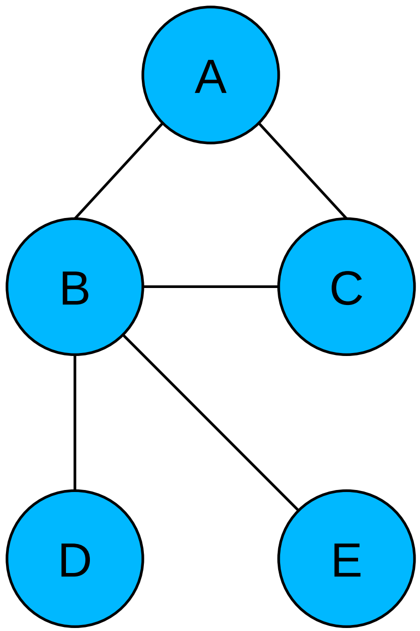
Introduction to Network Theory

Ernesto Estrada
Department of Mathematics & Statisitics
Department of Physics
University of Strathclyde
www.estradalab.org

Local metrics

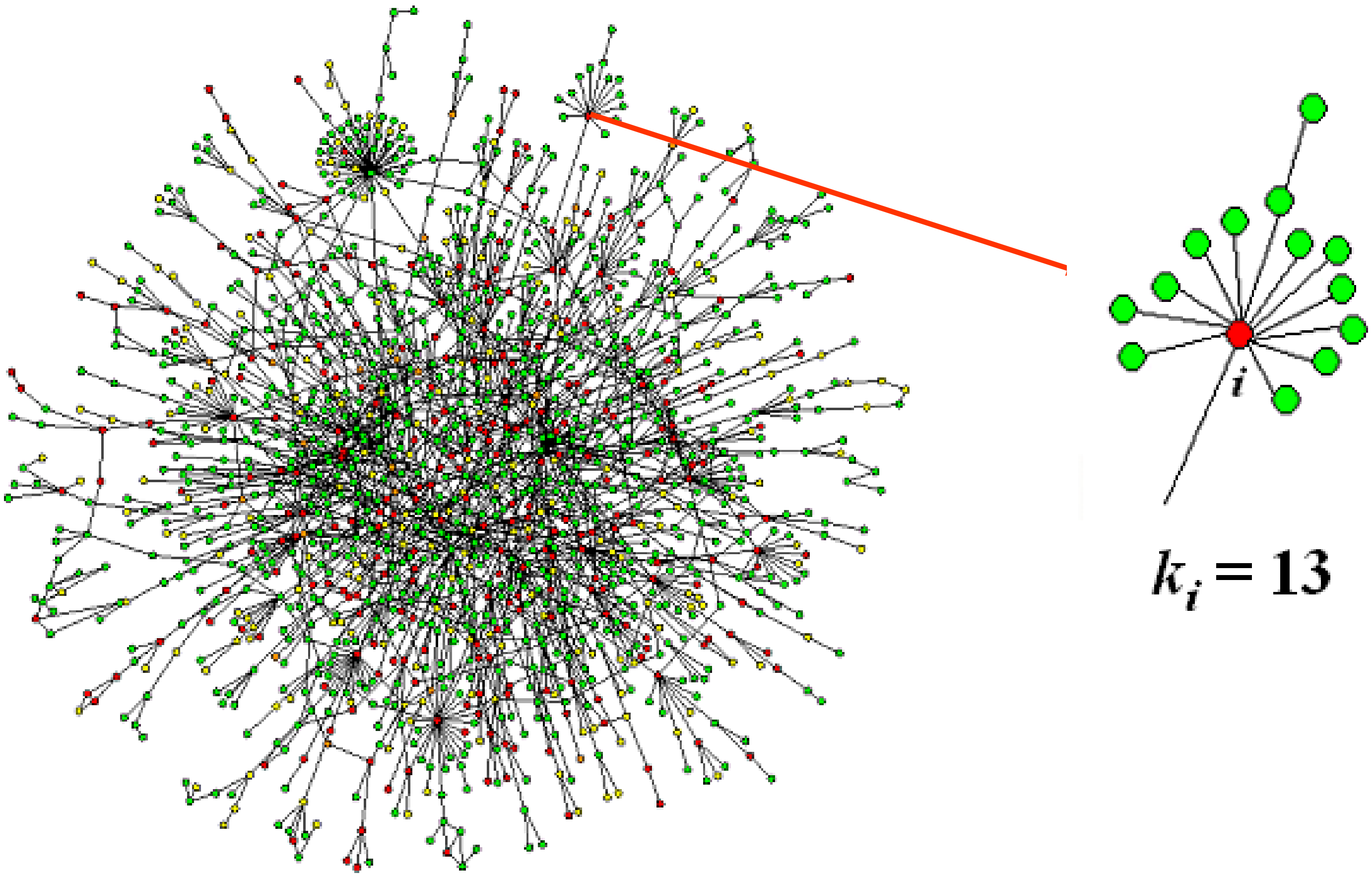
- Local metrics provide a measurement of a structural property of a single node
- Designed to characterise
  - Functional role – what part does this node play in system dynamics?
  - Structural importance – how important is this node to the structural characteristics of the system?

Degree Centrality



	A	B	C	D	E	degree
A	0	1	1	0	0	2
B	1	0	1	1	1	4
C	1	1	0	0	0	2
D	0	1	0	0	0	1
E	0	1	0	0	0	1

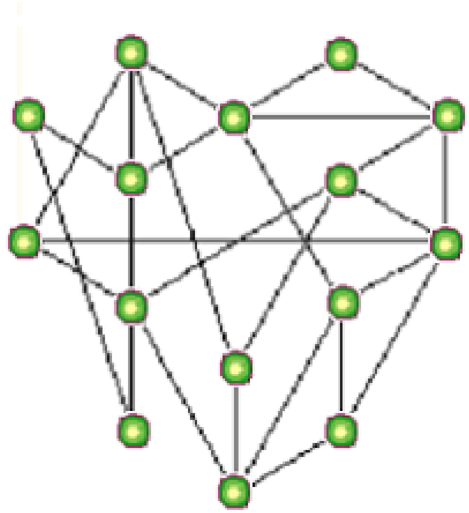
DEGREE DISTRIBUTION

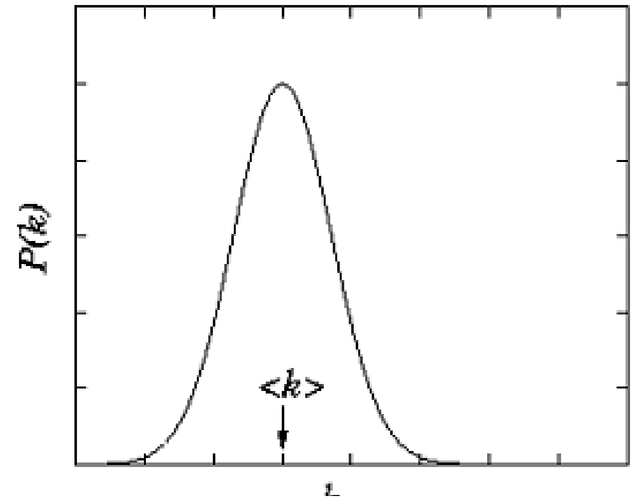


protein-protein interaction network (PIN)

