NETWORKS

Mario Nicodemi

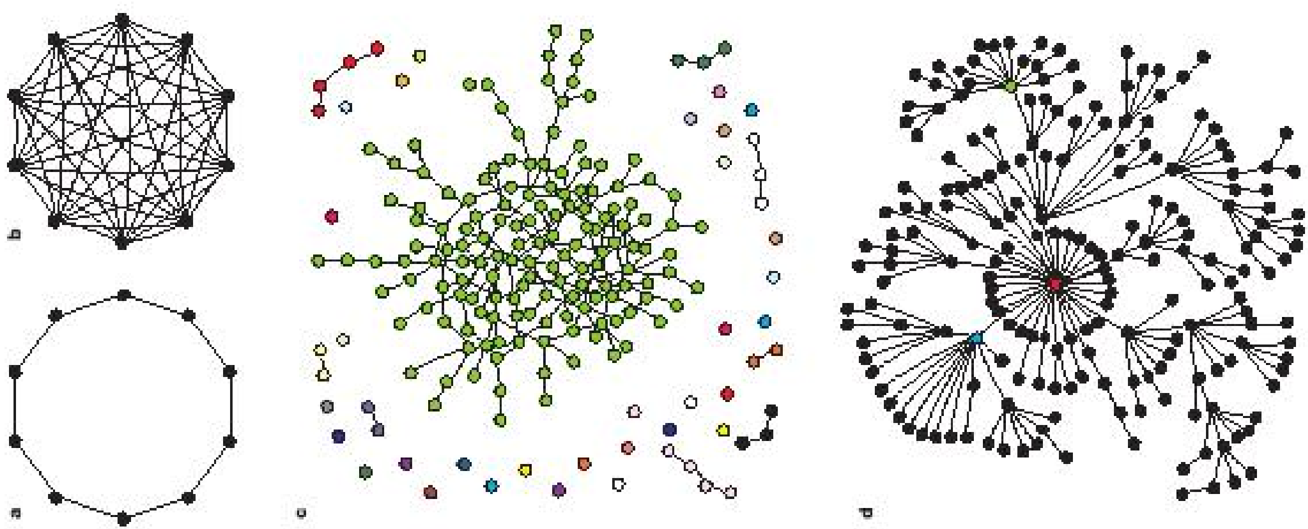
Complexity Science & Theor. Phys., University of Warwick"

Topic:

- Basic concepts on **networks** and graph structure

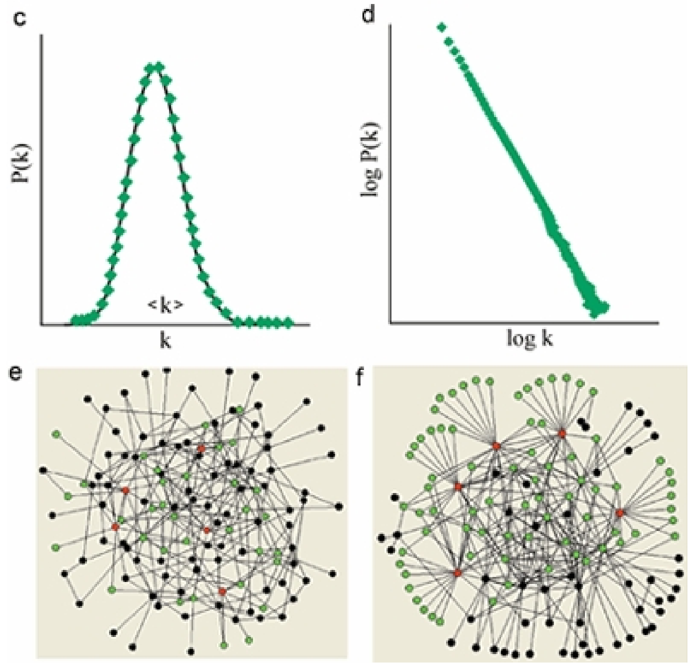
Ref.s: Barabasi&Bonabeau, Sci.Am. ('03). Watts&Strogatz, Nat. ('98). Strogatz, Nat. ('01).

