COMP 150 CSB –
Computational Systems Biology

Modularity Analysis in Metabolic Networks

Soha Hassoun

Department of Computer Science (primary)
Department of Chemical and biological Engineering
Department of Electrical and Computer Engineering

Tufts UNIVERSITY

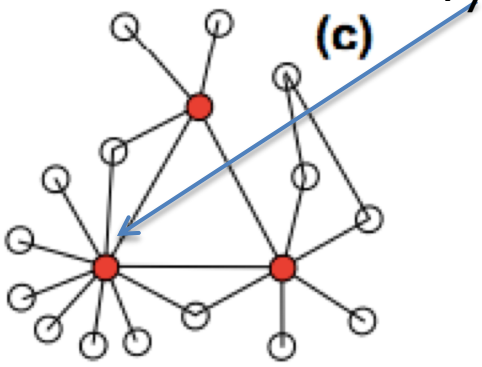
Reading

- Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., & Barabási, A. L. (2000). The large-scale organization of metabolic networks. Nature, 407(6804), 651

- Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., & Barabási, A. L. (2002). Hierarchical organization of modularity in metabolic networks. *science*, *297*(5586), 1551-1555

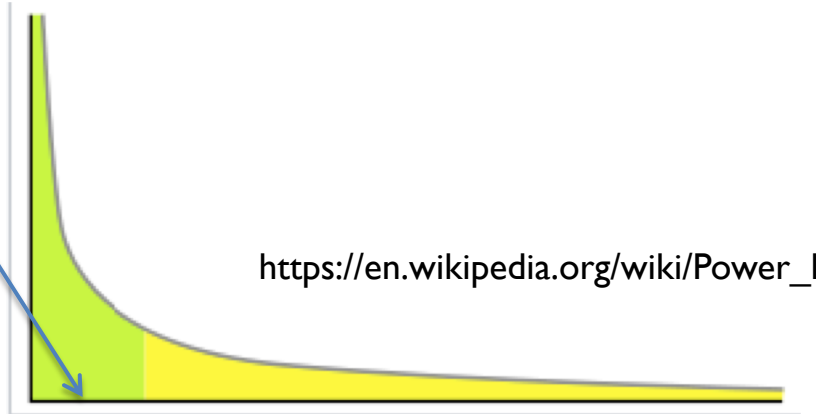
Observation - metabolic networks

- There are a few metabolites, referred to as hubs, that participate in many reactions, *and* interact with many substrates
  - e.g. Pyruvate has lots of connections
- Treat a metabolic network as a graph
  - For each node i, count the number of neighboring nodes that are connected to i via an edge.
  - Draw the corresponding histogram
  - Observation: some metabolites have lots of neighbors, while others have few



Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., & Barabási, A. L. (2000). The large-scale organization of metabolic networks. Nature, 407(6804), 651.