DEGREE DISTRIBUTION

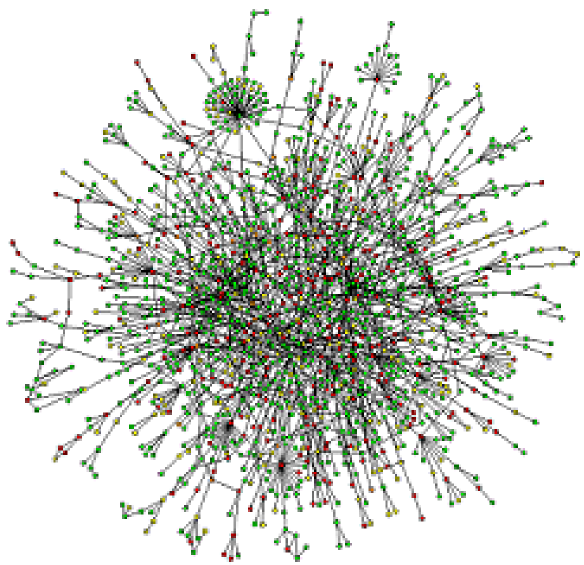
Random Network

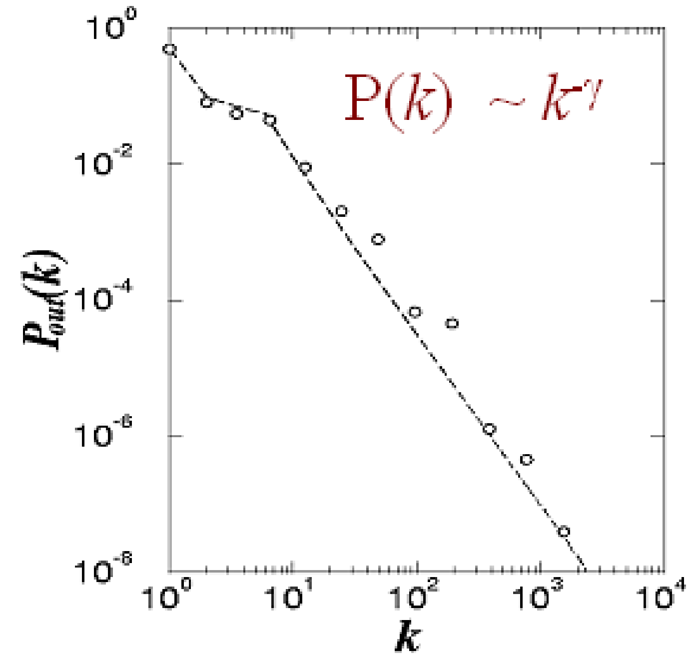




Expected

Real-World Network





Found

Albert, et al., *Nature*, 401, 1999, 130

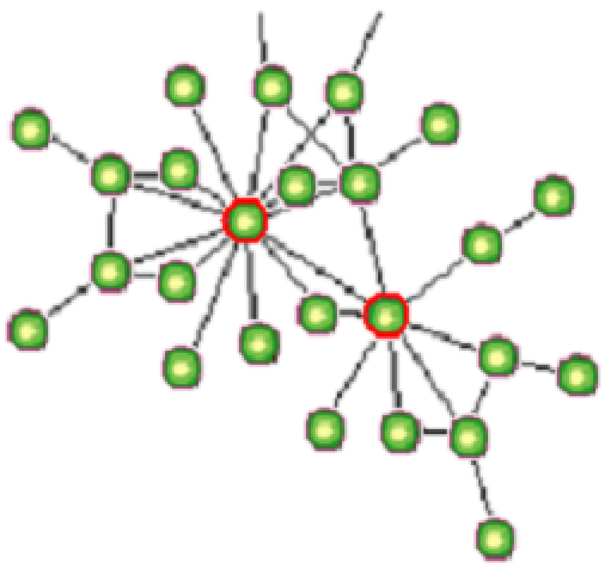
Scale-Freeness

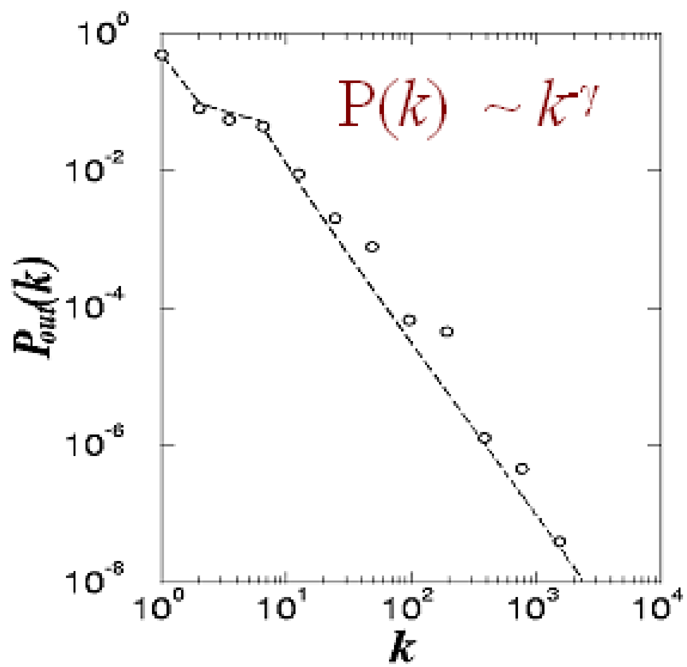
Power-law: $f(x) = ax^{\gamma}$ $\qquad \log f(x) = \gamma \log x + \log a$

