

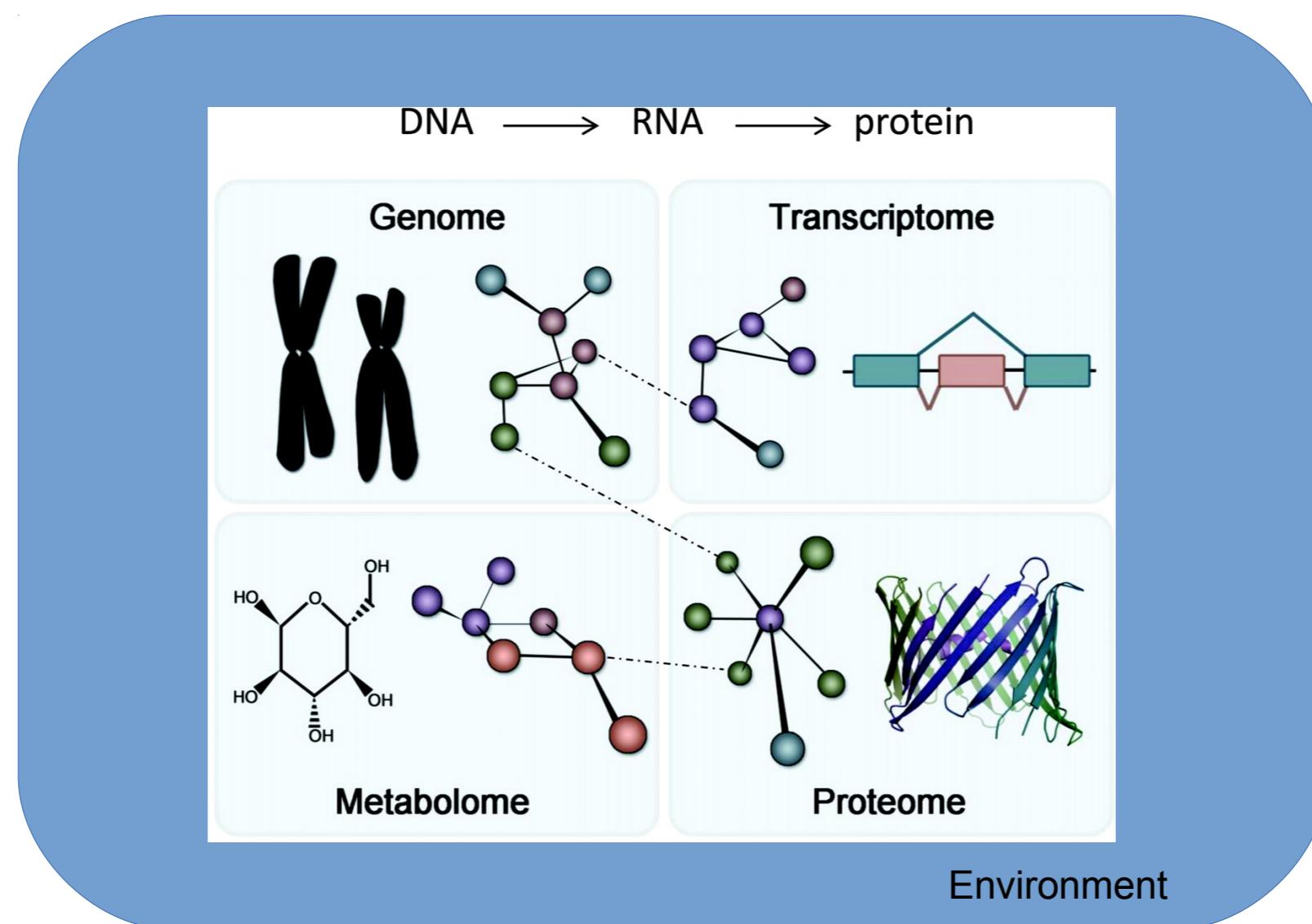
Bioinformatics 2 (INFR11005)

Lecture 5 Networks

Overview

- Introduction to Network Theory
- Statistical Properties of Networks
- Finding Structure
- Using Networks in Biology
- iPython Session

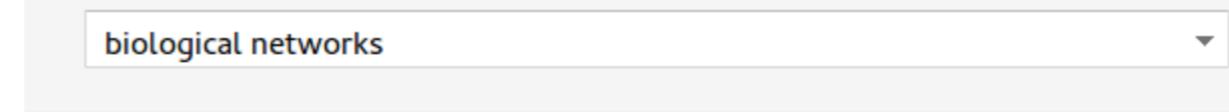
Focus on interactions



Franklin S , and Vondriska T M Circ Cardiovasc Genet 2011;4:576
<http://circgenetics.ahajournals.org/content/4/5/576.full>

Networks in biology

- Protein-protein interactions
 - Short interactions and modifications: signalling, e.g. phosphorylation
 - Long-lasting interactions: complex formation, transport, localisation
 - Modifications affect binding, so rewiring is possible
- Transcriptional regulation
 - Interactions between transcription factors and gene/transcription factor promotor regions
 - This network is continuously rewired!
- Metabolic networks
 - Biochemical reactions in cell to maintain function
 - Modulated by the proteome
- Neural network
- Species interactions
- Food networks
- ...



About 2,470,000 results (0.03 sec)

[Community structure in social and biological networks](#)

M Girvan, MEJ

Abstract A number of recent studies have focused on the statistical properties of complex systems such as social networks and the Worldwide Web. Researchers have particularly...
Cited by 4988

[Community structure in social and biological networks](#)

M Girvan, MEJ

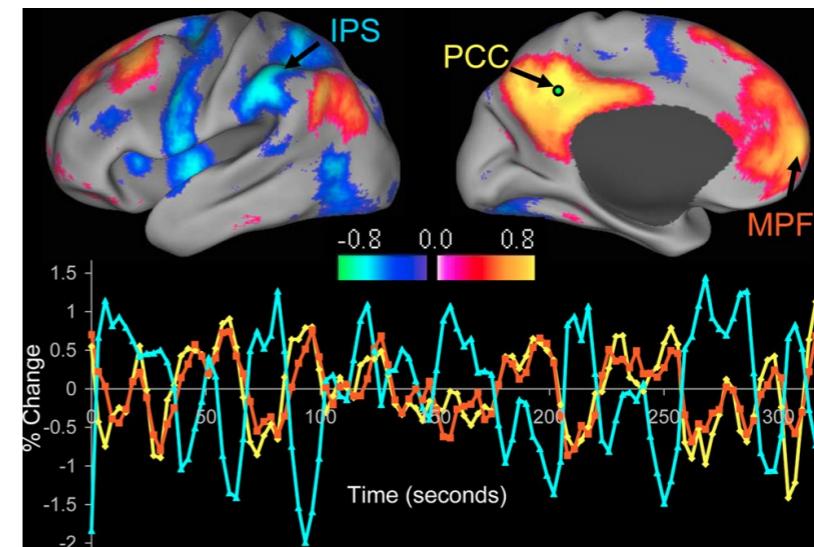
Abstract A number of recent studies have focused on the statistical properties of complex systems such as social networks and the Worldwide Web. Researchers have particularly...
Cited by 7832