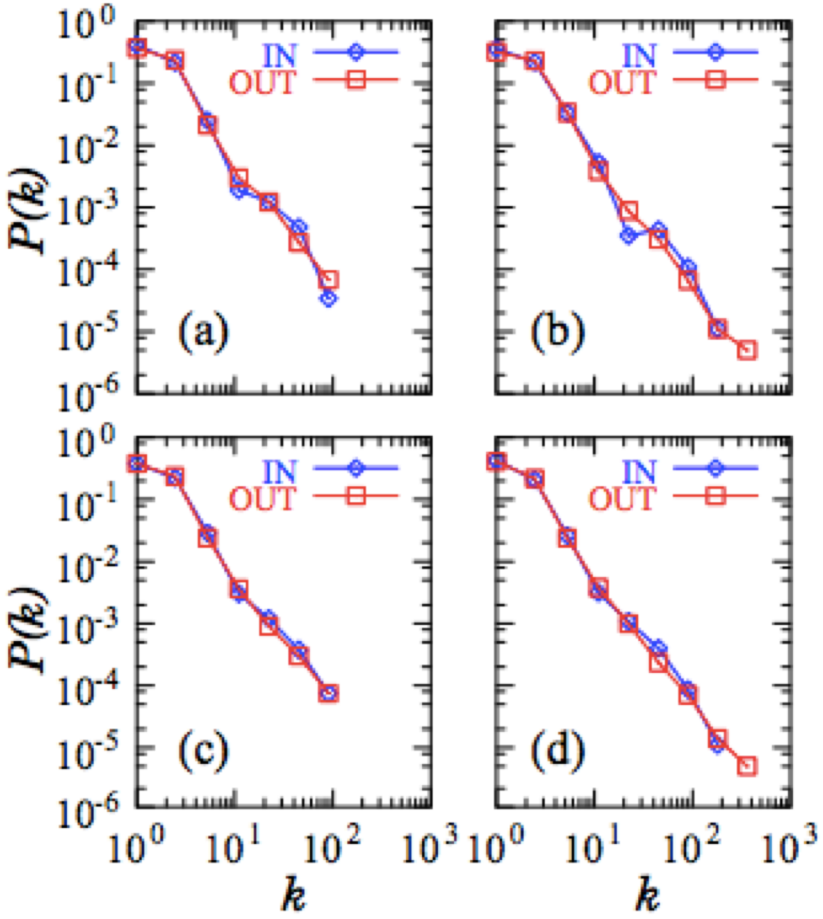
Pyruvate



https://en.wikipedia.org/wiki/Power_law

An example power-law graph, being used to demonstrate ranking of popularity. To the right is the long tail, and to the left are the few that dominate (also known as the 80–20 rule).

Let's examine metabolic networks for a few organisms



Connectivity distribution P(k) for the substrates in
(a) A. fulgidus (Archae)
(b) E. coli (Bacterium)
(c) C. elegans (Eukaryote),

shown on a log-log plot, counting separately the incoming (IN) and outgoing links (OUT) for each substrate, kin (kout).

(d) The connectivity distribution averaged over 43 organisms.

Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., & Barabási, A. L. (2000). The large-scale organization of metabolic networks. Nature, 407(6804), 651.

Observation for metabolic networks

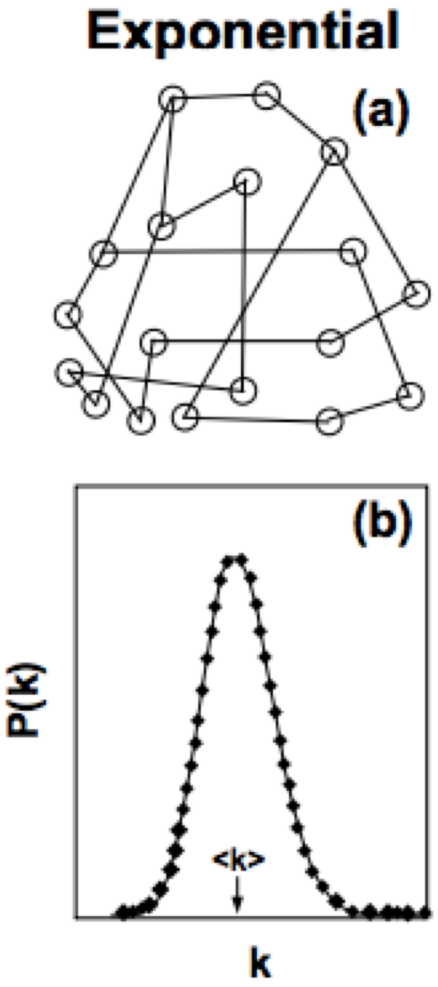
- Assume P(k) is the probability that a substrate can react with k other substrates
- The degree distribution $P(k)$ of a metabolic network decays as a power law $P(k) \sim k^{-\gamma}$ with $\gamma \sim = 2.2$ in all organisms

  ➔ Network topology is "scale-free"

- Many other networks are scale-free
  - Web links
  - Social networks
  - Paper citations

Contrast scale-free networks with random networks

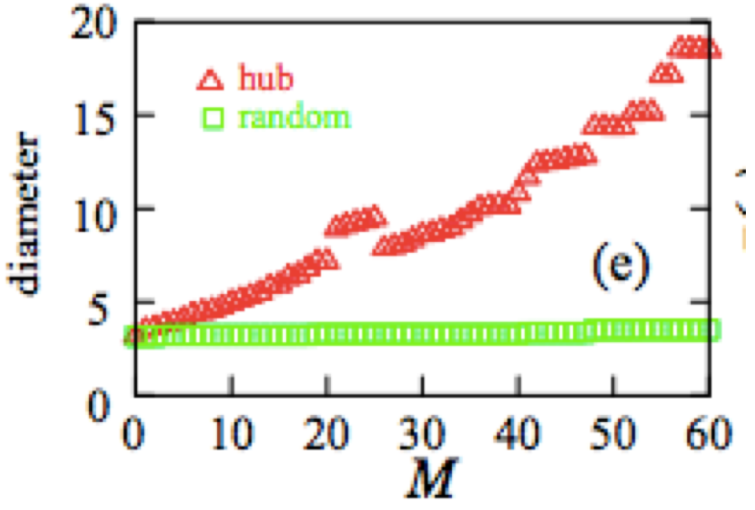
- Contrast with classical random network theory introduced by Erdös and Rényi (ER):
  - assumes that each pair of nodes in the network is connected randomly with probability p
  - despite the fundamental randomness of the model, most nodes have the same number of links, ⟨k⟩
  - Connectivity follows a Poisson distribution strongly peaked at ⟨k⟩ implying that the probability to find a highly connected node decays exponentially (i.e. $P(k) \sim e^{-k}$ for $k \gg \langle k \rangle$



Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., & Barabási, A. L. (2000). The large-scale organization of metabolic networks. Nature, 407(6804), 651.

