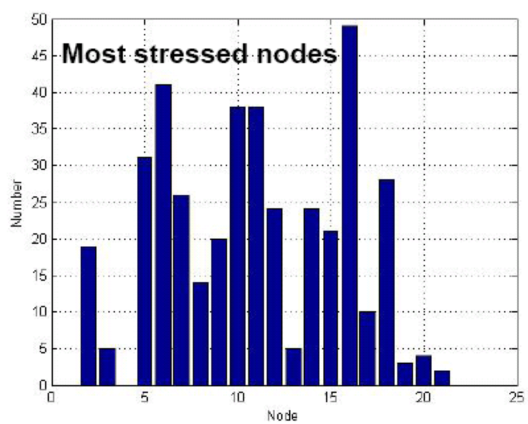
Network Characteristics: Shortest Path (II)

Example: Highway Network





Most stressed nodes: most utilized nodes in all shortest paths



Network Characteristics: Clustering Coefficient *C*

How interlinked are my friends?

C measures the density of connections around a particular node.
Suppose you have *z* close friends.

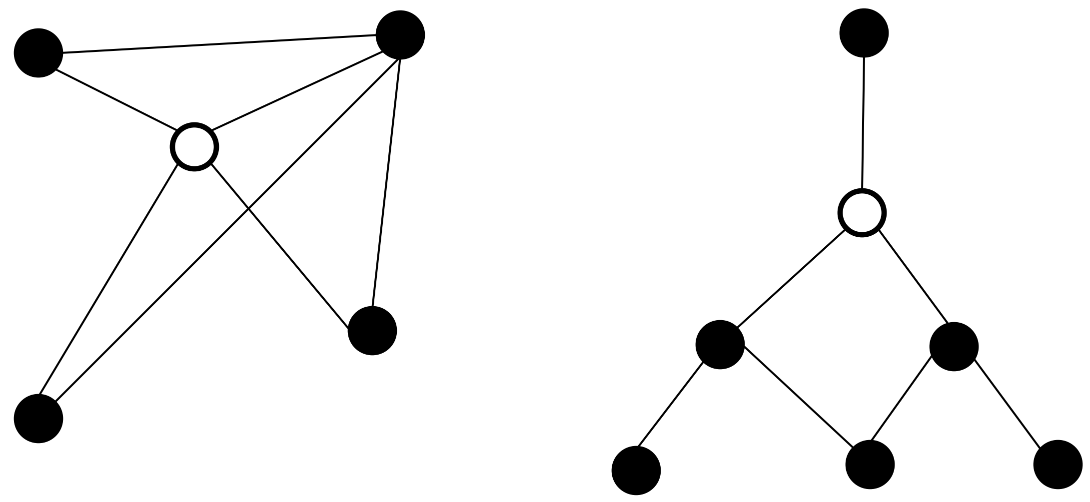
If they all are again friends among themselves there will be:

$$C_{max} = \frac{z(z-1)}{2}$$

links between them. Suppose that there are only *y* connections between them. *C* will be

$$C = \frac{2y}{z(z-1)} = \frac{y}{C_{max}}$$

Network Characteristics: Clustering Coefficient *C* (II)



How would you calculate the clustering coefficient for the white node?

Characteristics of Real-life Networks

	Network	Type	N	L	$\bar{k}$	ℓ	γ	C
Social	film actors	undirected	449 913	25 516 482	113.43	3.48	2.3	0.78
	company directors	undirected	7 673	55 392	14.44	4.60	–	0.88
	math coauthorship	undirected	253 339	496 489	3.92	7.57	–	0.34
	physics coauthorship	undirected	52 909	245 300	9.27	6.19	–	0.56
	biology coauthorship	undirected	1 520 251	11 803 064	15.53	4.92	–	0.60
	telephone call graph	undirected	47 000 000	80 000 000	3.16		2.1	
	email messages	directed	59 912	86 300	1.44	4.95	1.5/2.0	0.16
	email address books	directed	16 881	57 029	3.38	5.22	–	0.13
	student relationships	undirected	573	477	1.66	16.01	–	0.001
	sexual contacts	undirected	2 810				3.2	
Information	WWW `nd.edu`	directed	269 504	1 497 135	5.55	11.27	2.1/2.4	0.29
	WWW Altavista	directed	203 549 046	2 130 000 000	10.46	16.18	2.1/2.7	
	citation network	directed	783 339	6 716 198	8.57		3.0/–	
	Roget's Thesaurus	directed	1 022	5 103	4.99	4.87	–	0.15
	word co-occurrence	undirected	460 902	17 000 000	70.13		2.7	0.44
Technological	Internet	undirected	10 697	31 992	5.98	3.31	2.5	0.39
	power grid	undirected	4 941	6 594	2.67	18.99	–	0.080
	train routes	undirected	587	19 603	66.79	2.16	–	0.69
	software packages	directed	1 439	1 723	1.20	2.42	1.6/1.4	0.082
	software classes	directed	1 377	2 213	1.61	1.51	–	0.012
	electronic circuits	undirected	24 097	53 248	4.34	11.05	3.0	0.030
	peer-to-peer network	undirected	880	1 296	1.47	4.28	2.1	0.011
Biological	metabolic network	undirected	765	3 686	9.64	2.56	2.2	0.67
	protein interactions	undirected	2 115	2 240	2.12	6.80	2.4	0.071
	marine food web	directed	135	598	4.43	2.05	–	0.23
	freshwater food web	directed	92	997	10.84	1.90	–	0.48
	neural network	directed	307	2 359	7.68	3.97	–	0.28