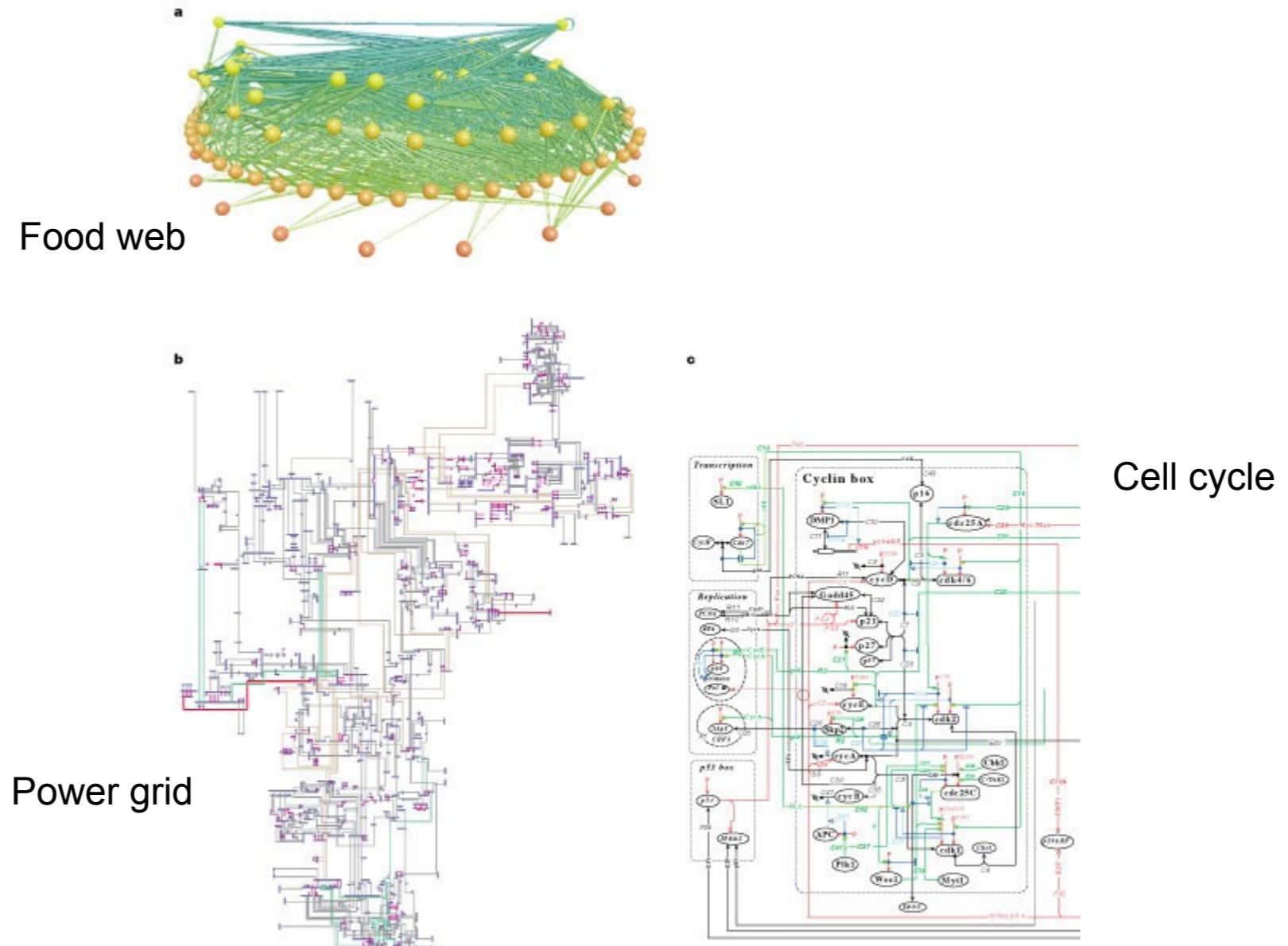
[Community structure in social and biological networks](#)

M Girvan, MEJ Newman - Proceedings of the national ..., 2002 - National ...
Abstract A number of recent studies have focused on the statistical properties of complex systems such as social networks and the Worldwide Web. Researchers have particularly...
Cited by 10683 Related articles All 52 versions

Networks in biology

- Physical networks
 - represent true known physical interactions, e.g. binding, eating, ...
- Functional networks
 - represent interactions found in data
 - may or may not be due to direct physical interaction
 - often inferred from time series data
 - correlation
 - Granger causality
 - transfer entropy
 - but correlation is not causation



