Hierarchical Organization of Complex Networks

E. Ravasz, A. L. Barabási, Z. N. Oltvai

University of Notre Dame

Northwestern University

• Starting Point: how is the Metabolism Organized?

Modular

Partitioned into spatially or chemically isolated functional blocks composed of several cellular components and carrying discrete biological function [1].

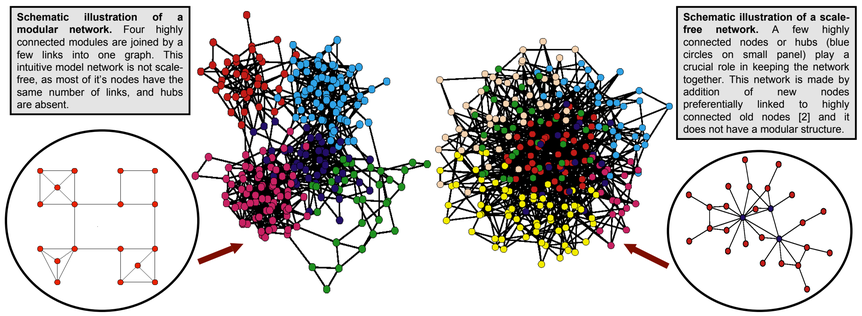
Modularity and scale-free property are antagonistic.

In the presence of hubs one cannot have isolated modules.

Scale-free

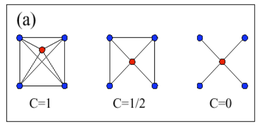
The components are dynamically interconnected, functional properties being encoded into a complex web of molecular interactions. A hierarchy of highly connected nodes with degree distribution decaying as a power law ($P(k) \sim k^{\gamma}$, $\gamma \approx 2.2$) integrate all substrates into a single scale-free web [2].

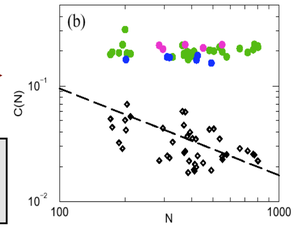
Schematic illustration of a modular network. Four highly connected modules are joined by a few links into one graph. This intuitive model network is not scale-free, as most of it's nodes have the same number of links, and hubs are absent.



Schematic illustration of a scale-free network. A few highly connected nodes or hubs (blue circles on small panel) play a crucial role in keeping the network together. This network is made by addition of new nodes preferentially linked to highly connected old nodes [2] and it does not have a modular structure.

• Clustering in the Metabolism





(a) Clustering coefficient. A node's clustering coefficient is the fraction of realized links between its neighbors. It offers a measure of the degree of interconnectivity in a node's neighborhood [3].

(b) The average clustering coefficient of 43 organisms. The measured values are independent of the system size (N) and an order of magnitude larger than the prediction of the simple scale-free model (dashed line) [3].

Size-independent clustering coefficient

The average clustering coefficient of the metabolic networks of 43 different organisms is independent of their size, in contrast with the predictions of the simple scale-free model [3].

• Hierarchy in the Metabolism

Clustering coefficient decreases with connectivity

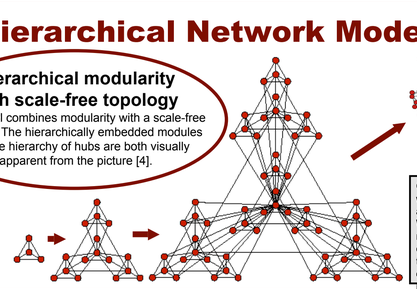
The clustering coefficients of 43 metabolic networks show power-law scaling, indicating the presence of hierarchical modularity in cellular metabolism.

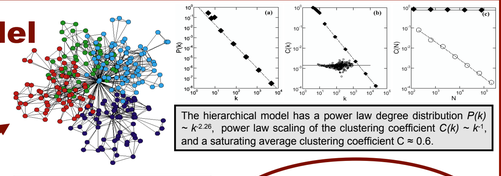
Species belonging to archaea, bacteria and eukaryotes are shown in purple, green and blue, respectively. The diamonds denote C for a scale-free network with the same parameters (node and link number) as observed in the 43 organisms.

• Hierarchical Network Model

Hierarchical modularity with scale-free topology

The model combines modularity with a scale-free topology. The hierarchically embedded modules and the hierarchy of hubs are both visually apparent from the picture [4].



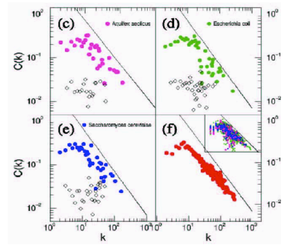


The hierarchical model has a power law degree distribution $P(k) \sim k^{-2.25}$, power law scaling of the clustering coefficient $C(k) \sim k^{-1}$, and a saturating average clustering coefficient $C \approx 0.6$.