Features of networks represented using the power law – small world characteristic

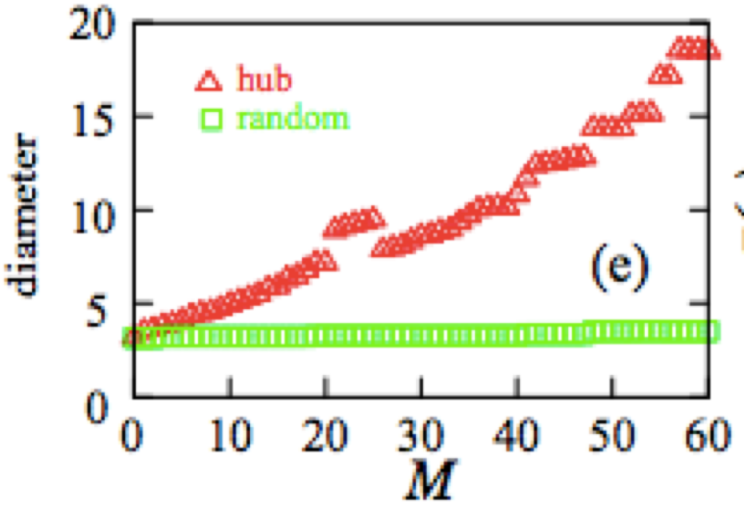
- Definition:
  - network diameter: shortest path (biochemical pathway) averaged over all pairs of nodes
- What happens to network diameter when removing nodes from the network:
  - Randomly – a random node is removed; recalculate diameter
  - Hub – remove most connected nodes first; recalculate diameter
- Show for *E. coli*
  - *successively remove top 60 nodes, one at a time*



Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., & Barabási, A. L. (2000). The large-scale organization of metabolic networks. Nature, 407(6804), 651.

Features of networks represented using the power law – small world characteristic

- Implications of prior experiment:
  - Removing hubs increases network diameter

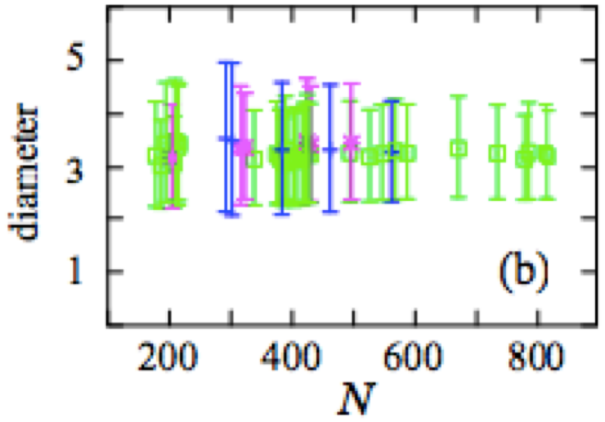


- Small-world characteristics
  - any two nodes in the system can be connected by relatively short paths along existing links

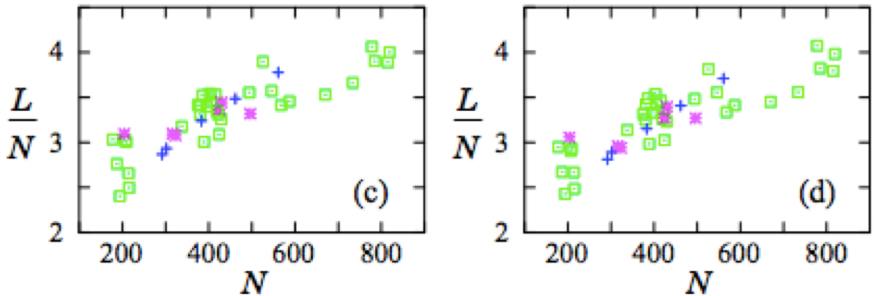
Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., & Barabási, A. L. (2000). The large-scale organization of metabolic networks. Nature, 407(6804), 651.