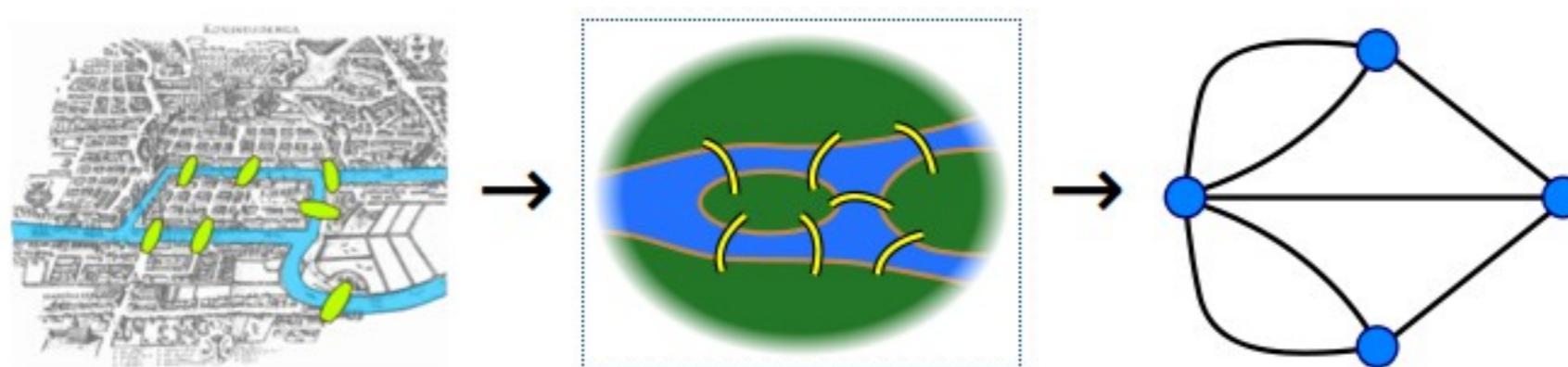


Strogatz, Nature, 2001

<http://www.nature.com/nature/journal/v410/n6825/full/410268a0.html>

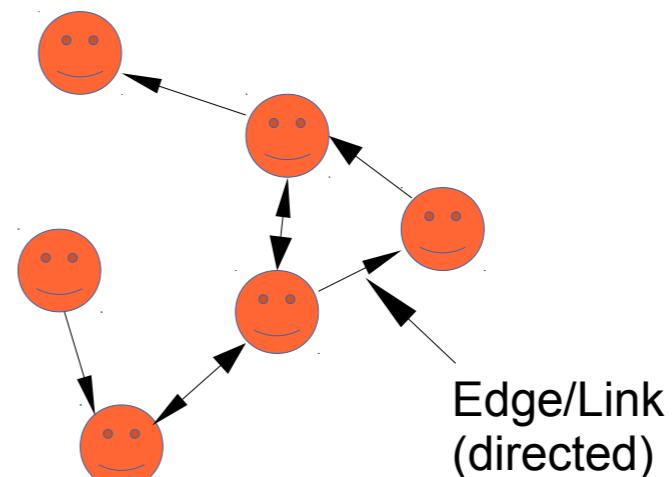
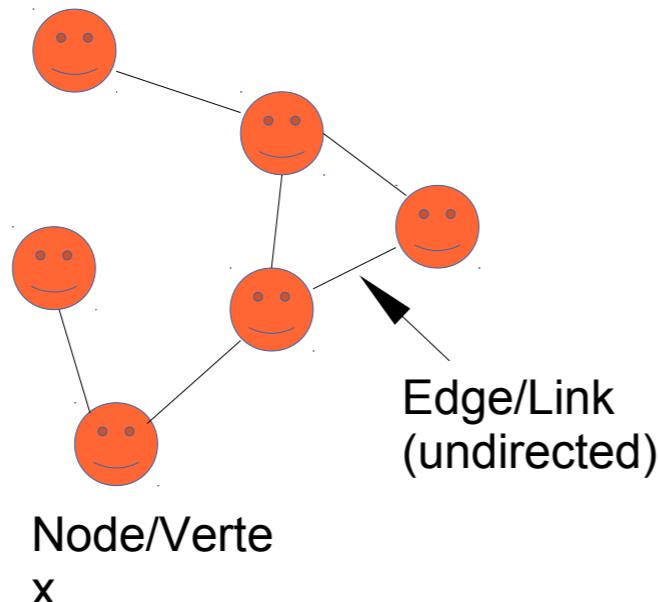
Graph Theory

- Introduced by Euler around 1735:
 - Kaliningrad (nee Koenigsberg) is on the river Pregolya (Pregel). There are four islands which could be reached by seven bridges.
 - Can he go to every place in the city and cross each bridge only once... i.e. go from island to island while traversing each bridge only once?

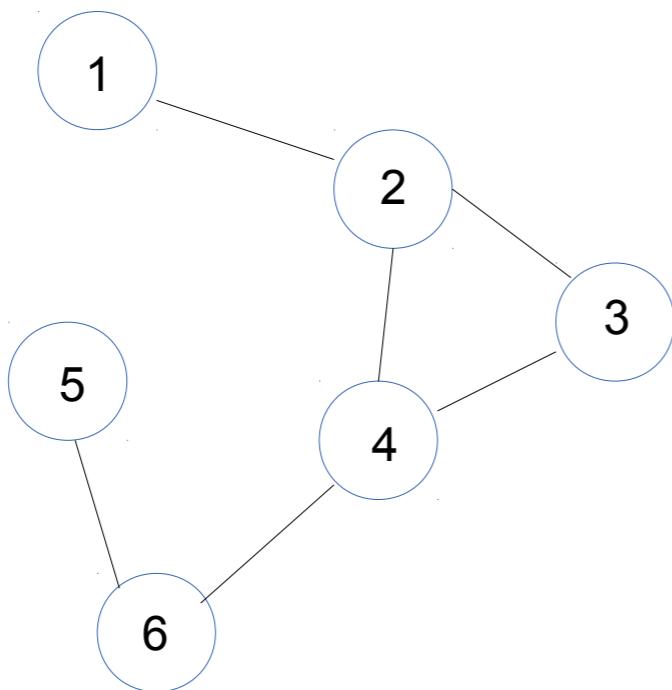


Graph Theory

- Networks are represented as a **graph** (N, g)
- **Nodes**: $N = \{ 1, \dots, n \}$
- **Edges**: $n \times n$ adjacency matrix $g = [g_{ij}], i,j \in N$
 - binary: $g_{ij} \in \{ 0, 1 \}$ is availability of edge from node i to j
 - weighted: $g_{ij} \in \mathbb{R}$
- **Undirected graph**: $g_{ij} = g_{ji}$ for all $i,j \in N$
- **Directed graph, digraph**: some $g_{ij} \neq g_{ji}$