Constructing the network. We start with four fully interconnected nodes and make three copies of this module. We connect each of the 3*3 external nodes of the new copies to the old central node. We repeat the copying of the now 16-node graph and connect all external nodes to the original middle node.

Clustering coefficient follows a power law

The scaling of the clustering coefficient indicates the presence of a hierarchy of modules of increasing size and decreasing interconnectedness [5].



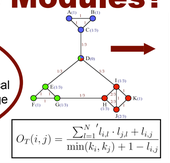
(d-f) The dependence of the clustering coefficient on the node's degree. **(c)** *Aqudex aeolicus* (archaea), **(d)** *Escherichia Coli* (bacterium), **(e)** *Saccharomyces cerevisiae* (eukaryote), (f) $C(k)$ averaged over the 43 organisms. The inset displays all 43 species together, the lines correspond to $C(k) \sim k^{-1}$.

[1] L.H.Hartwell, J.J.Hopfield, S.Leibler, A.W.Murray, *Nature* **402**, C47 (1999); H.Kitano, *Science* **295**, 1662 (2002); Y.I.Wolf, G.Karev, E.V.Koonin, *Bioessays* **24**, 105 (2002).
[2] H.Jeong, B.Tombor, R.Albert, Z.N.Oltvai, A.-L.Barabási, *Nature* **407**, 651 (2000); A.Wagner, D.A.Fell, *Proc.R.Soc.London Ser.B* **268**, 1803 (2001);
[3] D.J.Watts, S.H.Strogatz, *Nature* **393**, 440 (1998); R.Albert, A.L.Barabási, *Rev.Mod.Phys.* **74**, 47 (2002).
[4] A.-L.Barabási, E.Ravasz, T.Vicsek, *Physica A* **299**, 559 (2001).
[5] S.N.Dorogovtsev, A.V.Goltsev, J.F.F.Mendes, *Phys.Rev.E* **65**, 066122 (2002).
[6] E.Ravasz, A.L.Somera, D.Mongru, Z.N. Oltvai, A.L.Barabási, *Science* **297**, 1551 (2002).
[7] M.B.Eisen, P.T.Spellman, P.O.Brown, D.Botstein, *Proc.Natl.Acad.Sci.U.S.A.* **95**, 14863 (1998).
[8] R.Overbeek *et al.*,*Nucleic Acids Res.* **28**, 123 (2000).
[9] A.-L.Barabási, R.Albert, *Science* **286**, 509 (1999); R. Ferrer i Cancho, R.V. Solé, *Proc.R.Soc..London Ser.B* **268**, 2261 (2001);
[10] R.Albert, H.Jeong, A.-L.Barabási, *Nature* **401**, 130 (1999); A. Vázquez, R.Pastor-Satorras, A. Vespignani, *Phys.Rev.E* **65**, 066130 (2002).
[11] E. Ravasz, A.-L. Barabási, *Phys.Rev.E* **67**, 026112 (2003).

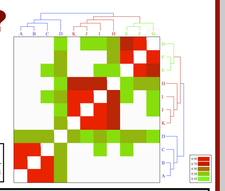
• How to Find the Modules?

Average linkage clustering on a network

We defined an overlap coefficient similarity between any pair of nodes in the network based on the topological overlap of their first neighbors. Then we applied average linkage hierarchical clustering [7] to group the nodes in a hierarchical fashion.



$$O_T(i,j) = \frac{\sum_{l=1}^{N} l_{i,l} \cdot l_{j,l} + l_{i,j}}{\min(k_i, k_j) + 1 - l_{i,j}}$$

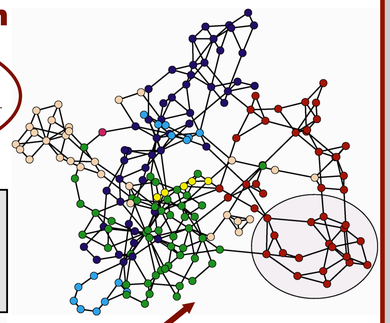


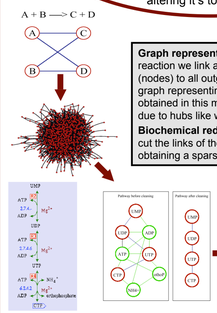
The overlap is a measure of common neighbors relative to the links of the smaller node ($l_{i,j}$ is 1 only if i and l are linked, k_i is the degree of node i.) The overlap values are color coded in the matrix, and used to obtain the hierarchical tree. (Red and blue numbers on the example show the overlap values and the clustering coefficients corresponding to the nodes [6].