□ Network architectures



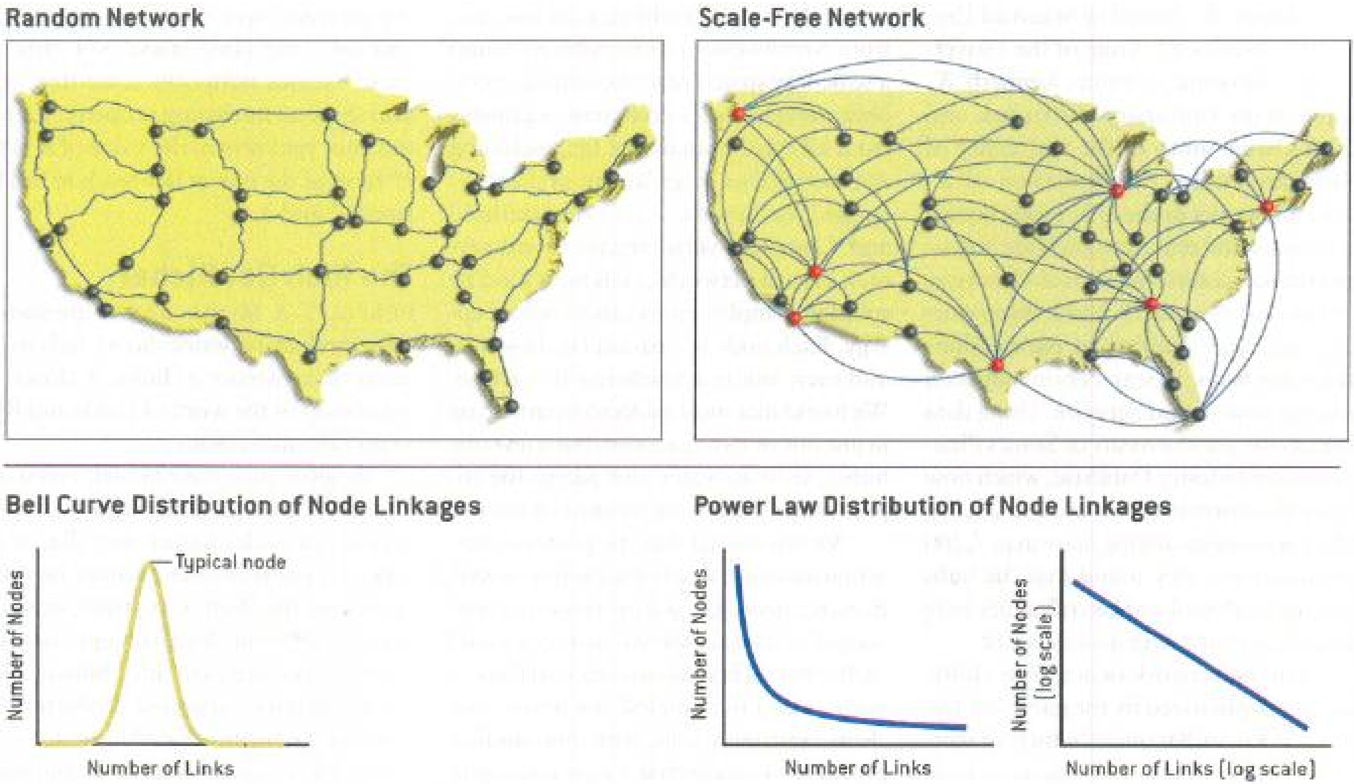
a Ring of ten nodes linked to nearest neighbours. **b** Fully connected network. **c** Random graph with N nodes, joined in pairs with m links. A single giant component appear if $m > N/2$ (here $N = 200$, $m = 193$). No dominant hubs. The degree distribution is Poisson. **d** Scale-free graph, grown by attaching new nodes at random to existing nodes (probability of attachment proportional to the degree of the target node). Hubs form & the degree distribution has a heavy tail. Colours indicate the three nodes with most links.

□ Random v.s. Scale free nets



e In this random-graph N nodes are linked in pairs with a probability p, so they have approx. the same number of links.
f The majority of nodes in a scale-free network have 1 or 2 links, but a few nodes have a large number; this guarantees that the system is fully connected. More than 60% of nodes (green) are reached from the 5 most connected nodes (red) compared with only 27% in the random network. This shows the key role that hubs play in the scale-free network.
c The network connectivity is characterized by the probability $P(k)$ that a node has k links. For random graphs $P(k)$ is peaked at $k =< k >$ and decays approx. as $P(k) \sim \exp(-k)$ for large k.
d In a scale-free network $P(k)$ has a power law $P(k) \sim k^{-\gamma}$ tail.