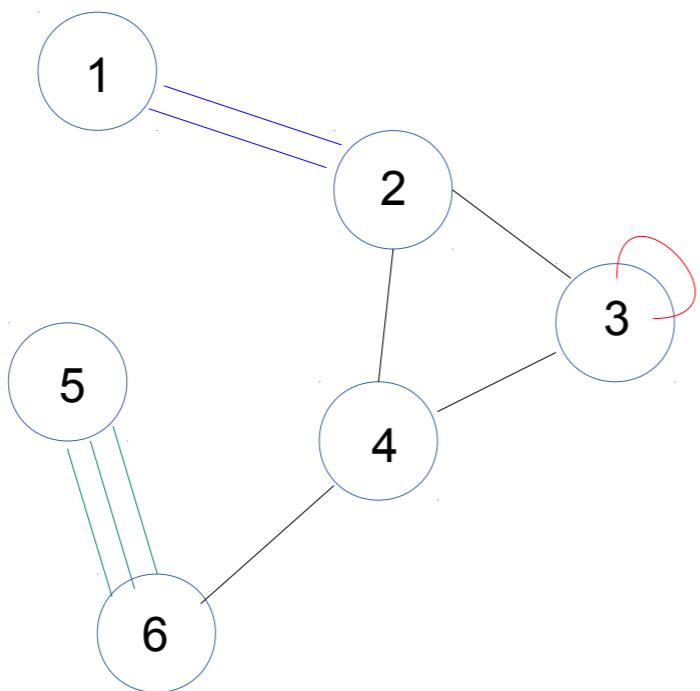
Adjacency Matrix



$$g = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

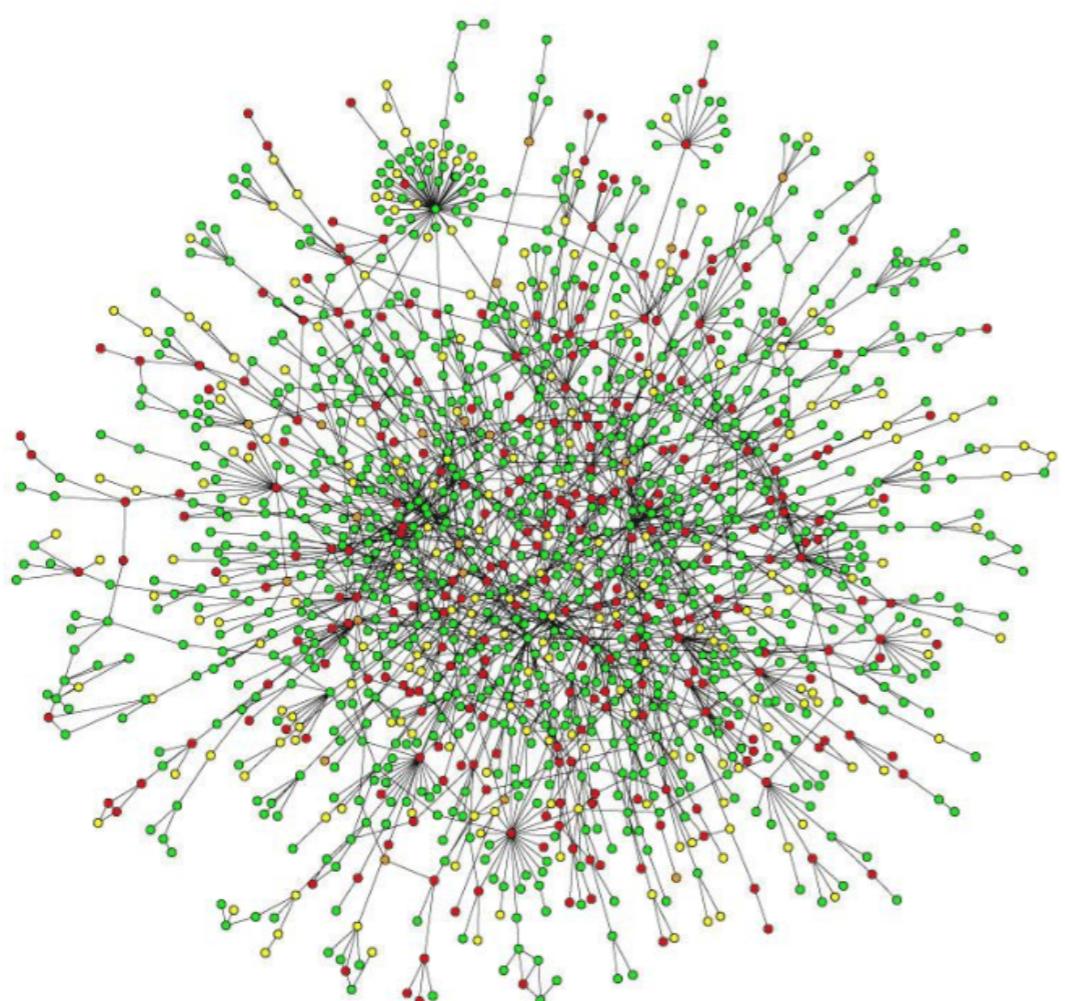
Adjacency Matrix

Multi-edges/Self-edges



$$g = \begin{pmatrix} 0 & 2 & 0 & 0 & 0 & 0 \\ 2 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 1 & 3 & 0 \end{pmatrix}$$

Yeast Protein Interaction Network



sparse matrix, connectivity along diagonal

Dongbo Bu. et al. Topological structure analysis of the protein-protein interaction network in budding yeast. Nucleic Acids Research, 2003, Vol. 31, No. 9 2443-2450

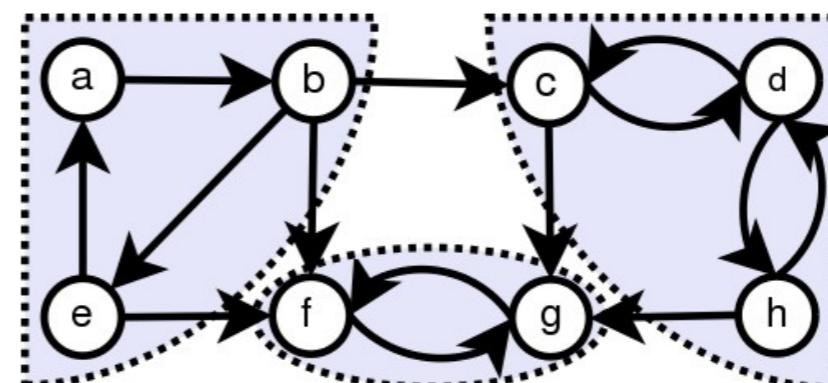
Finding protein interaction networks

The screenshot shows the homepage of the IntAct Molecular Interaction Database. At the top, there is a navigation bar with links for Home, Search, Browse, Data Submission, Downloads, Datasets, Statistics, FAQ, Developer Resources, Contact Us, About IntAct, and Feedback. Below the navigation bar, the main content area features a search form with a text input field, a 'Search' button, a 'Clear' button, and links for 'Show Advanced Fields' and 'MIQL syntax'. A 'reference' link is also present. To the right of the search form, there is a section titled 'Data Content' with statistics: Publications: 13115, Interactions: 457389, and Interactors: 83701. Further down, there is a 'Contributors' section listing MINT, UniProt, SIB, I2D, InnateDB, and Molecular Connections. The URL at the bottom of the page is <https://www.ebi.ac.uk/training/online/course/intact-molecular-interactions->.

<https://www.ebi.ac.uk/training/online/course/intact-molecular-interactions->

Network structure

- **Connected**
 - every two nodes in the network are connected by some path
- **Components**
 - the distinct maximally connected subgraphs
- A directed graph is
 - **connected** if the underlying *undirected* graph is connected
 - **strongly connected** if each node can reach every other node by a *directed* path.



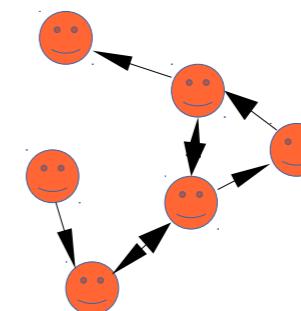
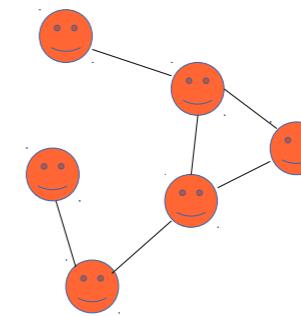
Strongly connected sub-graphs

Network measures and statistics

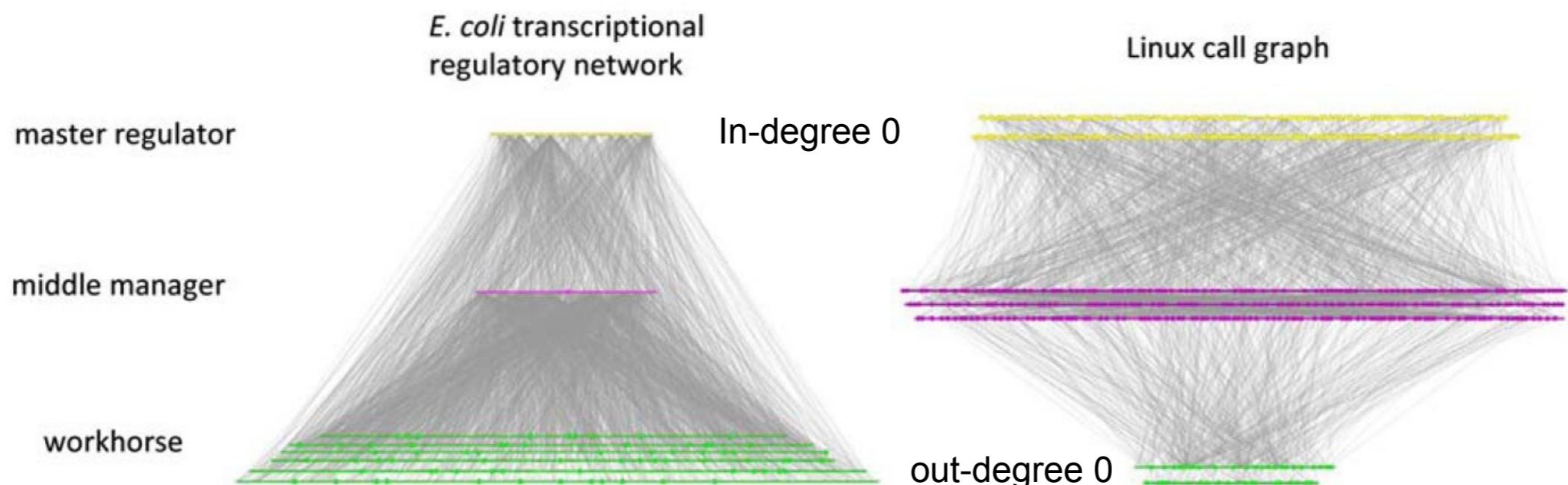
- Degree
- Diameter and average path length
- Clustering