Features of networks represented using the power law – small world characteristic

- Finding: diameter of the metabolic network is the same for 43 organisms, irrespective of the number of substrates found in the given species
  - N in the graph is the number of nodes
  - Error bars show standard deviation
  - Counter intuitive: expect diameter to increase with increasing network size
  - Archaea (magenta), bacteria (green), and eukaryotes (blue)



- Explained by:
  - The average number of reactions in which a certain substrate participates increases with the number of substrates found within the given organism
    - Shown: the average number of incoming links (c) or outgoing links (d) per node for each organism.



Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., & Barabási, A. L. (2000). The large-scale organization of metabolic networks. Nature, 407(6804), 651.

Which metabolites are hubs?

No.	Name	N	L(IN)	L(OUT)	R	E	γ_{in}	γ_{out}	D	Hub(IN)	Hub(OUT)
1	A. pernix	204	588	575	178	135	2.2	2.2	3.2	bacdelgfij	adbcegipfh
2	A. fulgidus	496	1527	1484	486	299	2.2	2.2	3.5	abcdghefjk	adbijchemf
3	M. thermoautotro phicum	430	1374	1331	428	280	2.2	2.2	3.4	abcdgefkhi	adbicejfkh
4	M. jannaschii	424	1317	1272	415	264	2.2	2.3	3.5	abcdgekfnh	adbciejkhf
5	P. furiosus	316	901	867	283	191	2.0	2.3	3.4	abcdgeknfh	dabceipjhf
6	P. horikoshii	323	914	882	288	196	2.0	2.2	3.4	abdcgefknl	dabciepjhq
7	A. aeolicus	419	1278	1249	401	285	2.1	2.2	3.3	bcadgefhki	adbciejgfh
8	C. pneumoniae	194	401	391	134	84	2.2	2.3	3.4	bdcagfleri	dabciergfp
9	C. trachomatis	215	479	462	158	94	2.2	2.4	3.5	bdacgfelrm	dbaciegrfp
10	Synechocystis sp.	546	1782	1746	570	370	2.0	2.2	3.3	abcdegfhjk	adbicjehfg
11	P. gingivalis	424	1192	1156	374	254	2.2	2.2	3.3	abdcgefnkh	adbceipjhg
12	M. bovis	429	1247	1221	391	282	2.2	2.2	3.2	abdcgelfkm	adbceifhjq
13	M. leprae	422	1271	1244	402	282	2.2	2.2	3.2	abcdgefkml	adbceifjhq
14	M. tuberculosis	587	1862	1823	589	358	2.0	2.2	3.3	adbcghemjk	adbjhmceit
15	B. subtilis	785	2794	2741	916	516	2.2	2.1	3.3	abdcjhmegf	adhbjcimef
16	E. faecalis	386	1244	1218	382	281	2.1	2.2	3.1	bdacgelfik	adbciefghj
17	C. acetobutylicum	494	1624	1578	511	344	2.1	2.2	3.3	abcdgefhlk	adbceijhfo
18	M. genitalium	209	535	525	196	85	2.4	2.2	3.5	bdcgzxuyos	adbcguvwos
19	M. pneumoniae	178	470	466	154	88	2.3	2.2	3.2	bcdgxoyasl	dabcgowvsr
20	S. pneumoniae	416	1331	1298	412	288	2.1	2.2	3.2	abdcgelfno	adbceifghj
21	S. pyogenes	403	1300	1277	404	280	2.1	2.2	3.1	abdcegfoln	adbceifohg
22	C. tepidum	389	1097	1062	333	231	2.1	2.2	3.3	badcgenfki	dabceipgfq
23	R. capsulatus	670	2174	2122	711	427	2.1	2.2	3.4	abcdhgefjk	adbjhicmet
24	R. prowazekii	214	510	504	155	100	2.3	2.3	3.4	bdacfegilm	dabicfemgt
25	N. gonorrhoeae	406	1298	1270	413	285	2.1	2.2	3.2	abdcgefkjh	adbiechfjg
26	N. meningitidis	381	1212	1181	380	271	2.2	2.2	3.2	abdcegfkli	adbecifhjg
27	C. jejuni	380	1142								
28	H. pylori	375	1181								

~4% of all substrates that are found in all 43 organisms are present in all species. These substrates represent the most highly connected substrates found in any individual organism, indicating the generic utilization of the same substrates by each species.

Table 1.