Scale invariance: $\log f(cx) = \gamma \log x + \log ca$

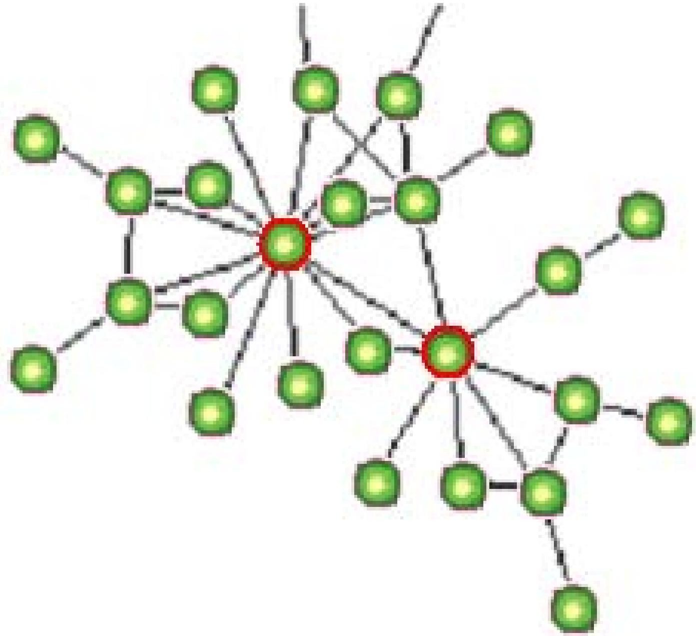
$$f(cx) \propto f(x)$$

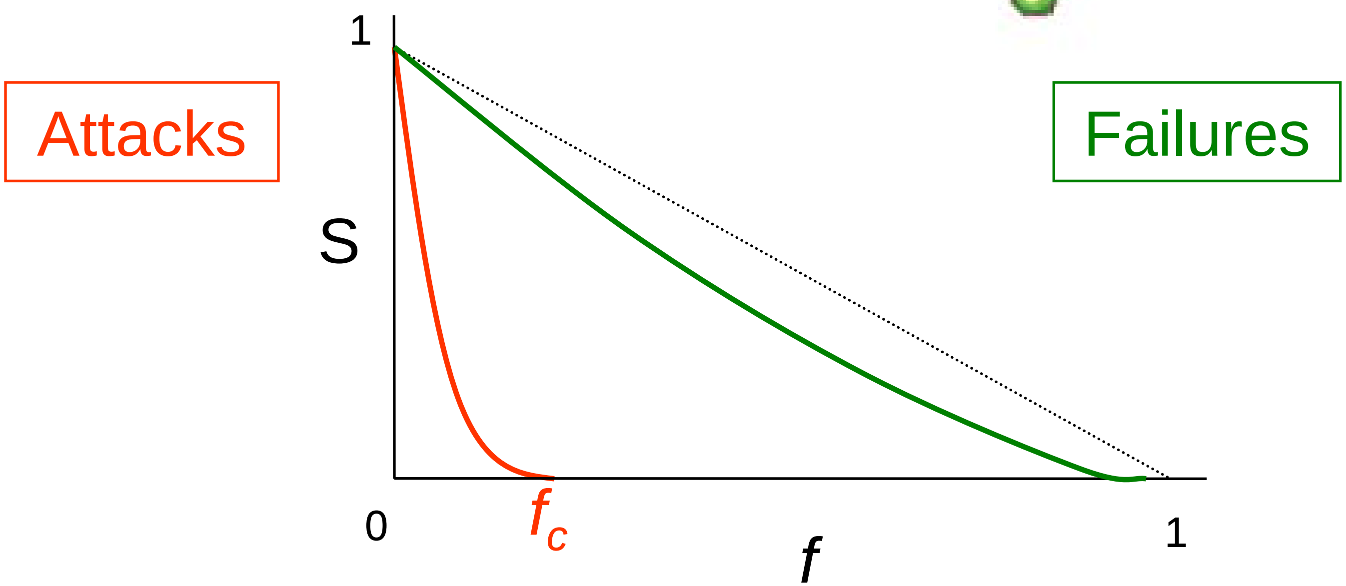
Scale-free Network





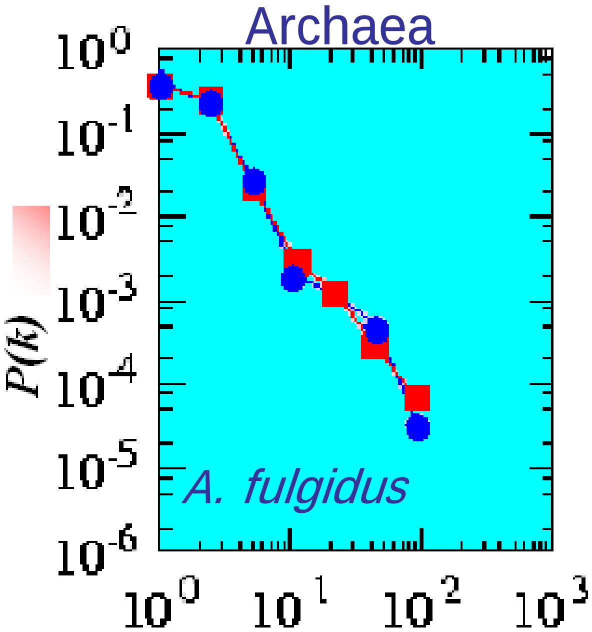
Consequences

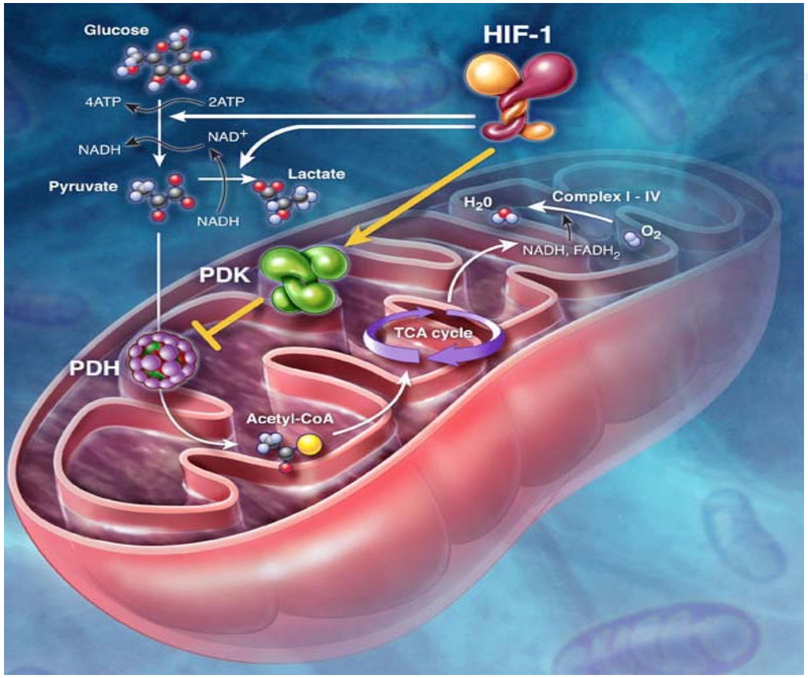


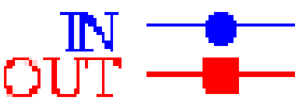


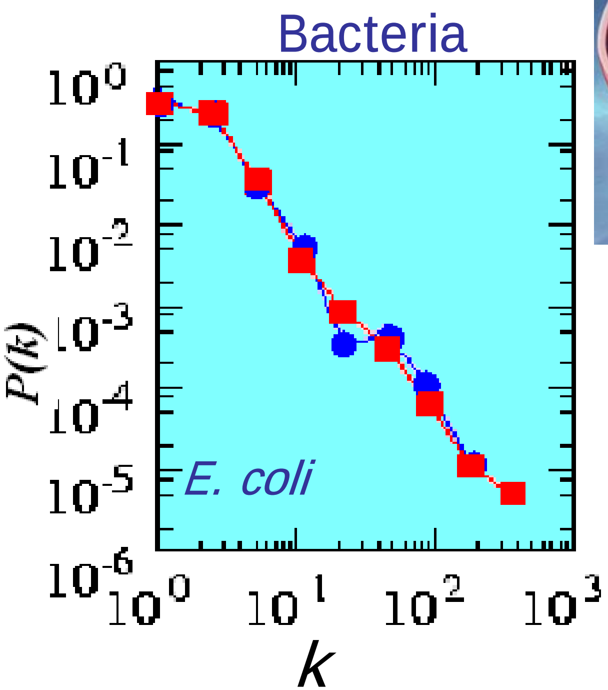
Albert, Jeong, Barábasi, *Nature* 406, 2000, 378