Node degree

- **Neighbourhood** of node i : set of nodes i is connected to
- Undirected graphs:
 - **Degree**: number of edges of node i (cardinality of I)
 $d_i = \sum_j g_{ij} = \sum_j g_{ji}$
- Directed graphs:
 - **In-degree**, incoming edges: $d_{I,i} = \sum_j g_{ji}$
 - **Out-degree**, outgoing edges: $d_{O,i} = \sum_j g_{ij}$



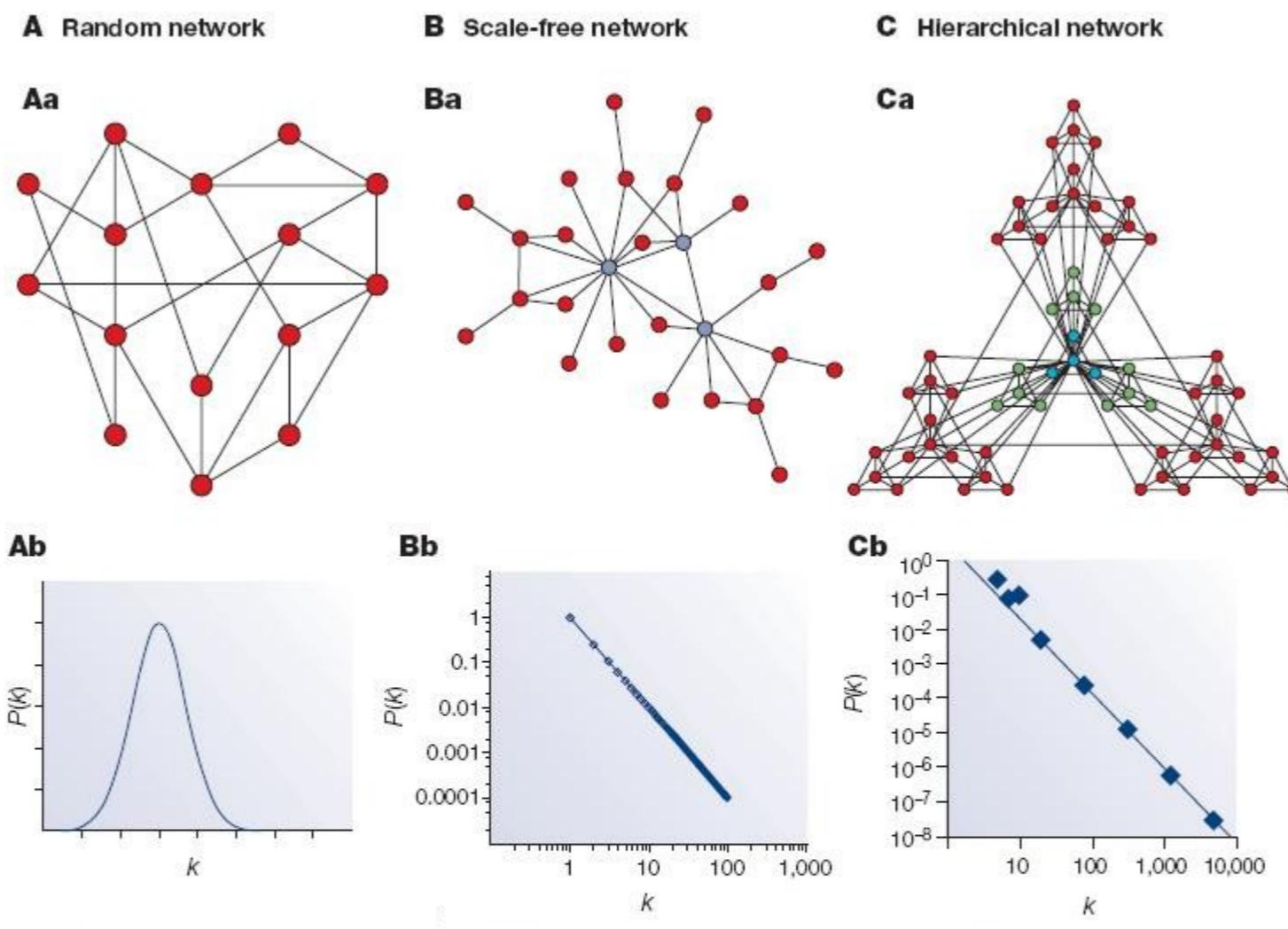
Comparing networks



Yan, K. K., Fang, G., Bhardwaj, N., Alexander, R. P., & Gerstein, M. (2010). Comparing genomes to computer operating systems in terms of the topology and evolution of their regulatory control networks. *Proceedings of the National Academy of Sciences*, 107(20), 9186-9191.