Summary of the characteristics of the 43 investigated organisms. For each organism we show the number of substrate (N), number of links (L), number of individual reactions or temporary substrate-enzyme complexes (R), number of enzymes (E), the exponent γ_{in} and γ_{out} and the diameter of the metabolic network (D). In the last two columns we list the ten substrates with the largest number of incoming (IN) and outgoing (OUT) links. The letters correspond to: a=H_2O, b=ADP, c=orthophosphate, d=ATP, e=L-glutamate, f=$NADP^+$, g=pyrophosphate, h=NAD^+, i=NADPH, j=NADH, k=CO_2, l=NH_4^+, m=CoA, n=AMP, o=pyruvate, p=L-glutamine, q=2-oxoglutarate, r='alpha'-D-glucose 1-phosphate, s=phospho`enol`pyruvate, t=acetyl-CoA, u=H^+, v=uridine, w=cytidine, x=UMP, y=CMP, z=glycerol, α=D-fructose 6-phosphate. The color code of the fields denotes the different domains of life such a magenta = Archae green = Bacterium sky blue =Eukaryote.

Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., & Barabási, A. L. (2000). The large-scale organization of metabolic networks. Nature, 407(6804), 651.
Supplementary material.

What about modularity in metabolic networks?

- We already saw that there are distinct functional and chemical units within networks
  - Provides specialized function
  - Sometimes repeated motifs. Examples:
    - Reaction modules from KEGG
    - Substrate cycles: a set of reactions that forms a loop and does not lead to a net production or consumption of the participating metabolites

- How to reconcile the facts that metabolic networks are modular AND scale-free?
  - Let's examine how they are BOTH!