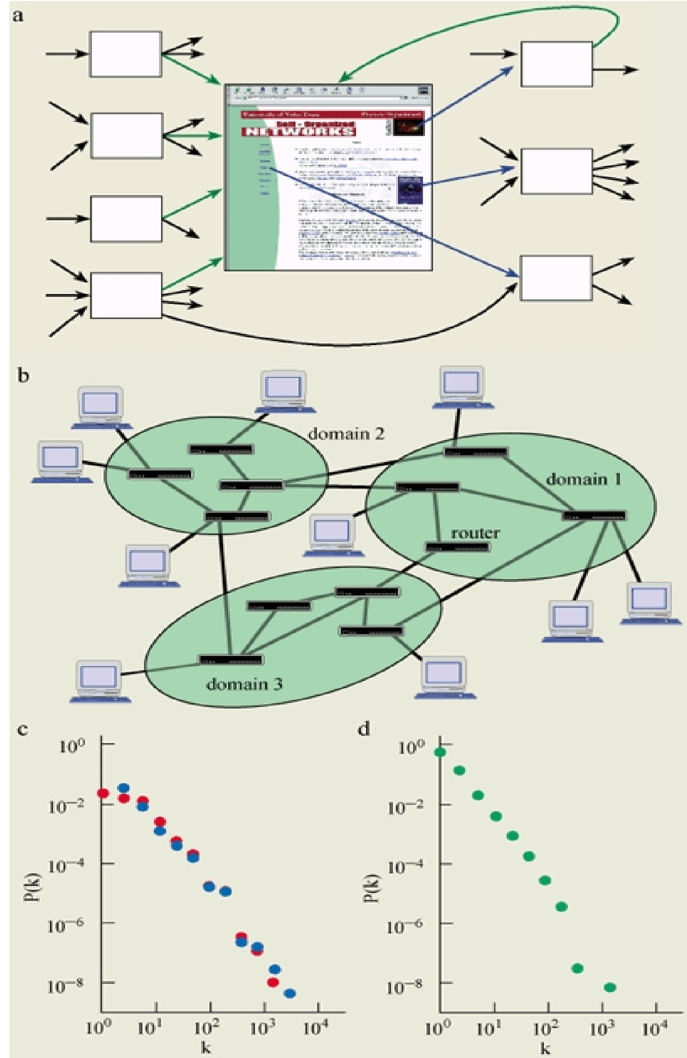
☐ Random v.s. Scale free nets



Main network features: **hubs** absence/presence; **degree distribution**.

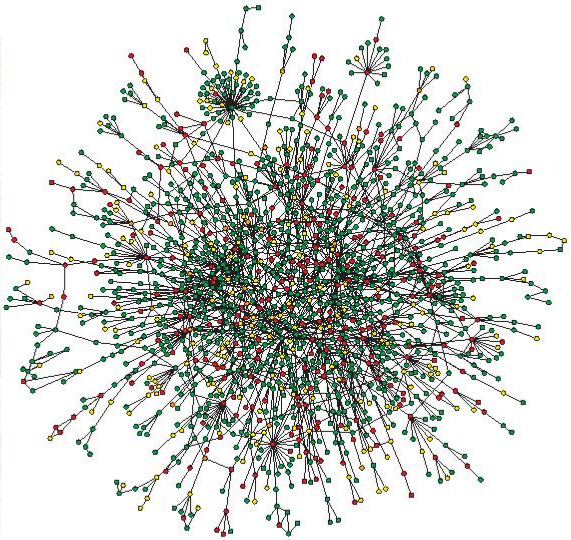
☐ Example: WWW

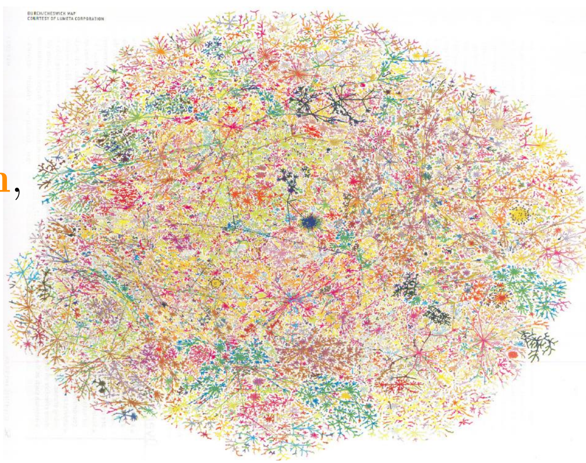
a The nodes of the WWW are web documents, identified by an unique uniform resource locator (URL). Outgoing links to other pages are shown as blue arrows, incoming links are green arrows.
b The Internet is a net of routers connected by physical or wireless links and are grouped into domains.
c The probability that a Web page has k_{in} (blue) or k_{out} (red) links follows a power law.
d The degree distribution of the Internet is power law, where k (green) denotes the number of links a router has to other routers.