Degree Distribution

- $P(d)$: relative frequencies of nodes with degree d



Scale free networks

- Degree distribution: $P(d) = A d^{-\alpha}$

- Looks the same on every scale

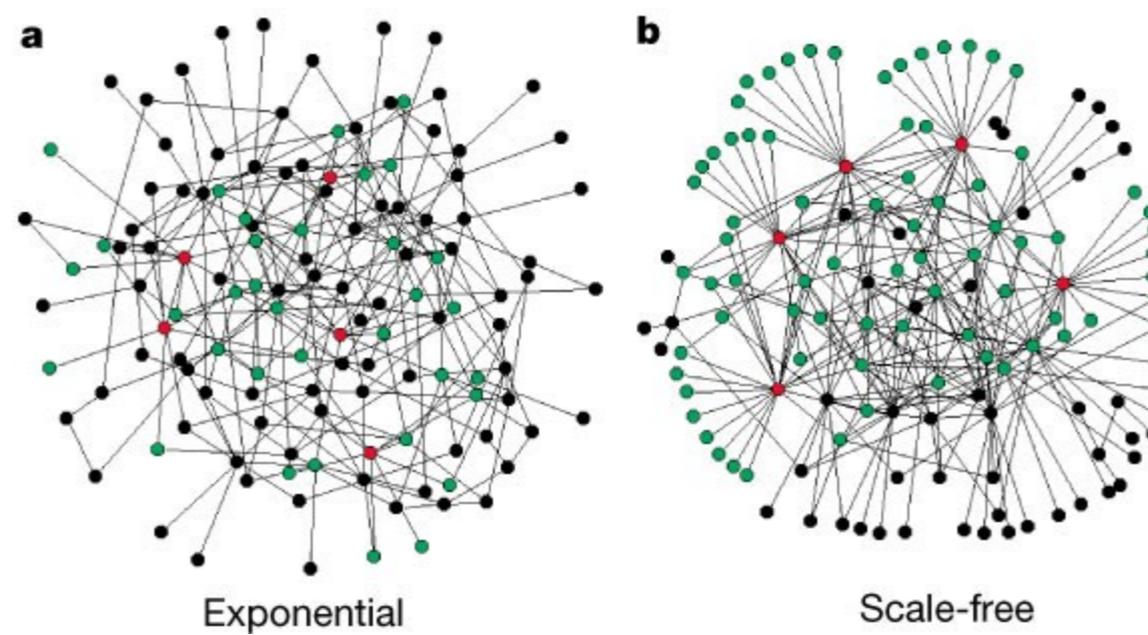
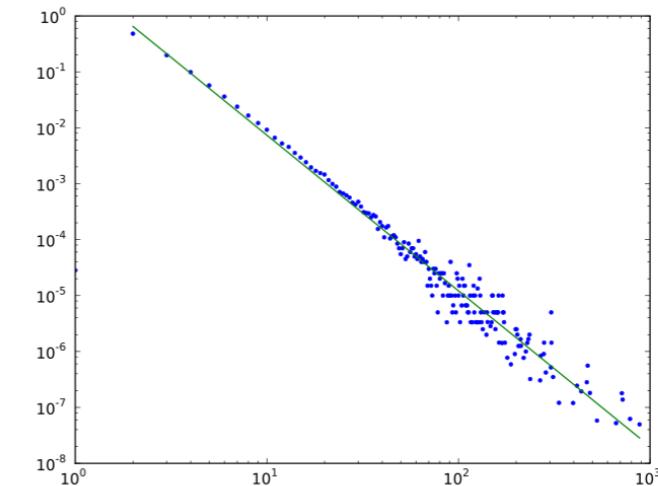
$$P(cd) = A(cd)^{-\alpha} = Ac^{-\alpha} d^{-\alpha} \sim P(d)$$

- de Solla Price (1965) found that citation distribution between papers follows a power law

- Barabási (1999) found that the WWW has power law degree distribution (<http://arxiv.org/abs/cond-mat/9910332>)

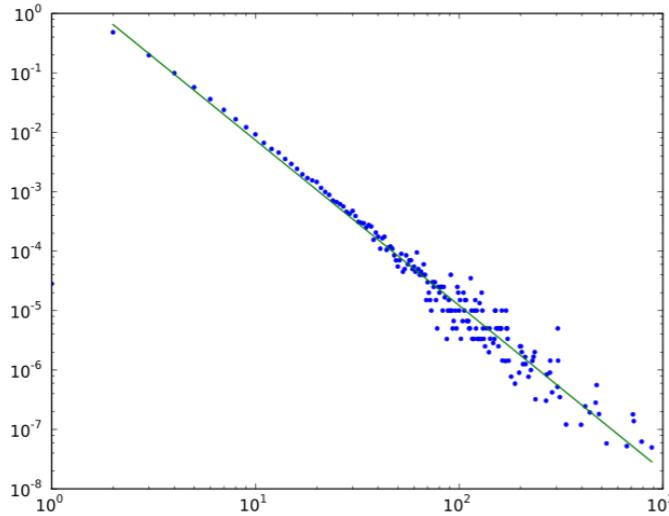
Scale free networks

- Degree distribution: $P(d) = A d^{-\alpha}$
- Hubs have more connections than other nodes
- Robustness: most nodes are not important for connectedness

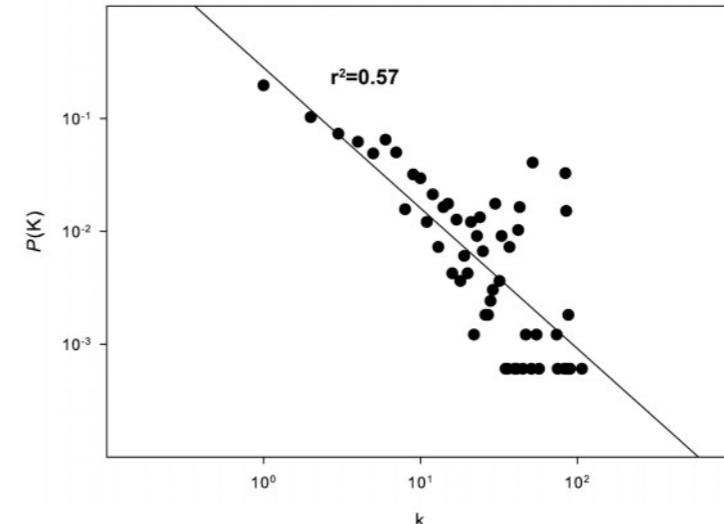


Albert, Jeong and Barabási, Nature 406, 378-382 (2000)

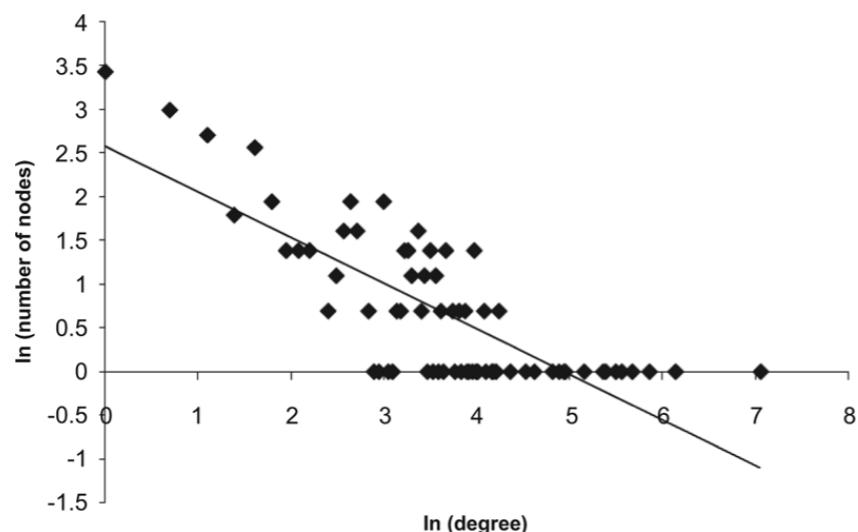
Power law? Really?