Scale-free vs Modular

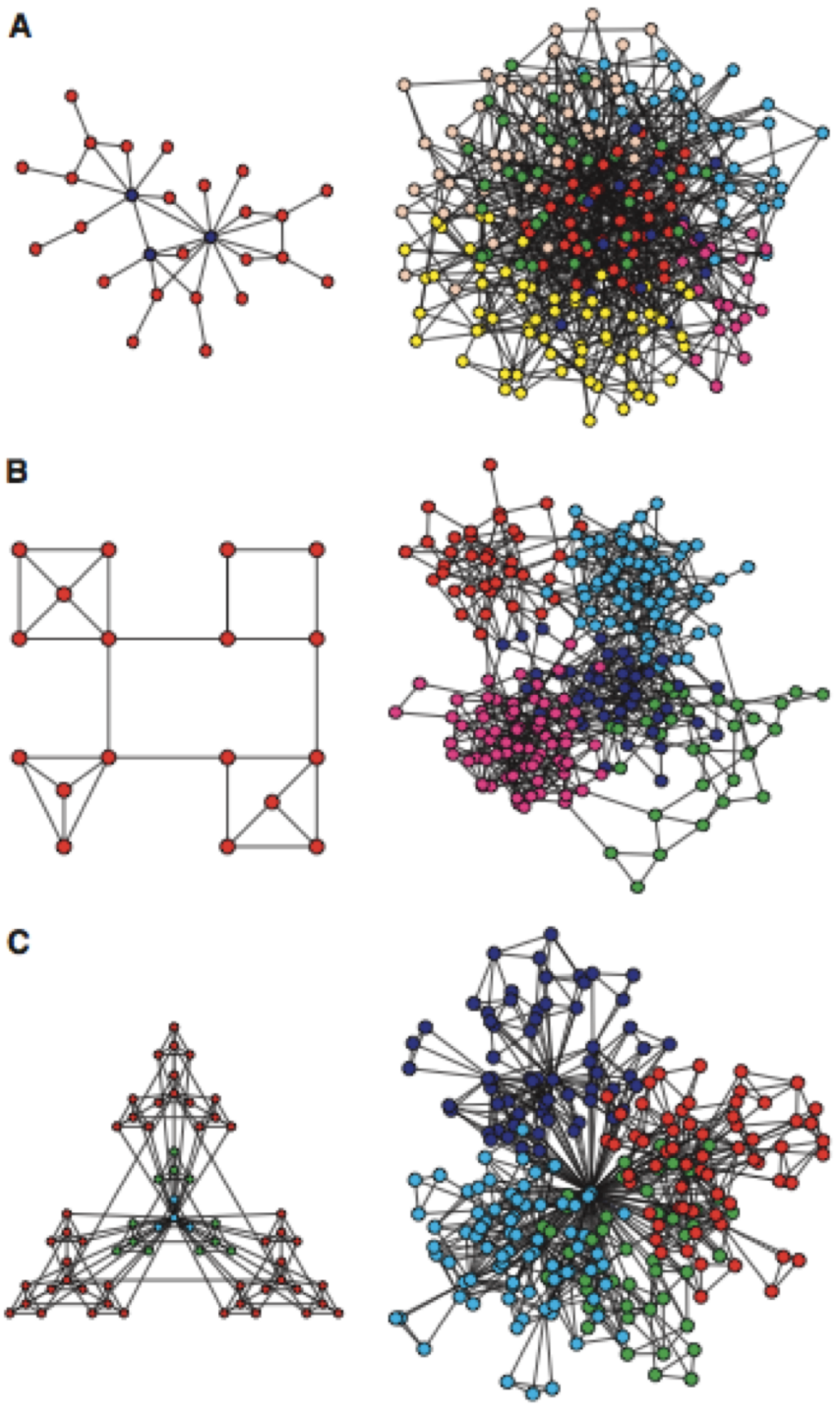
A – scale-free with hubs

The addition of a new node requires that existing nodes with higher degrees of connectivity have a higher chance of being linked to new nodes

B – modular distinct groupings

No hubs. A clustering algorithm groups nodes into different modules

C – Both modular AND scale free, **forming hierarchical modules**

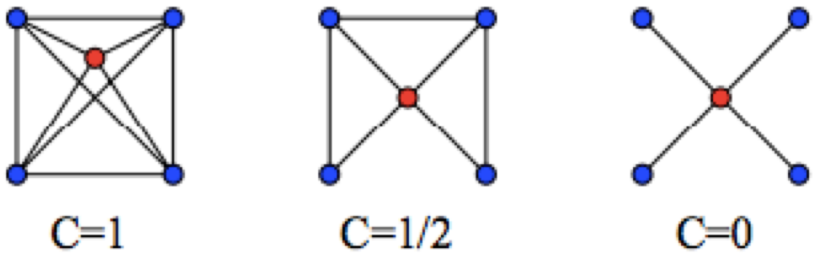


Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., & Barabási, A. L. (2002). Hierarchical organization of modularity in metabolic networks. *science*, 297(5586), 1551-1555.

Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., & Barabási, A. L. (2002). Hierarchical organization of modularity in metabolic networks. *science*, *297*(5586), 1551-1555.

Evidence of Modularity

- Clustering coefficient offers a measure of the degree of interconnectivity in the neighborhood of a node
  - a node whose neighbors are all connected to each other has C =1 (left), whereas a node with no links between its neighbors has C = 0 (right).



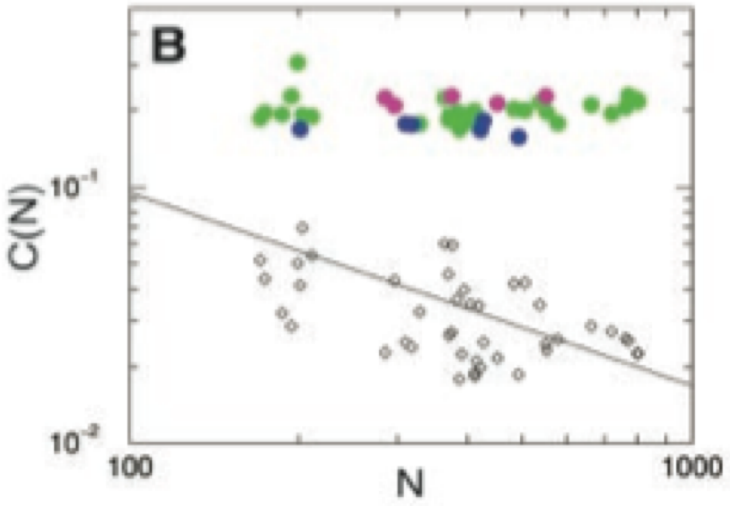
Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., & Barabási, A. L. (2002). Hierarchical organization of modularity in metabolic networks. *science*, *297*(5586), 1551-1555.

- $Ci = 2n/k_i(k_i - 1)$
  - n denotes the number of direct links connecting the k_i nearest neighbors of node
  - Left example: C_rednode = 2 * 6 / (4 * 3) = 1
  - Middle example: C_rednode = 2 * 3/ (4 *3) = ½
  - Right example: C_rednode = 2* 0/ (4 * 3) = 0

Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., & Barabási, A. L. (2002). Hierarchical organization of modularity in metabolic networks. *science*, *297*(5586), 1551-1555.