Examples of Scale-Free Networks

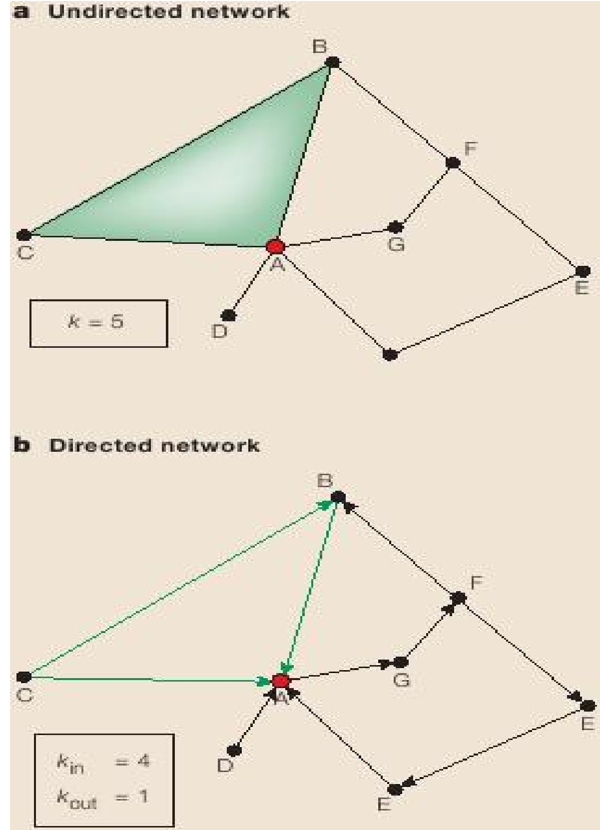
NETWORK	NODES	LINKS
Cellular metabolism	Molecules involved in burning food for energy	Participation in the same biochemical reaction
Hollywood	Actors	Appearance in the same movie
Internet	Routers	Optical and other physical connections
Protein regulatory network	Proteins that help to regulate a cell's activities	Interactions among proteins
Research collaborations	Scientists	Co-authorship of papers
Sexual relationships	People	Sexual contact
World Wide Web	Web pages	URLs





Top right Yeast protein interaction network: largest cluster shown ($\sim 78\%$ of all proteins). Node colour $\Leftrightarrow$ phenotypic effect of removing that protein (**lethal**, **non-lethal**, **slow growth**, **unknown**): approx. linear correl. between lethality&connectivity (*?!*).
Bottom right This Internet map (Feb. '03) traces the shortest routes from a test Web site to about 10^5 others, using like colors for similar Web addresses.