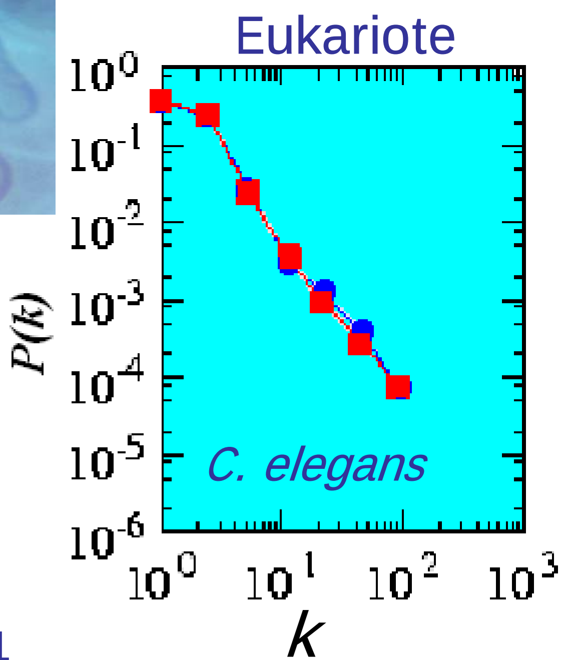
Metabolic Networks







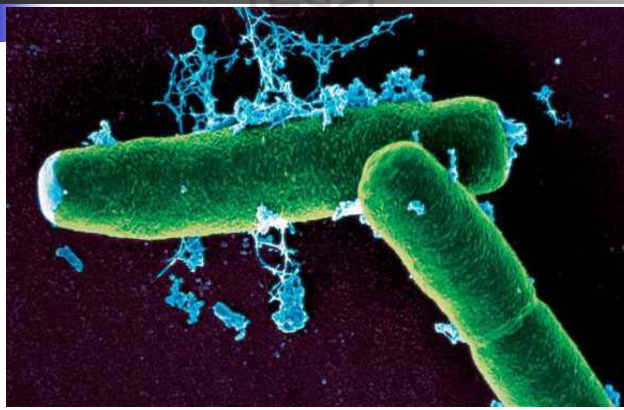




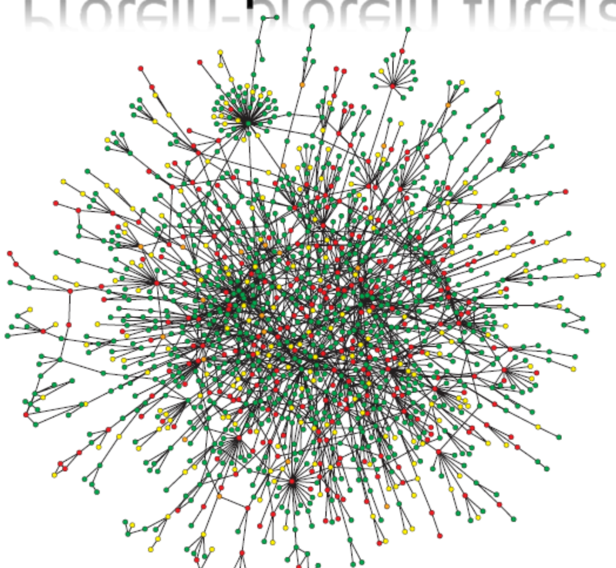
Jeong, et al., *Nature*, 407, 2000, 651

Protein Essentiality

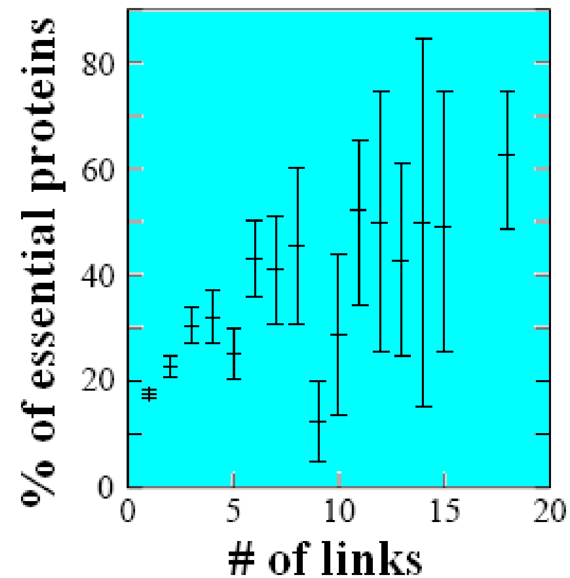
Yeast



Protein-protein Interactions



% Essential proteins vs. Protein Degree



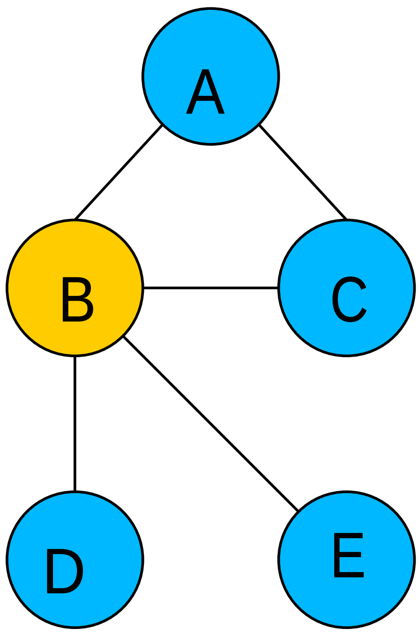
H. Jeong, et al., Nature 411, 41 (2001)

Betweenness centrality

- The number of shortest paths in the graph that pass through the node divided by the total number of shortest paths.

$$BC(k) = \sum_i \sum_j \frac{\rho(i,k,j)}{\rho(i,j)}, \quad i \neq j \neq k$$

Betweenness centrality



- Shortest paths are:
  - AB, AC, <u>ABD</u>, <u>ABE</u>, BC, BD, BE, <u>CBD</u>, <u>CBE</u>, <u>DBE</u>

$$\rho(A,B,D)=1; \quad \rho(A,D)=1$$
$$\rho(A,B,E)=1; \quad \rho(A,E)=1$$
$$\rho(C,B,D)=1; \quad \rho(B,D)=1$$
$$\rho(C,B,E)=1; \quad \rho(C,E)=1$$
$$\rho(D,B,E)=1; \quad \rho(D,E)=1$$

  - B has a BC of 5

Betweenness centrality

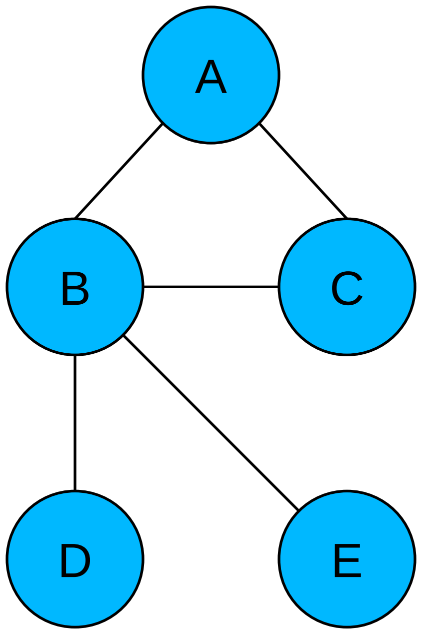
- Nodes with a high betweenness centrality are interesting because they
  - control information flow in a network
  - may be required to carry more information
- And therefore, such nodes
  - may be the subject of targeted attack

Closeness centrality

- The normalised inverse of the sum of topological distances in the network.