Clustering Coefficients

- Clustering coefficient averaged for all nodes of the network is a measure of the network's potential modularity
- Examine 43 organisms
  - Compare against "expected" for a scale-free network of similar size
    - N = number of nodes in the graph
    - Colors: archaea (purple), bacteria (green), and eukaryotes (blue)
    - diamonds denote *C* for a scale-free network with the same parameters (*N* and number of edges)
    - Line indicates indicates dependence of clustering coefficient on the network size for a module-free scale-free network

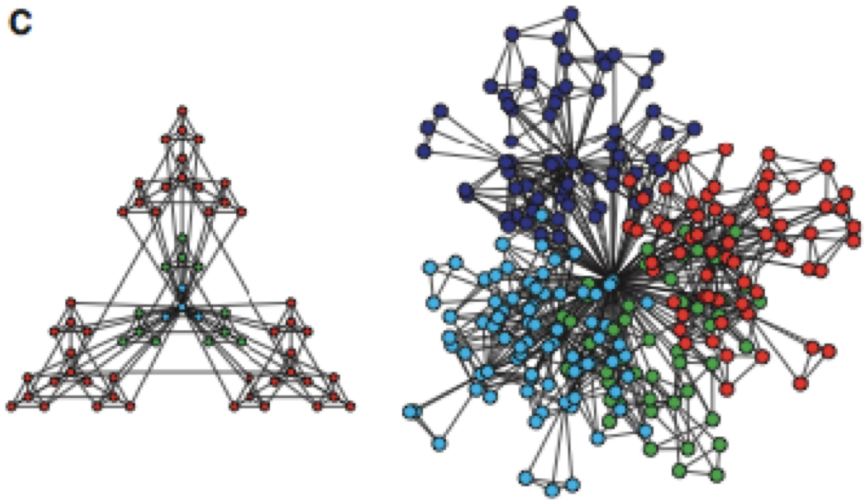


  - The average clustering coefficient is about an order of magnitude larger than that expected for a scale-free network of similar size

Clustering coefficient *vs* number of links in scale-free networks

- The nodes at the center of the:
  - numerous 4-node modules have a clustering coefficient C=3/4,
  - those at the center of a 16-node module have k = 13 and C = 2/13,
  - and those at the center of the 64-node modules have k = 40 and C = 2/40

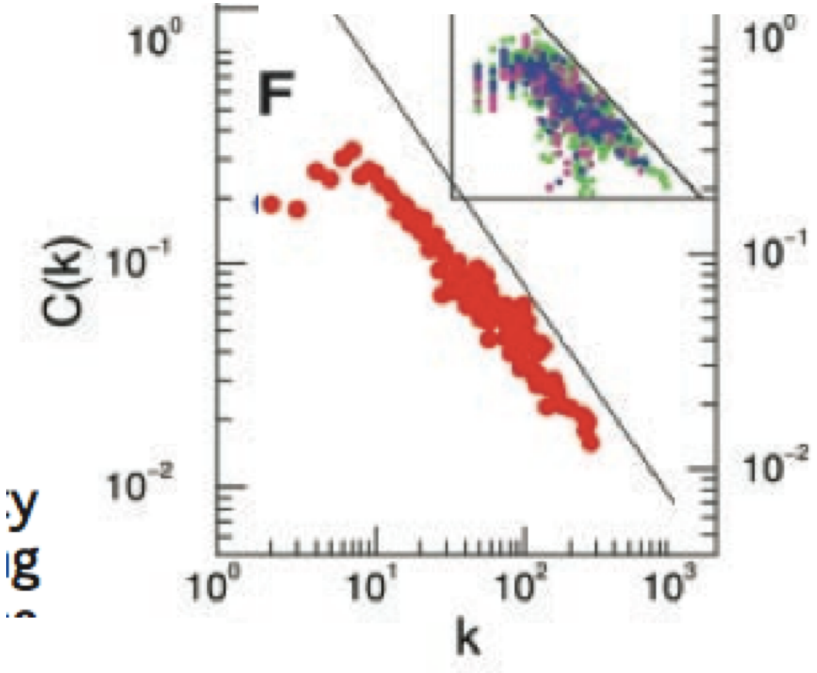
The higher the node's connectivity, the smaller its clustering coefficient



Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., & Barabási, A. L. (2002). Hierarchical organization of modularity in metabolic networks. *science*, *297*(5586), 1551-

Clustering coefficient *vs* number of links in modular scale-free networks

- Indeed, there is dependence of the clustering coefficient on the node's degree for 43 organisms
- The line correspond to $C(k) \sim k^{-1}$
- Inset shows individual organisms, while red dots shows average across all species



Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., & Barabási, A. L. (2002). Hierarchical organization of modularity in metabolic networks. *science*, *297*(5586), 1551-