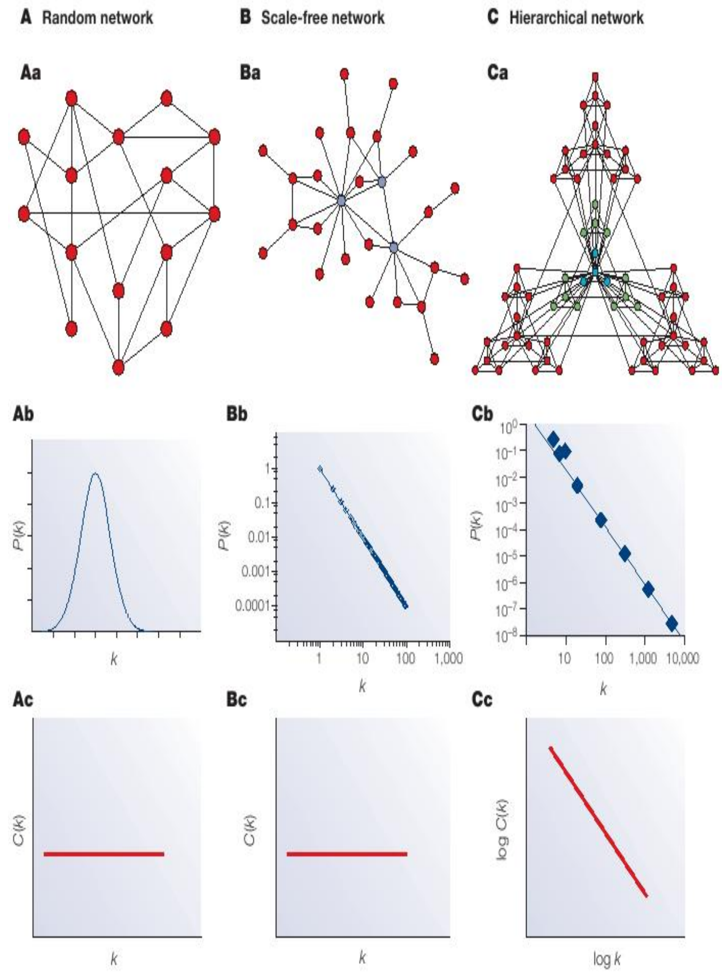
□ Basic concepts



Degree (or connectivity), k, of a node is the number of its links. F.ex., in the figure, node A has $k = 5$.
Degree distribution, $P(k)$, is the probability that a node has k links.
Path lenght, a measure of node distance, is the path with the smallest number of links between 2 nodes (in directed networks, the distance from A to B can be different from the one from B to A). Its average over the nodes is the **mean path length**, $< l >$, a measure of a network overall navigability.
Clustering coefficient In a networks, if node A is connected to B, and B is connected to C, A may also have a direct link to C. This is quantified by the clustering coefficient $C_I = 2n_I/k_I(k_I - 1)$, where n_I is the number of links connecting the k_I neighbours of node I to each other. So, C_I is the num. of 'triangles' through node I ($k_I(k_I - 1)/2$ is the max possible num. of triangles through I). F.ex., only one pair of node A five neighbours in the figure are linked together (B and C), which gives $n_A = 1$ and $C_A = 2/20$; none of node F neighbours link to each other, giving $C_F = 0$. The **average clustering coefficient**, $< C >$, characterizes the overall tendency of nodes to form clusters or groups.
An important measure of the net structure is $C(k)$, the average clustering coefficient of nodes with k links. For many real networks $C(k)\ k^{-1}$, which is an indication of a network hierarchical character. $P(k)$ and $C(k)$, capturing generic features, are used to classify networks, as $< k >$, $< l >$ and $< C >$ characterize a specific network.

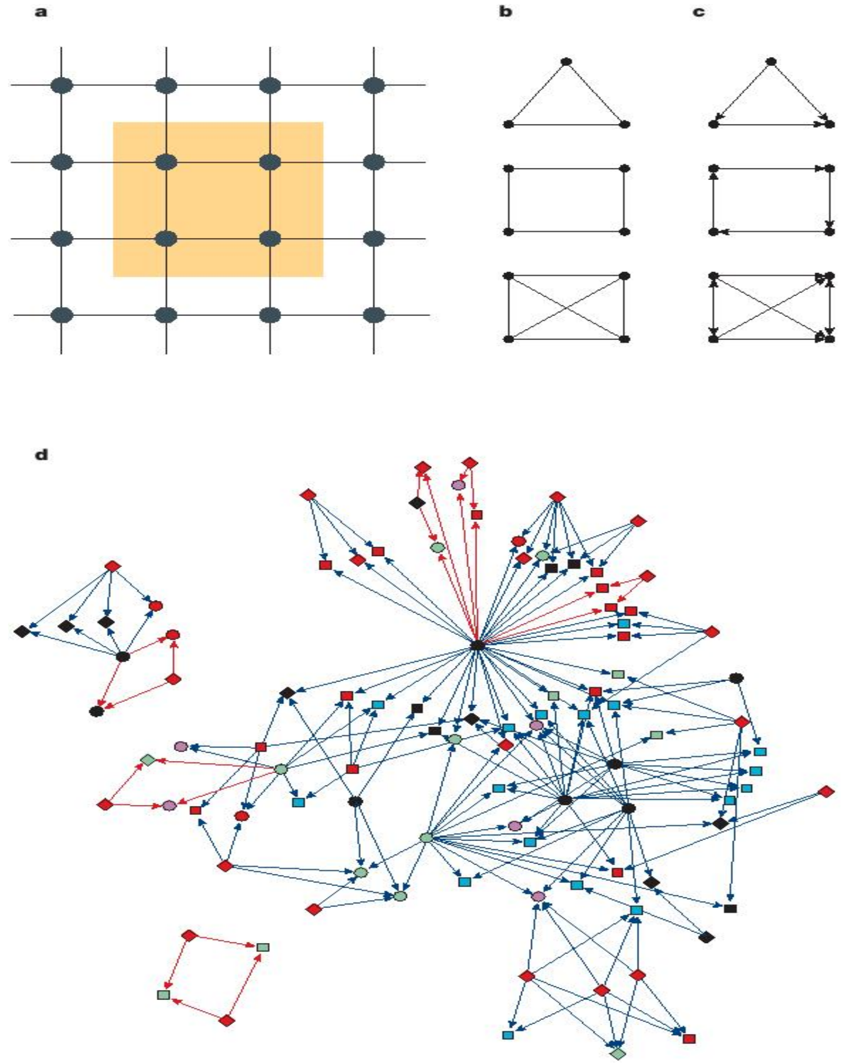
☐ Network structures

Random networks The Erdös-Rényi net has N nodes connected in pairs with probability p, which creates a graph with approximately $pN(N-1)/2$ random links. $P(k)$ is Poisson (Fig. Ab), with a typical $< k >$; $C(k)$ is independent on k (Fig. Ac); the mean path length is $l \sim \ln N$ ('small-world').
Scale-free networks have $P(k) \sim k^{-\gamma}$. Few nodes are highly connected (hubs, blue nodes in Ba). The Barabasi-Albert grown net (shown) does not have an inherent modularity, so its $C(k)$ is indep. of k (Fig. Bc). Scale-free nets with degree exponents $2 < \gamma < 3$ are observed in most nets. $l \sim \ln \ln N$, shorter than Random Nets.
Hierarchical networks have coexistence of modularity, local clustering and scale-free topology. They are formed by clusters combined in an iterative manner, with hierarchical hubs and $C(k) \sim k^{-1}$.