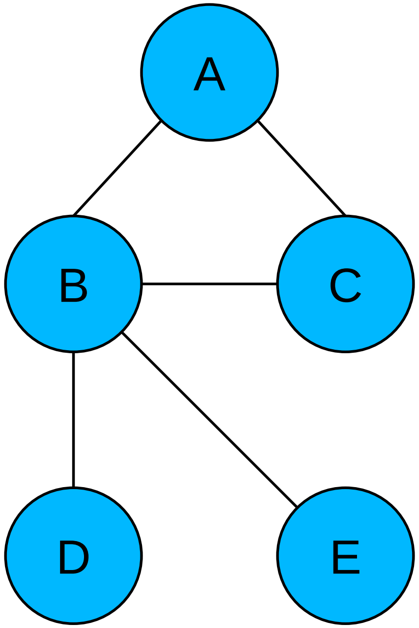
$$CC(i) = \frac{N-1}{\sum_j d(i,j)}$$

Closeness centrality



	A	B	C	D	E	$\sum_{j=1}^{n} d(i,j)$
A	0	1	1	2	2	6
B	1	0	1	1	1	4
C	1	1	0	2	2	6
D	2	1	2	0	2	7
E	2	1	2	2	0	7

Closeness centrality



Closeness

0.67

1.00

0.67

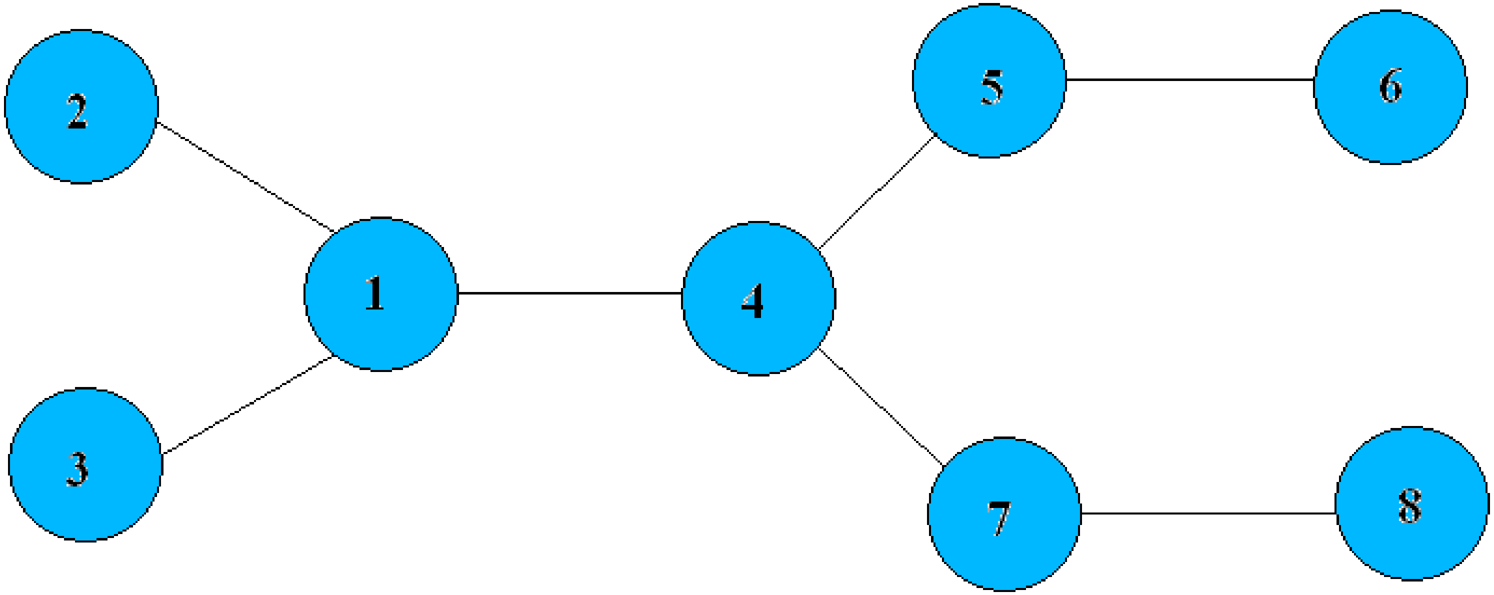
0.57

0.57

Closeness centrality

- Node B is the most central one in spreading information from it to the other nodes in the network.

Degree: Difficulties



Extending the Concept of Degree

Make x_i proportional to the average of the centralities of its i's network neighbors

$$x_i = \frac{1}{\lambda} \sum_{j=1}^{n} A_{ij} x_j$$

where λ is a constant. In matrix-vector notation we can write

$$\mathbf{x} = \frac{1}{\lambda} \mathbf{A}\mathbf{x}$$

In order to make the centralities non-negative we select the ***eigenvector*** corresponding to the ***principal eigenvalue*** (Perron-Frobenius theorem).

Eigenvalues and Eigenvectors

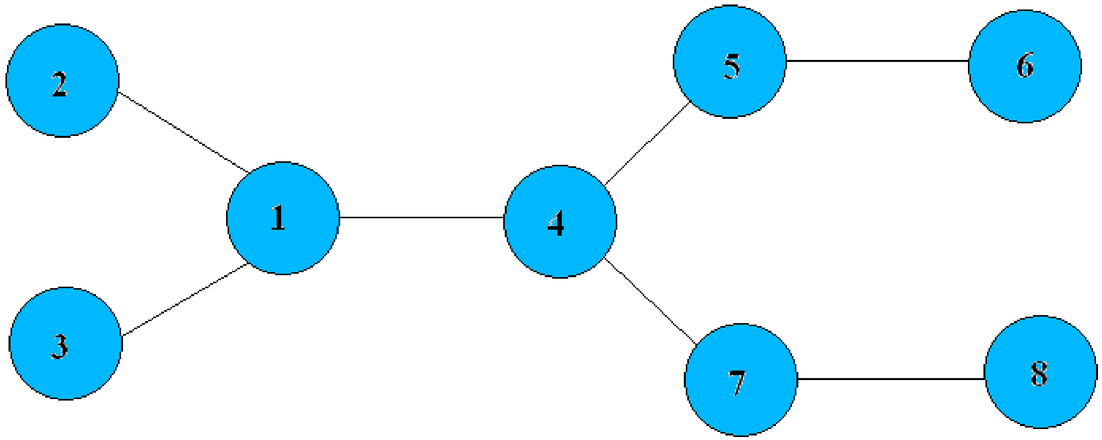
- The value λ is an eigenvalue of matrix A if there exists a non-zero vector x, such that $Ax=\lambda x$. Vector x is an eigenvector of matrix A
  - The largest eigenvalue is called the principal eigenvalue
  - The corresponding eigenvector is the principal eigenvector
  - Corresponds to the direction of maximum change

Eigenvector Centrality

- The corresponding entry of the principal eigenvector of the adjacency matrix of the network.