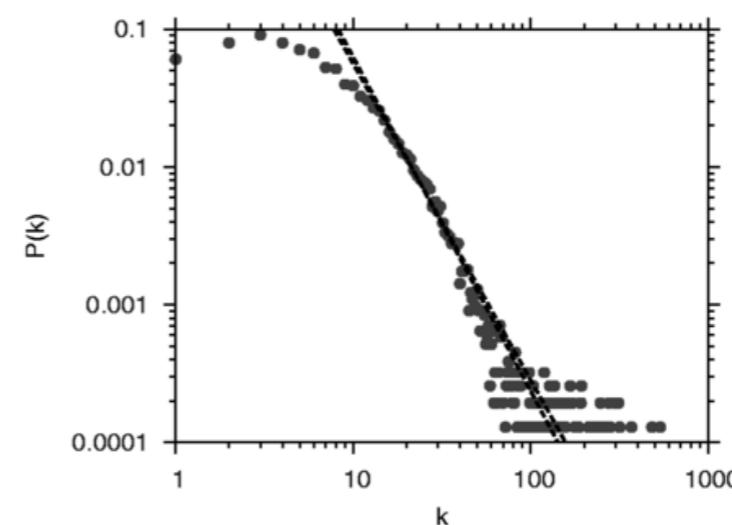
http://en.wikipedia.org/wiki/Barab%C3%A1si%E2%80%93Albert_model



<http://www.tbiomed.com/content/8/1/47/figure/F2>



<http://www.emeraldinsight.com/journals.htm?articleid=1827246&show=htm>



<http://nar.oxfordjournals.org/content/early/2012/10/04/nar.gks904/F5.expansion.htm>

Diameter, Path length

- **Shortest path** between i and j : $L(i, j)$
- **Diameter**: $D = \max_{i,j} L(i, j)$
- **Average path length**:
$$\bar{L} = \frac{\sum_{i \geq j} L(i, j)}{\frac{n(n-1)}{2}}$$
- **Small world networks**: $L \sim \log N$
 - This includes Erdős–Rényi (Random) graphs

6 degrees of separation

What is the average path length between any two people?

Network theory suggested this may be surprisingly small (e.g. work by Pool & Kochen, *Networks* 1978).

Stanley Milgram's experiment:

- sent 96 letters to random people in Omaha
- asked them to forward it to someone in Boston
- but: only pass it on to someone they knew
- receiver then asked to repeat this

18 letters arrived at destination

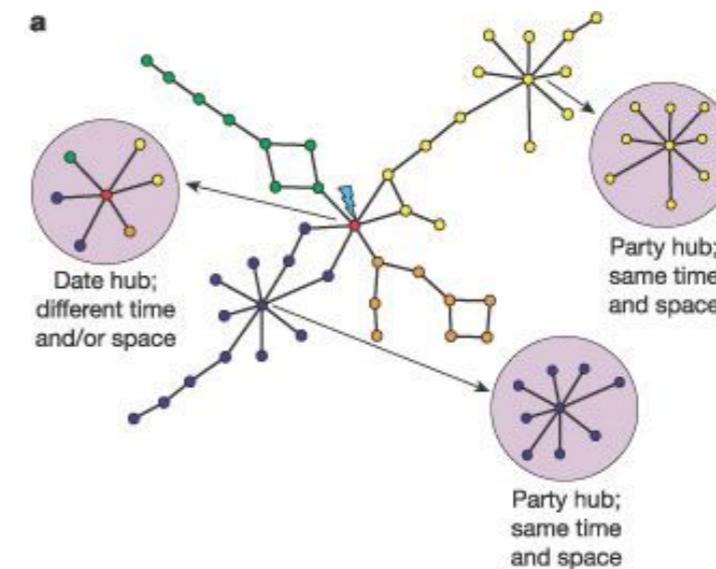
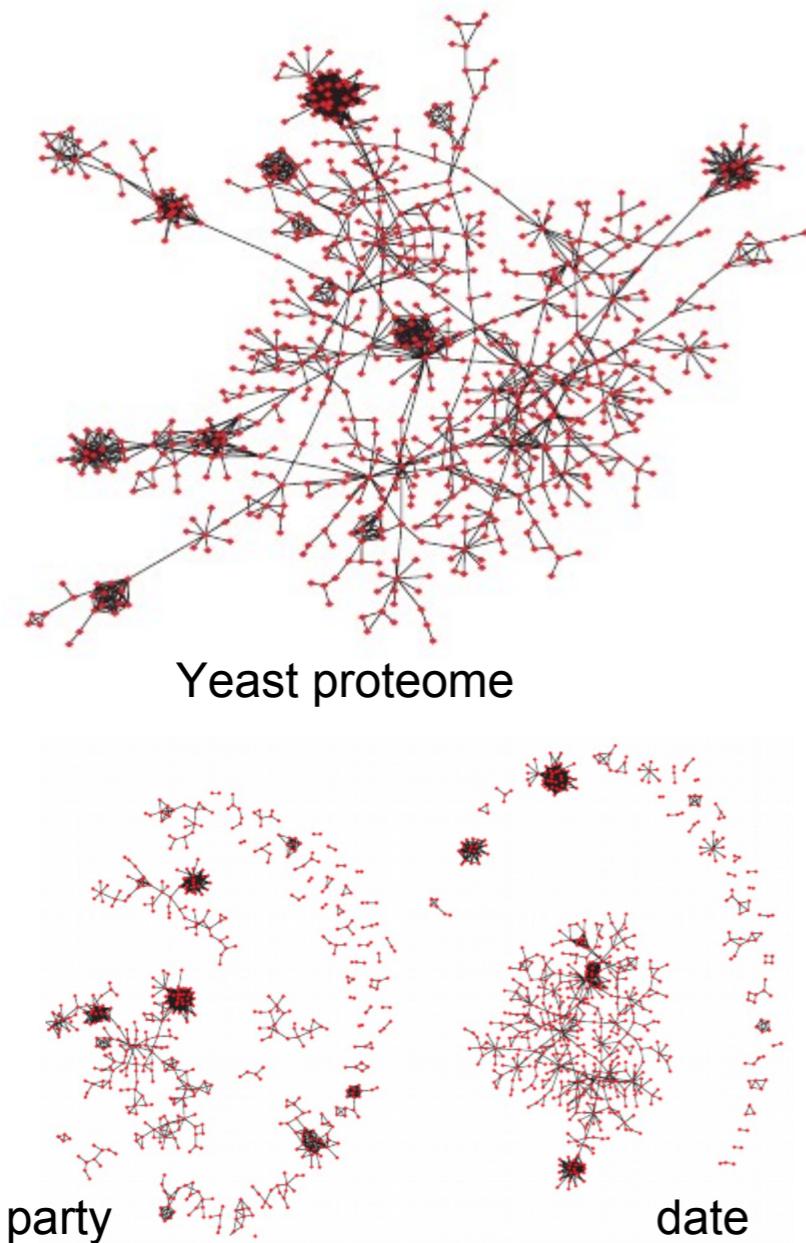
Mean path length: **5.9 steps**