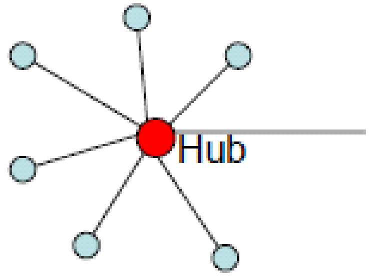
Exponent γ indicated only if it is a scale-free network

Source: Newman, SIAM Rev. 45, 167 (2003)

Random Failure and Attack Tolerance (I)

type of impact	exponential network	scale-free network
random	robust	extreme robust
malicious attack	robust	extreme vulnerable

scale-free network:



the chance to destroy the hub with a random attack is 1:7

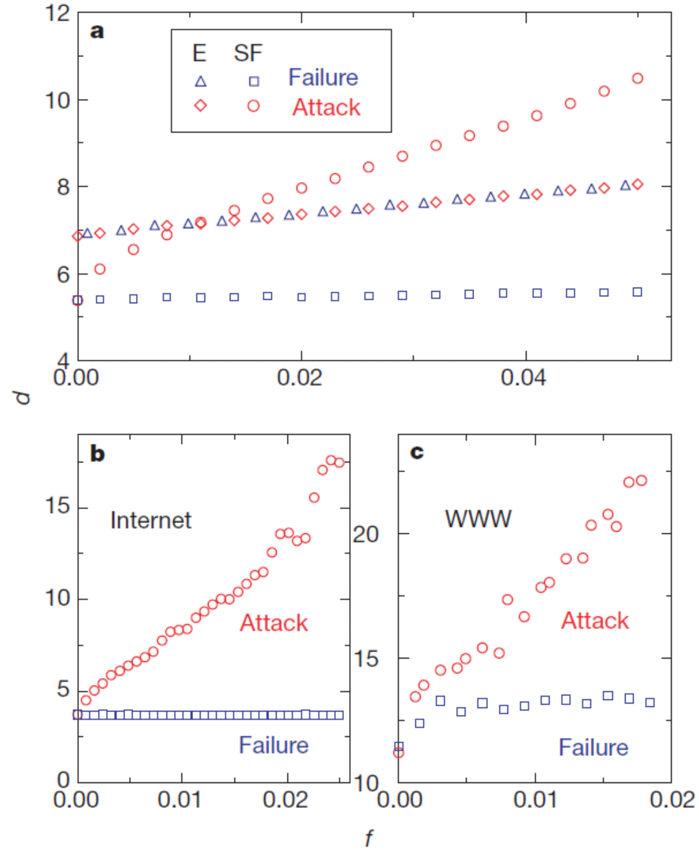
a malicious attack to the hub destroys the connection to six nodes

Random Failure and Attack Tolerance (II)

d: average length of the shortest paths between any two nodes in the network

f: fraction of the removed nodes

E: exponential
SF: scale-free



Albert, Jeong, Barabasi, Nature 406, 378, 2000

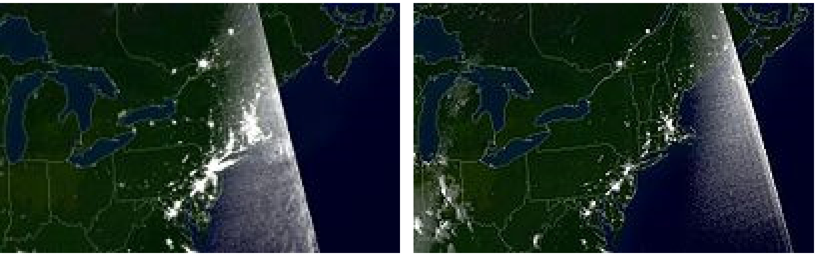
Picture: Flickr.com

Cascading Failures in Infrastructure Networks

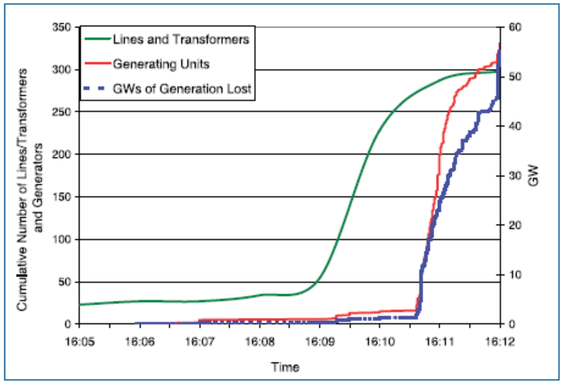
- are often the result of a relatively slow system degradation escalating into a fast avalanche of component failures, potentially leading to a complete loss of service
- while the first few outages might even be independent of each other, the causal failure chains usually become more pronounced in the course of the events, ending up in a fully cascading regime.