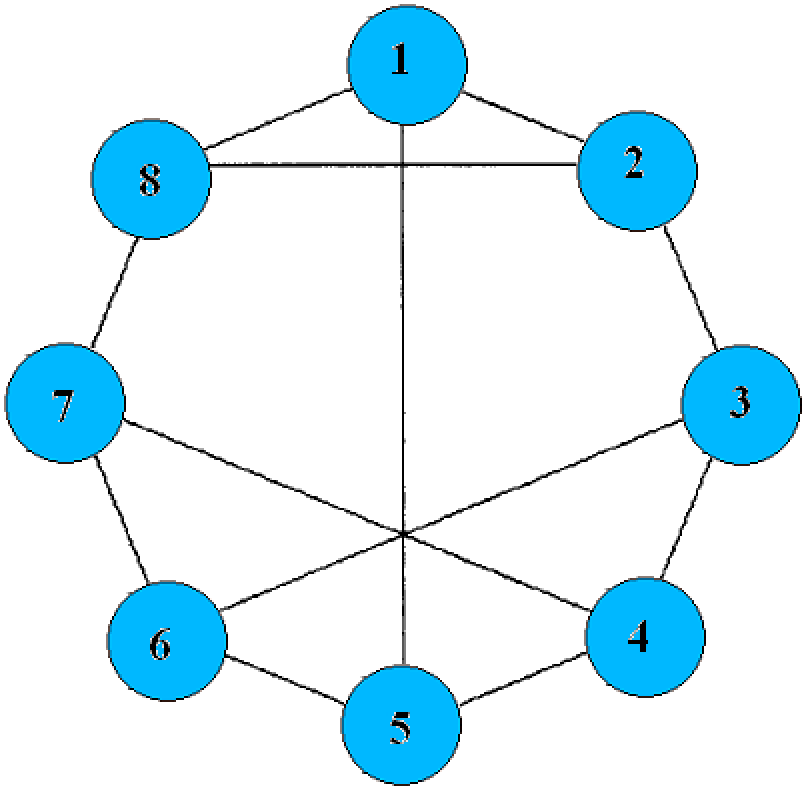
- It assigns relative scores to all nodes in the network based on the principle that connections to high-scoring nodes contribute more.

Eigenvector Centrality



Node	EC
1	0.500
2	0.238
3	0.238
4	0.575
5	0.354
6	0.354
7	0.168
8	0.168

Eigenvector Centrality: Difficulties



In ***regular networks*** all nodes have exactly the same value of the ***eigenvector centrality***, which is equal to $1/\sqrt{n}$

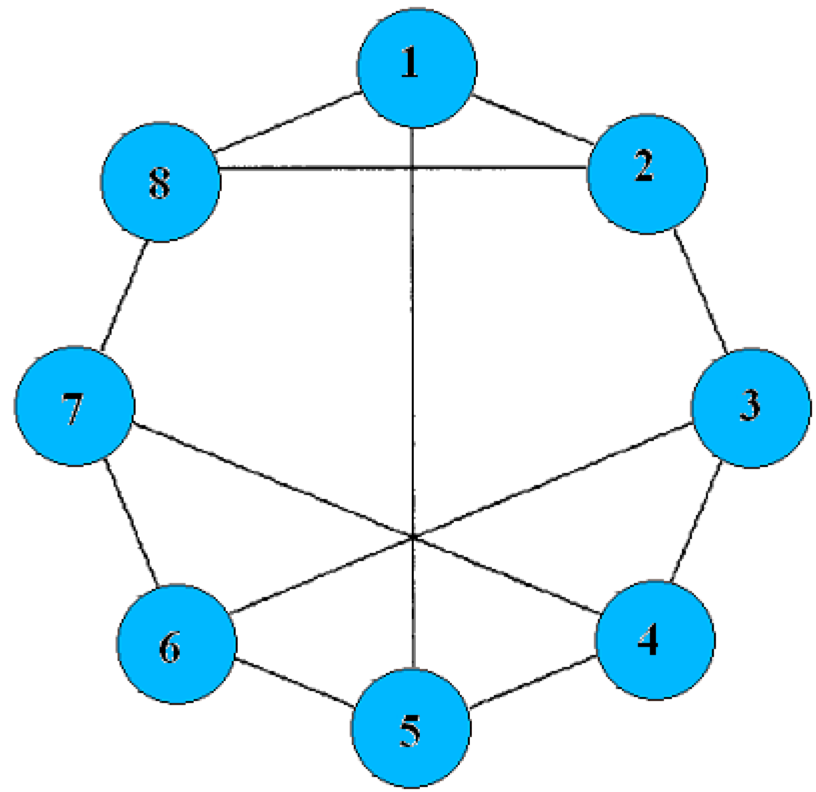
Subgraph Centrality

A ***closed walk of length k*** in a graph is a succession of k (not necessarily different) edges starting and ending at the same node, e.g.

1,2,8,1 (length 3)

4,5,6,7,4 (length 4)

2,8,7,6,3,2 (length 5)



Subgraph Centrality

The number of ***closed walk of length k*** starting at the same node i is given by the ii-entry of the kth power of the adjacency matrix

$$\mu_k(i) = \left(\mathbf{A}^k\right)_{ii}$$

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & \cdots \\ 1 & 0 & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

$$\mathbf{A}^2 = \begin{pmatrix} \mu_2(1) & & \\ & \mu_2(2) & \\ & & \ddots \end{pmatrix}$$

$$\mathbf{A}^3 = \begin{pmatrix} \mu_3(1) & & \\ & \mu_3(2) & \\ & & \ddots \end{pmatrix}$$

Subgraph Centrality

- We are interested in giving weights in decreasing order of the length of the ***closed walks***. Then, visiting the closest neighbors receive more weight that visiting very distant ones.