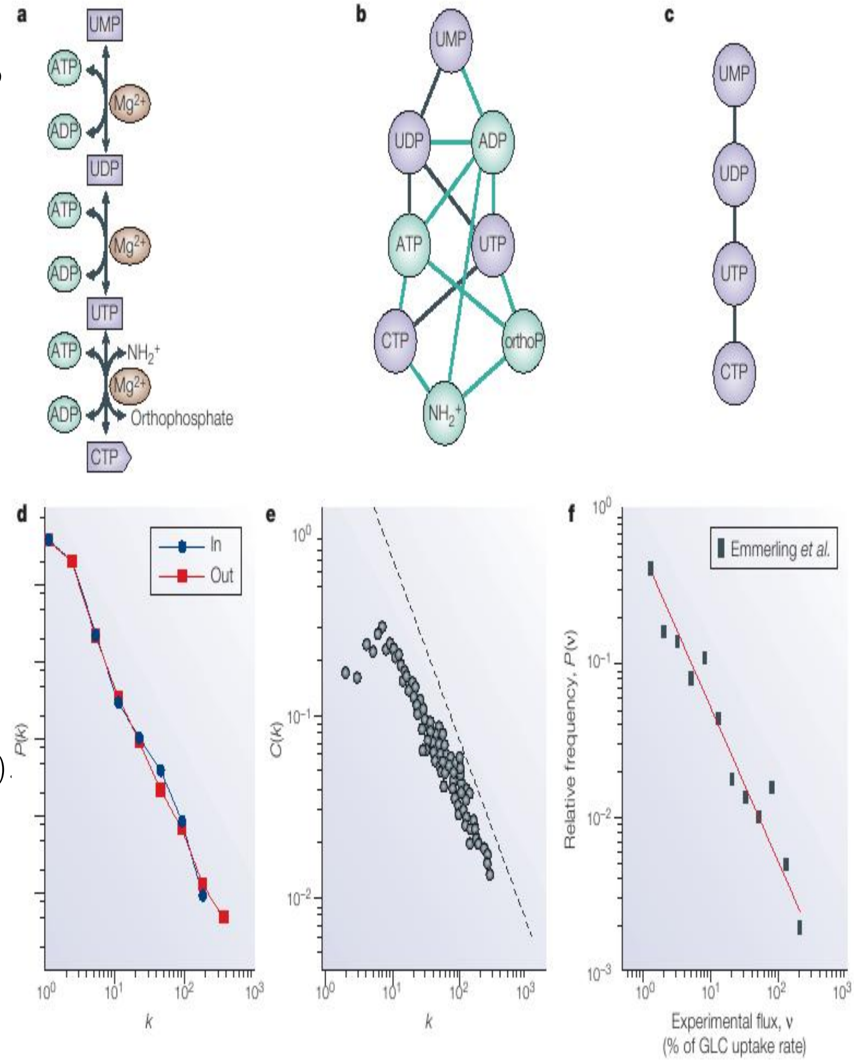
□ Motifs

Subgraph is a subset of nodes connected to each other in a specific wiring diagram. F.ex., in the figure four nodes that form a square (yellow) represent a subgraph of a square lattice. The number of distinct subgraphs grows exponentially with N. **Motifs** are subgraphs over represented as compared to a randomized version of the same network (keeping the number of nodes, links and $P(k)$ unchanged). **Motif clusters** Clustering of motifs into motif clusters seems to be a general property of all real networks.