• Reduced Graph Representation of the *E. Coli* Metabolism

Biochemical and topological reduction

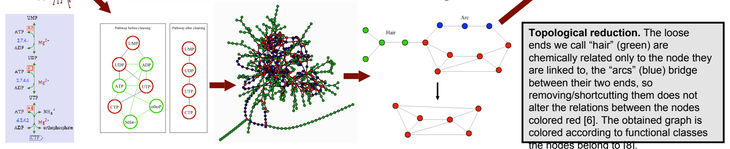
Removing highly connected common metabolites from reactions reveals the relevant biochemical connections. Removing non-branching arcs and loose ends simplifies the network without altering it's topology [6].





Graph representation of a reaction. In each reaction we link all incoming substrates (nodes) to all outgoing ones [2]. The red graph representing the E. Coli metabolism obtained in this manner is highly integrated due to hubs like water, ATP, ADP, P etc.

Biochemical reduction. In each pathway we cut the links of these helper substrates, thus obtaining a sparse graph [6].

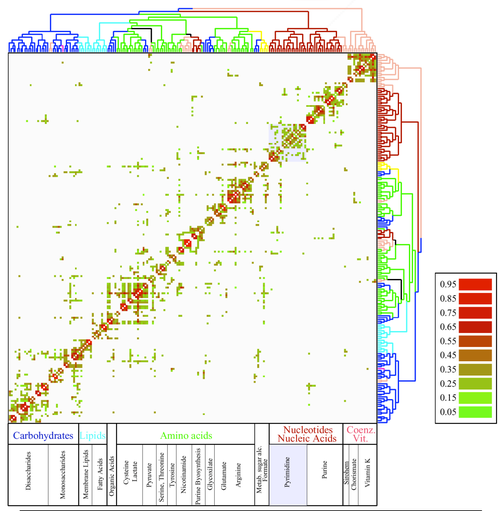


Topological reduction. The loose ends we call "hair" (green) are chemically related only to the node they are linked to, the "arcs" (blue) bridge between their two ends, so removing/shortcutting them does not alter the relations between the nodes colored red [6]. The obtained graph is colored according to functional classes the nodes belong to [8].

• Modules of the *E. Coli* Metabolism

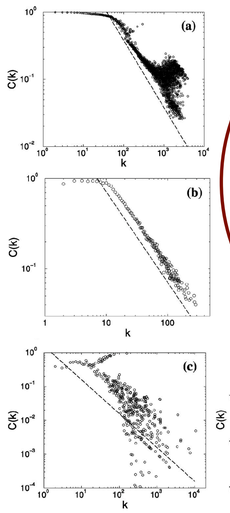
Modularity and biological function

Clustering reveals the hierarchically modular structure of the metabolism, visualized on the tree as well as in the overlap matrix structure. The higher level modules correlate with known functional classes [6].



Hierarchical tree representing the E. Coli metabolic network. The color coding of the branches corresponds to known functional classes of the metabolites [8], and the matrix represents the overlap between substrates. A good example of hierarchically modular structure is the *carbohydrate* branch, with the highly overlapping *dissaccharides* branch embedded into it [6].

• Hierarchy all around us!



The scaling law for the clustering coefficient holds for many networks!

Nodes in the actor network, semantic web, WWW and the Internet at AS level all have clustering coefficients that decrease at least two orders of magnitude with their degree [11]. This scaling indicates the presence of hierarchically embedded communities.

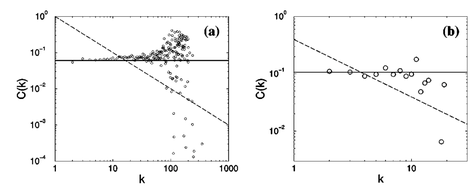
The scaling of $C(k)$ with k for four hierarchical networks. **(a)** Actor network, two actors being connected if they acted in the same movie according to the www.IMDB.com database [9]; **(b)** The semantic web, connecting two English words if they are listed as synonyms in the Merriam Webster dictionary [9]; **(c)** The World Wide Web, based on the data collected in Ref. [10]; **(d)** Internet at the autonomous system level, each node representing a domain, connected if there is a communication link between them [10]. The dashed line in each figure has slope -1.

Communities and clustering in complex networks

Complex networks such as the WWW, social networks, actor networks are fundamentally modular. Groups of co-workers and organizations in social networks, web page systems of similar topics on the WWW, actors playing in movies of the same genre in the actor network are all examples of such modules [9,10]. This modularity is the underlying reason for the high average clustering coefficient measured on these networks.

Some physically constrained networks do not show hierarchy

The Internet viewed at the router level, as well as the power grid network of the western US do not obey the scaling law for the clustering coefficient [11]. It is very likely that the length constrains on links, related to their cost, leads to the absence of hierarchical organization.



The $C(k)$ function for two non-hierarchical networks. **(a)** Internet at router level [10]; **(b)** The power grid of Western United States [3];