- The subgraph centrality is then defined as the following weighted sum

$$EE(i) = c_0 \left(A^0\right)_{ii} + c_1 \left(A\right)_{ii} + c_2 \left(A^2\right)_{ii} + c_3 \left(A^3\right)_{ii} + c_4 \left(A^4\right)_{ii} + \cdots$$

$$= \sum_{l=0}^{\infty} c_l \mu_l(i)$$

Subgraph Centrality

- By selecting $c_l=1/l!$ we obtain

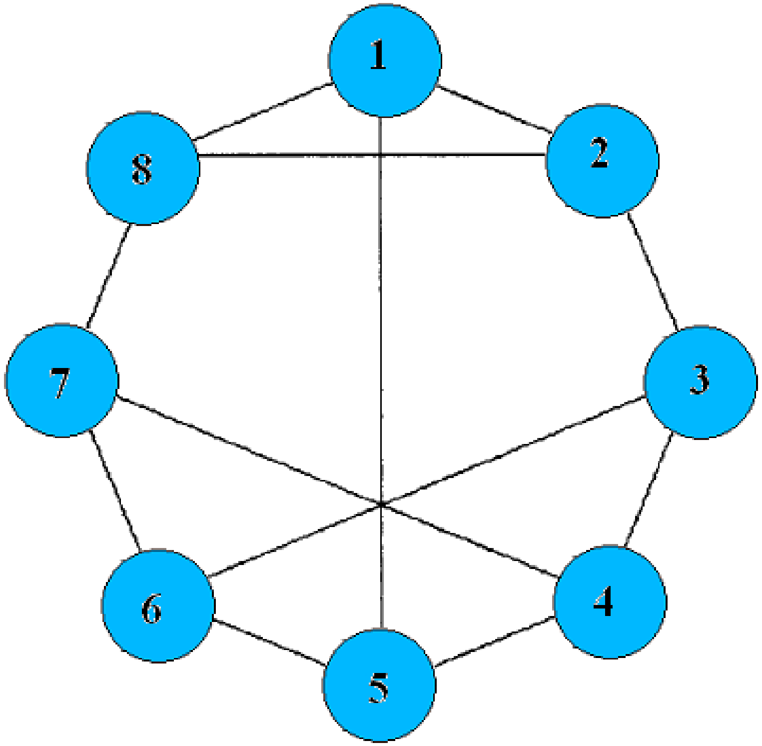
$$EE(i)=\sum_{l=0}^{\infty}\frac{\mu_l(i)}{l!} \qquad EE(i)=\left(e^{\mathbf{A}}\right)_{ii}$$

where e^A is the exponential of the adjacency matrix.

- For simple graphs we have

$$EE(i)=\sum_{j=1}^{n}\left[x_j(i)\right]^2 e^{\lambda_j}$$

Subgraph Centrality



$$DC(i) = DC(j), \forall i \in V, j \in V$$

$$CC(i) = CC(j), \forall i \in V, j \in V$$

$$EC(i) = EC(j), \forall i \in V, j \in V$$

Nodes	BC(i)
1,2,8	9.528
4,6	7.143
3,5,7	11.111

Nodes	EE(i)
1,2,8	3.902
4,6	3.705
3,5,7	3.638

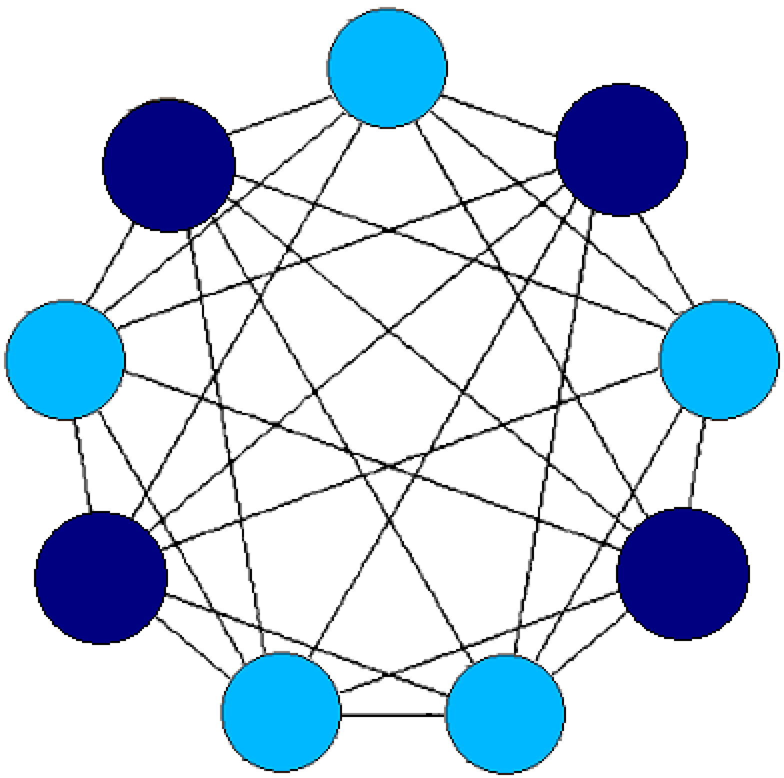
Subgraph Centrality

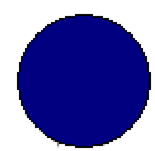
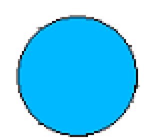
$$DC(i) = DC(j), \forall i \in V, j \in V$$

$$CC(i) = CC(j), \forall i \in V, j \in V$$

$$BC(i) = BC(j), \forall i \in V, j \in V$$

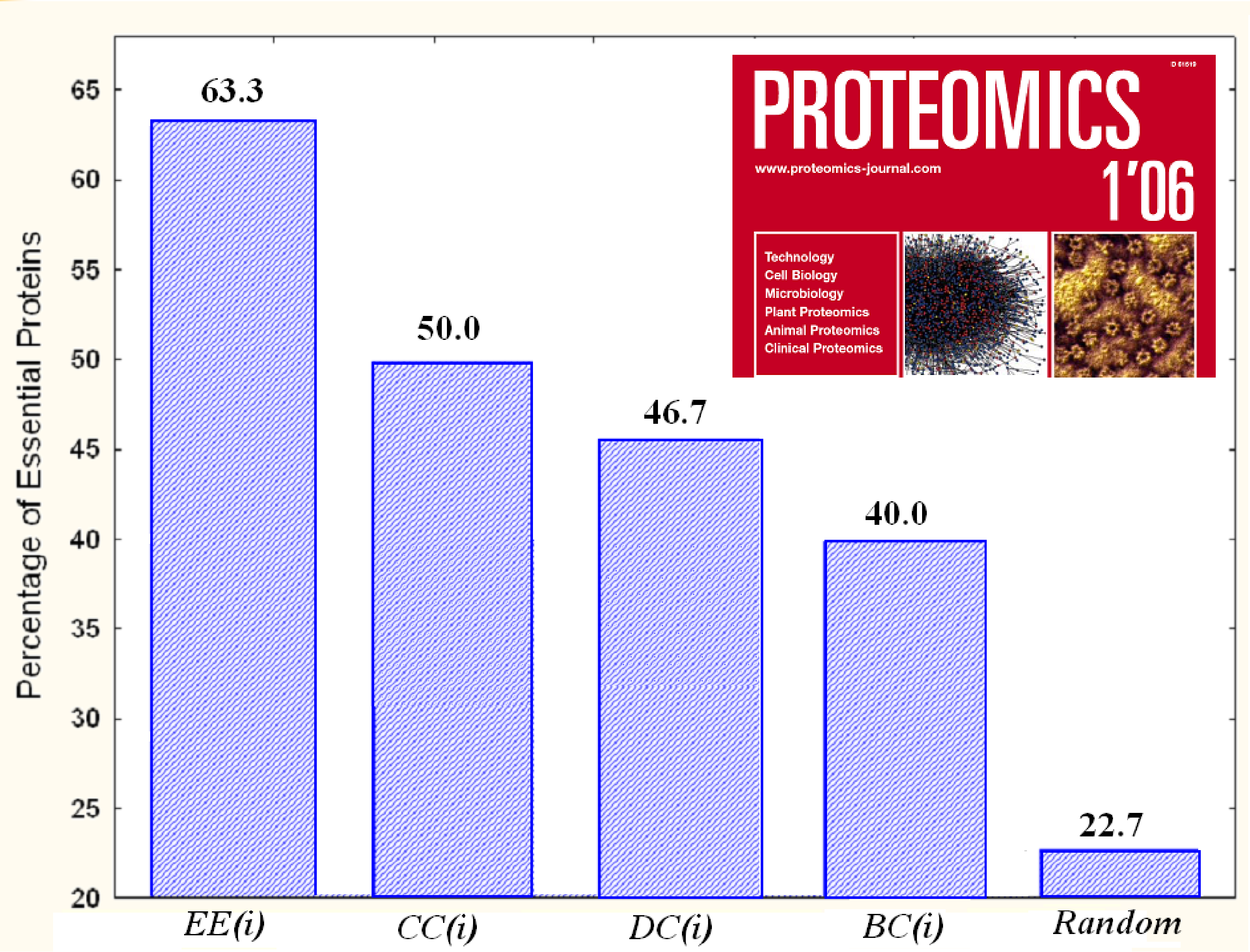
$$EC(i) = EC(j), \forall i \in V, j \in V$$