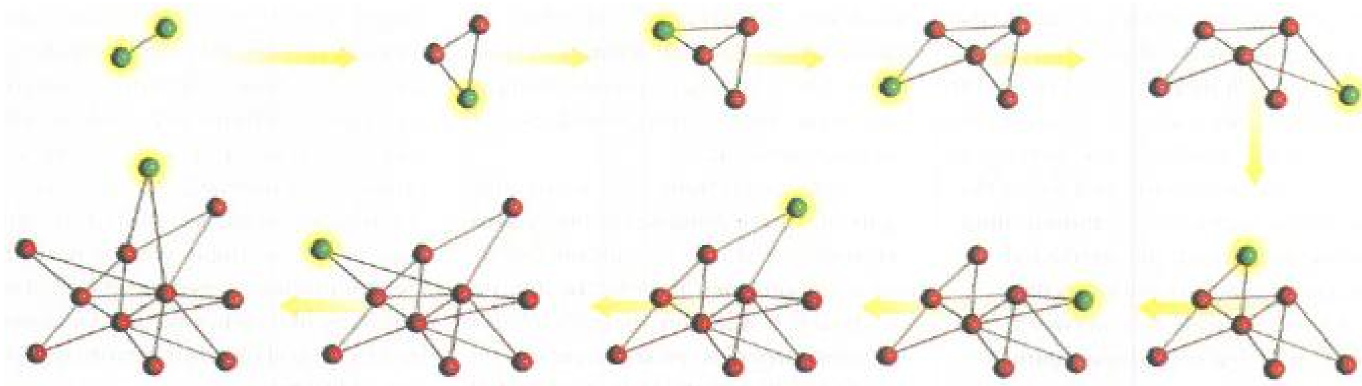


☐ Metabolic nets

A simple pathway (catalysed by Mg2+-dependant enzymes) is illustrated **(a)**. In the most abstract approach **(b)** all interacting metabolites are considered equally. The links between nodes represent reactions that interconvert one substrate into another. For many applications it is useful to ignore co-factors, such as ATP, which results in **(c)** a graph with only the main source metabolites to the main products.
$P(k)$ of metabolic networks **(d)** and clustering coefficient $C(k)$ **(e)**. (data shown in d and e are averages over 43 organisms, Barabasi Nat.Bio.). The flux distribution in the central metabolism of Escherichia coli follows a power law, which indicates that most reactions have small metabolic flux, whereas a few reactions, with high fluxes, carry most of the metabolic activity **(f)**.



□ Preferential attachment growth

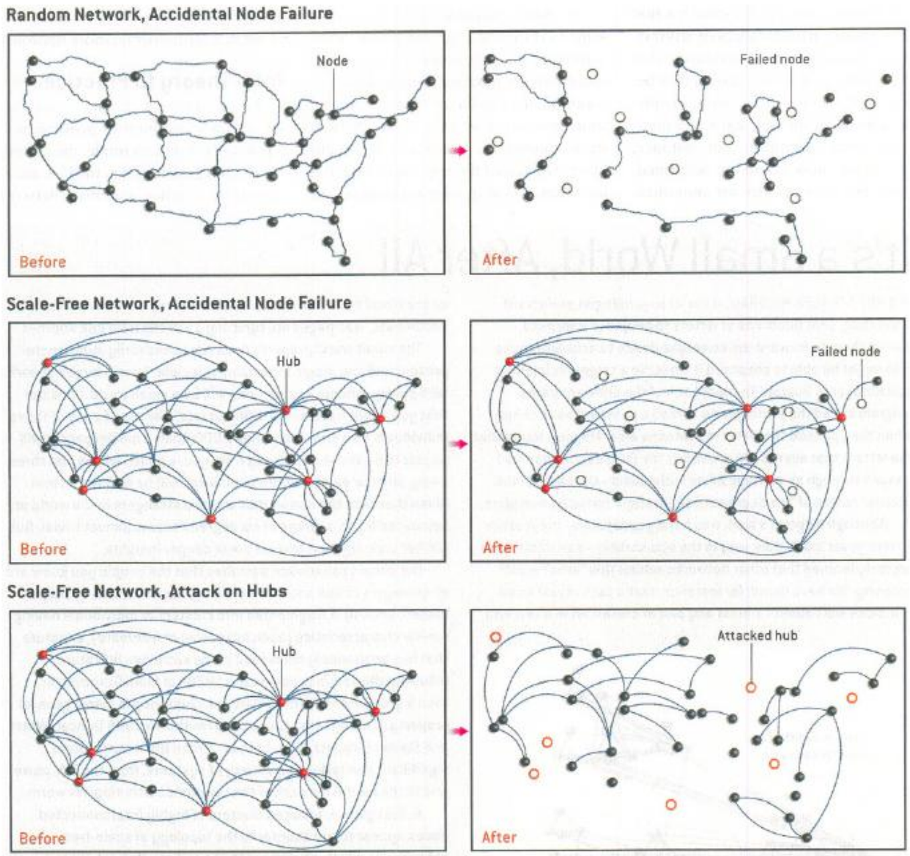


- Example of SFN incremental growth from 2 to 11 nodes. When deciding where to establish a link, a new node (green) prefers to attach to an existing node (red) that already has many connections. These basic mechanisms - growth and preferential attachment - lead to the system's being dominated by hubs, and power law degree distribution ($\gamma = 3$).

- Variant I: if the attachm. prob., p_i, to node i is not linear in k_i, only one single major hub emerges (*"winner take all"*, no scale free).

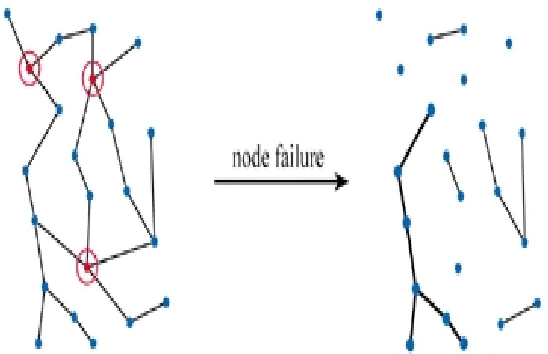
- Variant II: if $p_i \sim \eta_i k_i$, with η_i rnd from $\mathcal{P}(\eta)$, the γ depends on $\mathcal{P}$.