Example: The North American blackout 2003

The slow degradation started around noon with the outage of a system monitoring tool, further progressed during the afternoon through the independent outage of a generator and several transmission lines and finally evolved into the full cascade at around 16:00



Satellite image: day before and the night of the blackout



Component outages
(U.S.-Canada Power System Outage Task Force, 2004)

How to Analyze Cascading Failures?

Example: A simple load redistribution model (Motter and Lai)[1]

Model:

- The **load *L*** at a node is the total number of shortest paths passing through the node.
- The **capacity *C*** of a node is the maximum load that the node can handle. In man-made networks, the capacity is limited by cost. Thus, it is assumed that the capacity C_j of node j is proportional to its initial load L_j:

$$C_j = (1+\alpha)L_j, \quad j = 1,2,\ldots N$$

- The removal of nodes, in general, changes the distribution of shortest paths.
- The load at a particular node can then change; if it increases and becomes larger than the capacity C_j, the corresponding node fails.
- Any failure leads to a new redistribution of loads and, as a result, subsequent failures can occur $\longrightarrow$ ***cascading failure***
- Measure for the size of a cascade:

$$G = N'/N$$

N and N' are the number of nodes in the largest component before and after the cascade, respectively.

[1] Motter, A. E., Lai Y.-C., Cascade-based attacks on complex networks, Physical Review E 66, 065102

How to Analyze Cascading Failures?

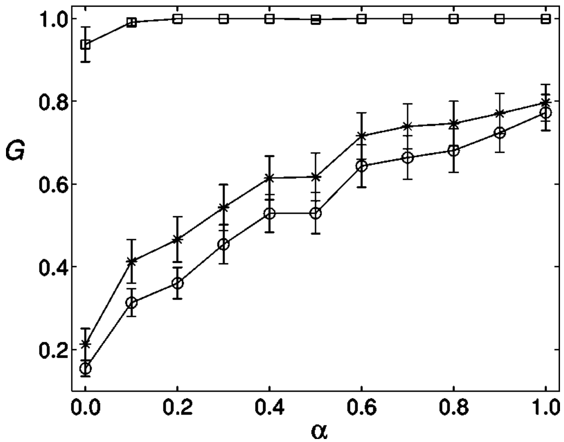
Example: A simple load redistribution model (Motter and Lai)

What can we learn from such a simple, abstract modeling approach?

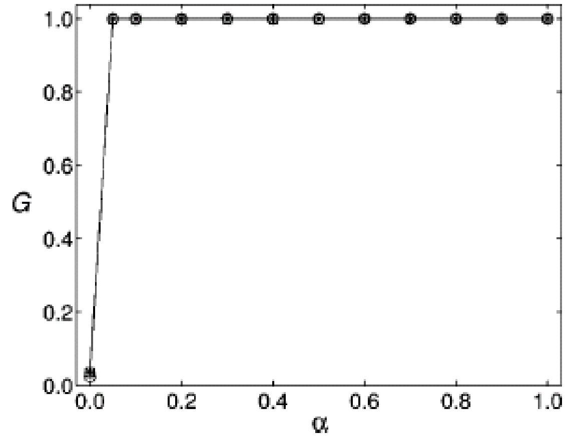
global cascades occur if

1. the network exhibits a highly heterogeneous distribution of loads (i.e. heterogeneous networks such as scale-free networks)
2. the removed node is among those with higher loads.

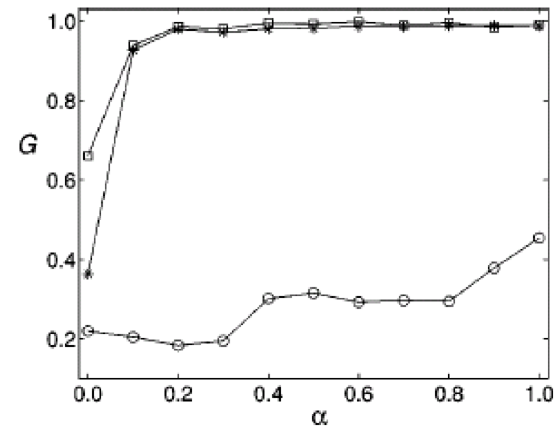
Otherwise, cascades are not expected.



scale-free networks (N~5000)



homogeneous networks (e.g. random graph) (N=5000)



western U.S. power transmission grid (N=4941)

The cascades are triggered by the removal of single nodes chosen at random (squares), or among those with largest degrees (asterisks) or highest loads (circles)

Recommended literature on network theory:

Dorogovtsev, S. N. and Mendes, J. F. F., "Evolution of Networks - from Biological Nets to the Internet and WWW”, (Oxford University Press, Oxford, 2003)

Barrat, A. and Barthelemy, M. and Vespignani, A., “Dynamical processes on complex networks”, (Cambridge University Press, 2008)

Newman, M., „The structure and function of complex networks”, (SIAM Rev. 45, 167, 2003)