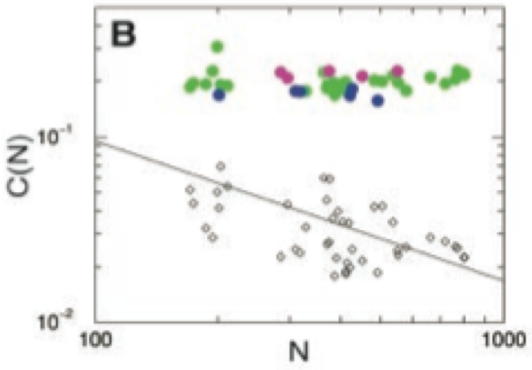
Metabolic networks are BOTH modular and Scale-Free

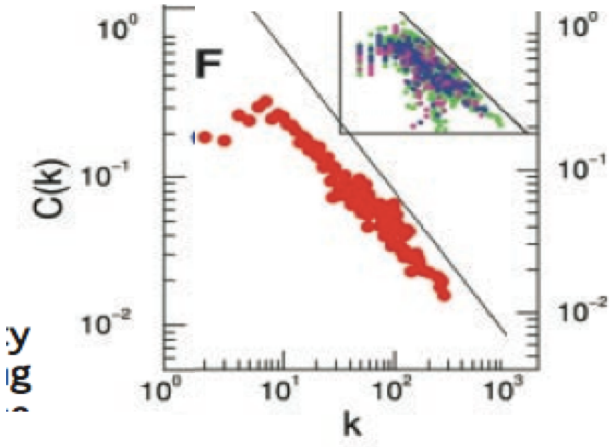
Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., & Barabási, A. L. (2002). Hierarchical organization of modularity in metabolic networks. *science*, *297*(5586), 1551-1555.

- Evidence for both being modular and for being scale-free

Modular: High and similar clustering coefficient for all N

Scale-free: Clustering coefficient of a node with k links follows the scaling law $C(k) = k^{-1}$

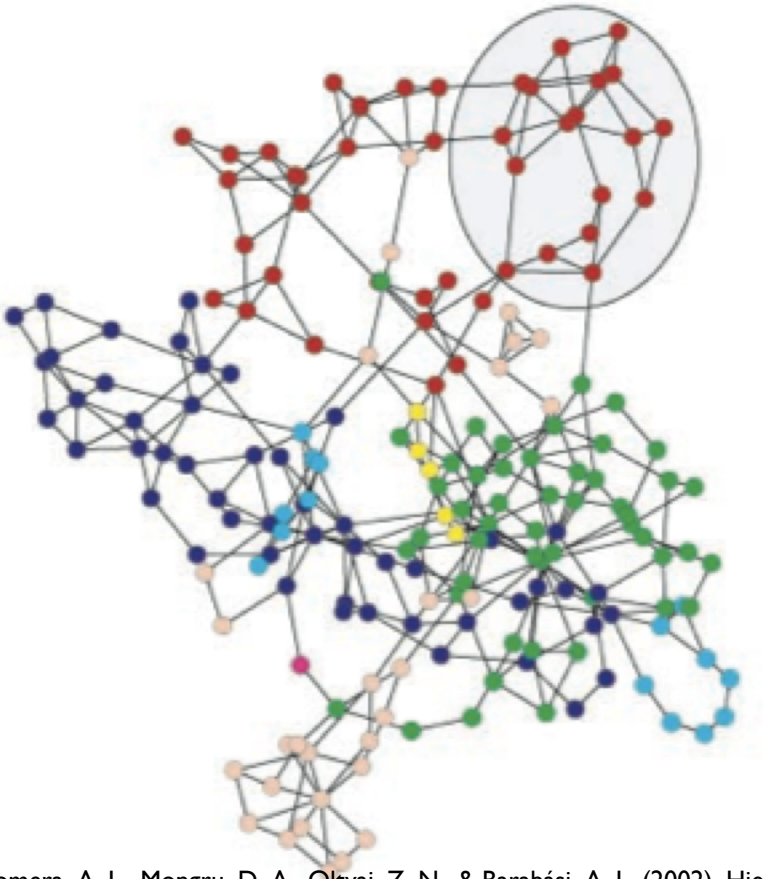




- Networks are organized into many small, highly connected topologic modules that combine in a hierarchical manner into larger, less cohesive units

A biological question: are identified hierarchical modules biologically meaningful?

- Examine *E. coli*
  - uncover that hierarchical modularity closely overlaps with known metabolic functions
  - using hierarchical clustering (grouping of similar elements)
  - carbohydrate metabolism (blue); nucleotide and nucleic acid metabolism (red); protein, peptide, and amino acid metabolism (green); lipid metabolism (cyan); aromatic compound metabolism (dark pink); monocarbon compound metabolism (yellow); and coenzyme metabolism (light orange)



Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., & Barabási, A. L. (2002). Hierarchica organization of modularity in metabolic networks. *science*, *297*(5586), 1551-1555.