□ Nets vulnerability

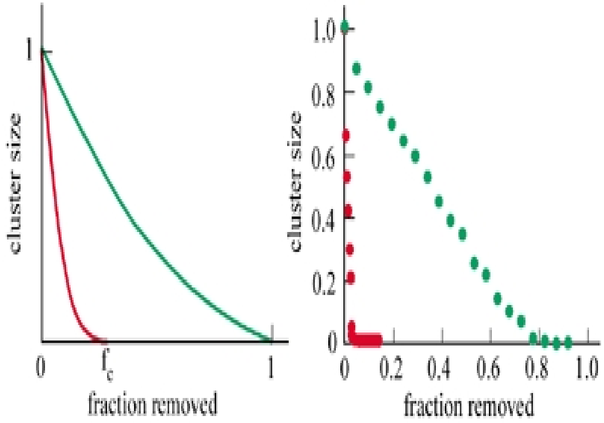


Accidental failure of nodes in a random network (top panels) can fracture the system into non-communicating islands. Scale-free networks are more robust to such failures (middle panels). But they are highly vulnerable to attacks against hubs (bottom panels).

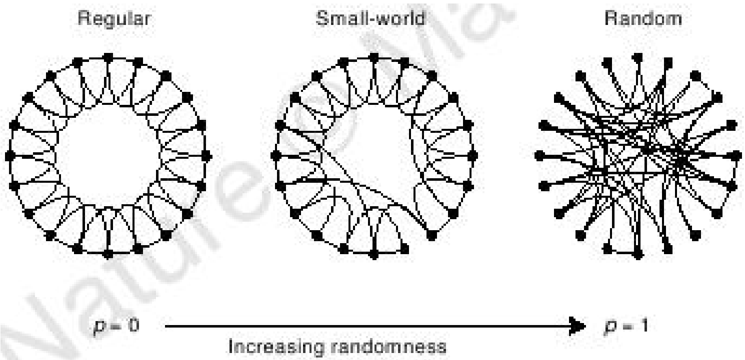
□ Robustness to failure

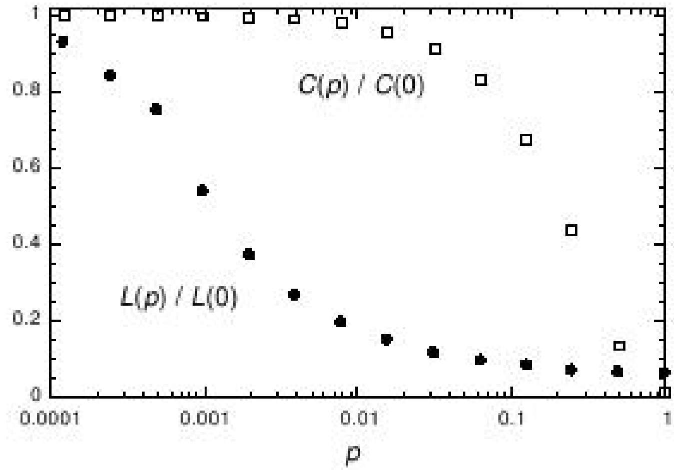


Top Removing just 3 (i.e., 14%, circled) nodes breaks the network into several smaller clusters.
Bot.Left Percolation theory predicts that a random network (red) breaks into tiny clusters if a critical fraction, f_c, of nodes is removed. In scale-free nets the cluster size only falls to zero when all the nodes have been disconnected (green). However, if the most-connected nodes are removed then the scale-free net break at a small f_c.
Bot.right By randomly removing domains from the Internet, more than 80% of the nodes have to fail before the network fragments (green). However, the same effect is achieved by removing just a small fraction of the most connected nodes are targeted (red).



☐ 'Small World'





Left Random rewiring of a ring of N vertices linked to k nearest neighbours. Choose a vertex and an edge to its n.n.. With probability p, reconnect this edge to a random vertex (duplicates forbidden). Repeat clockwise around the ring until one lap is completed. Next, consider edges to 2nd n.n., etc... (as there are $Nk/2$ edges in the graph, rewiring stops after $k/2$ laps.) For intermediate p, the graph is a **small-world network**: highly clustered like a regular graph, yet with a small path length, like a random graph.

Right The path length $L(p)$ is the averaged number of edges in the shortest path between two vertices. The clustering coefficient $C(p)$ is the average number of 'triangles' in the graph. As p increases, $L(p)$ rapidly drops, corresponding to the onset of the small-world phenomenon. Meanwhile, $C(p)$ is almost constant at its value for the regular lattice, indicating that the transition to 'small world' is almost undetectable at the local level.