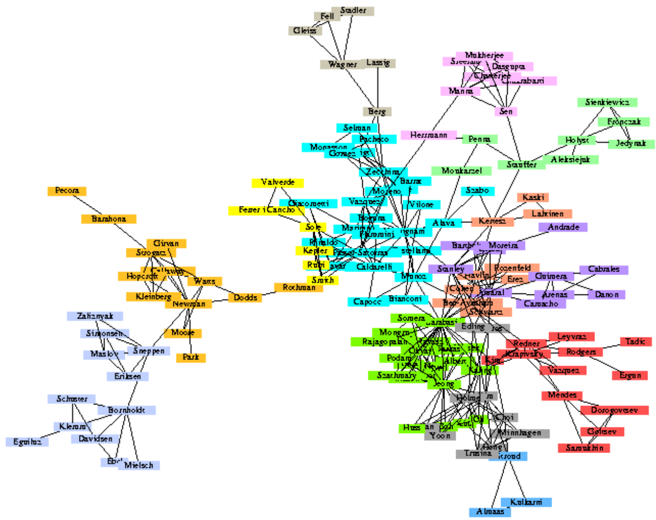
Nodes	EE(i)
	45.696
	45.651

Centrality and Protein Essentiality

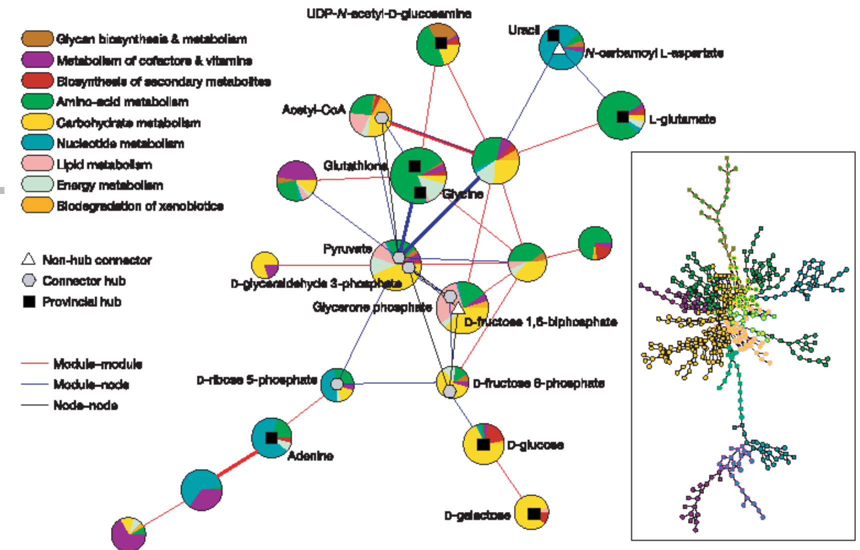
Estrada, E., Proteomics 6, 35 (2006)



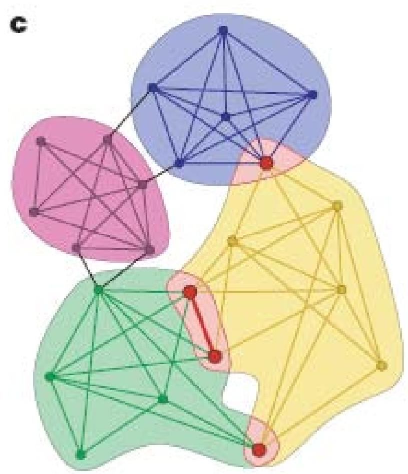
Communities



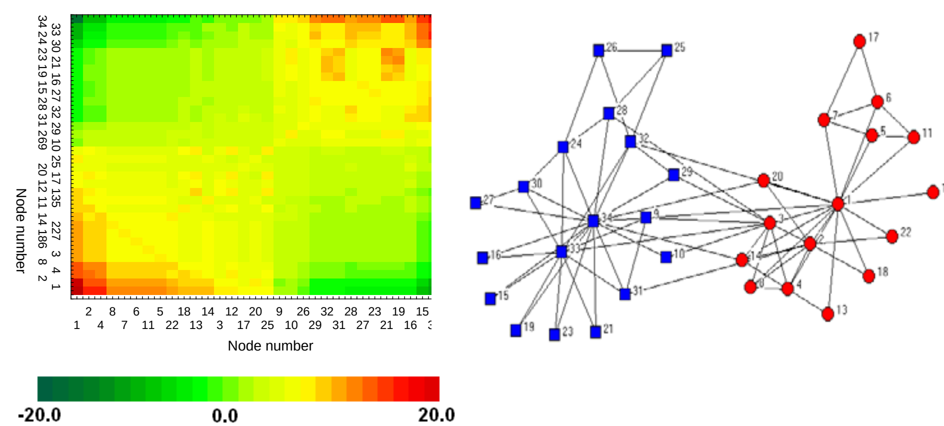
Girvan & Newman, *PNAS-USA* 99, 2002, 7821



Guimerá & Amaral, *Nature* 433, 2005, 895



Palla, et al., *Nature* 435, 2005, 814



Estrada & Hatano, Phys. Rev. E 77, 2008, 036111

References

- Aldous & Wilson, *Graphs and Applications. An Introductory Approach*, Springer, 2000.
- Wasserman & Faust, *Social Network Analysis*, Cambridge University Press, 2008.
- Estrada & Rodríguez-Velázquez, *Phys. Rev. E* **2005**, 71, 056103.
- Estrada & Hatano, *Phys. Rev. E*. **2008**, 77, 036111.

Exercise 1

Identify the most central node according to the following criteria:

(a) the largest chance of receiving information from closest neighbors;

(b) spreading information to the rest of nodes in the network;

(c) passing information from some nodes to others.