Caveats: spatial proximity, lost letters, bias

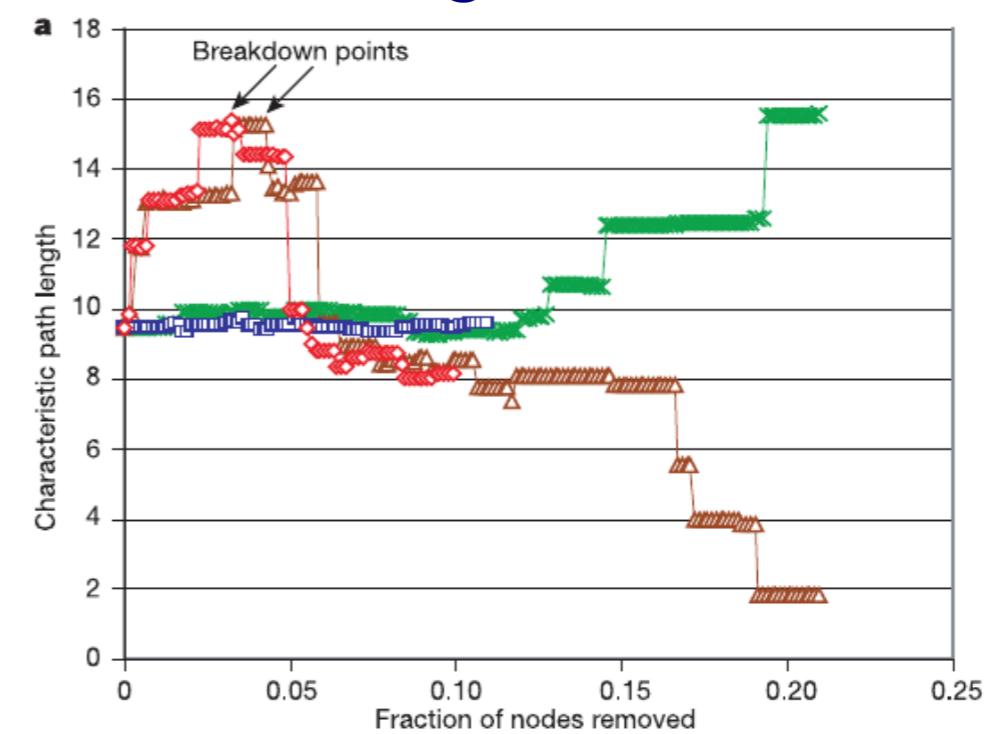


Network Integrity

Temporal Network Analysis



@different times
random, hubs, party hubs, date hubs
@same time



Han et al. Evidence for dynamically organised modularity in the yeast protein-protein interaction network. Nature. 2004 Jul 1;430(6995):88-93

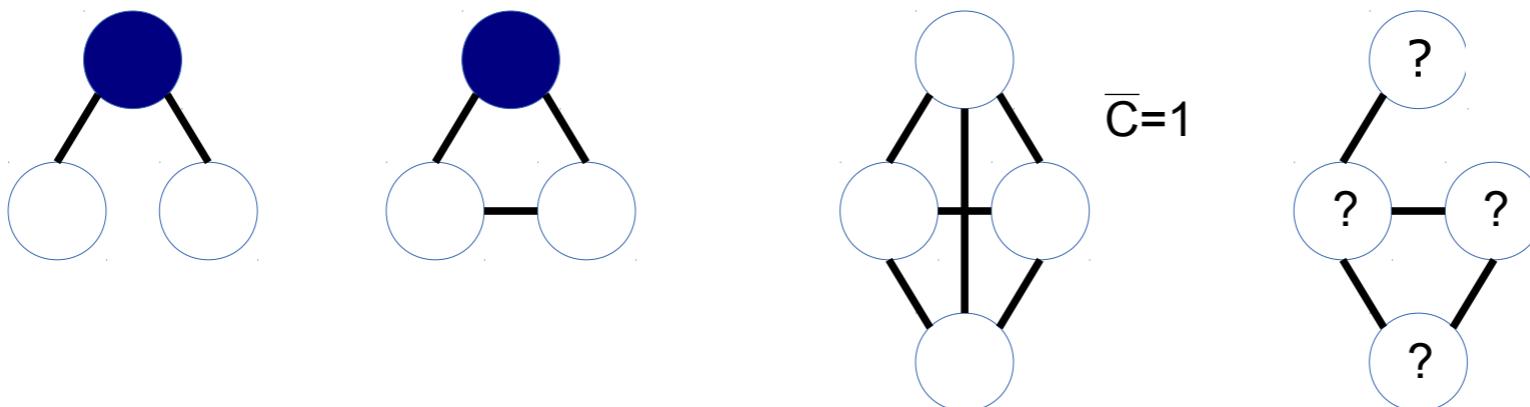
Local clustering coefficient

- On a per-node basis:

$$C_i(g) = \frac{\text{number of connected pairs at } i}{\text{number of pairs at } i}$$

- Average clustering coefficient:

$$\bar{C}_i = \frac{1}{n} \sum_i C_i(g)$$



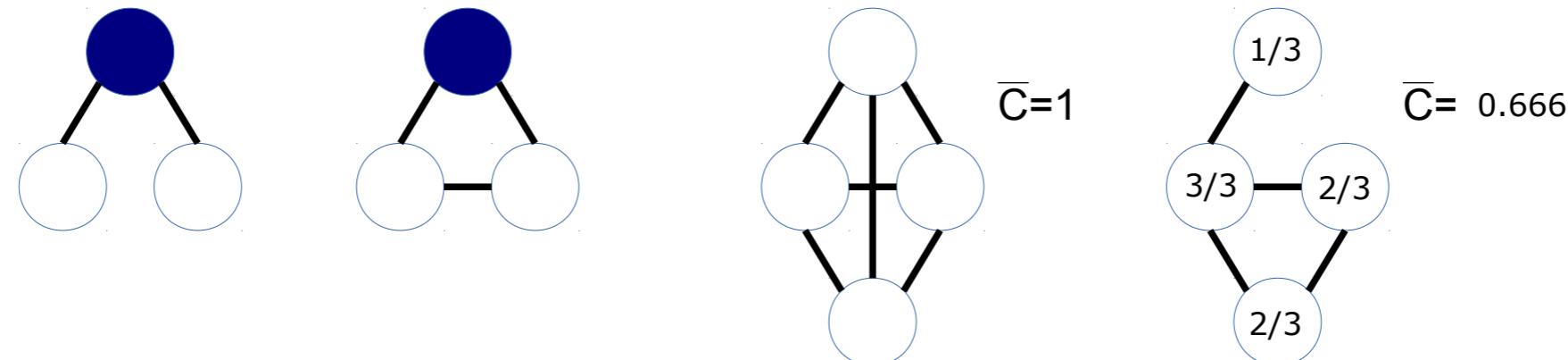
Local clustering coefficient

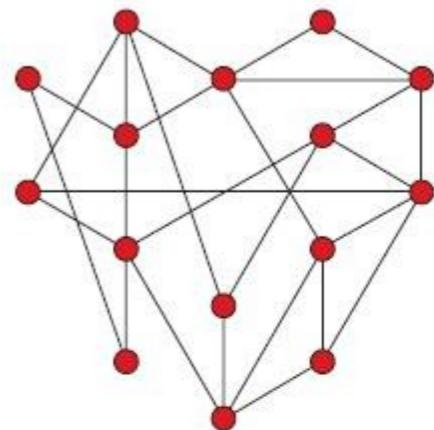
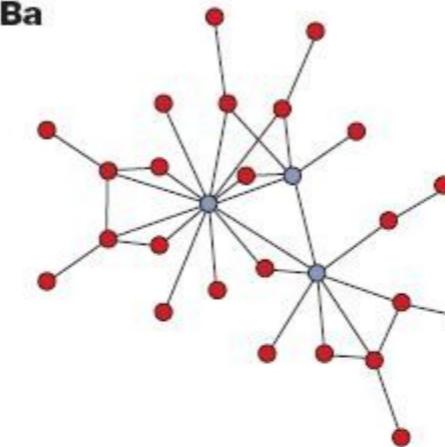
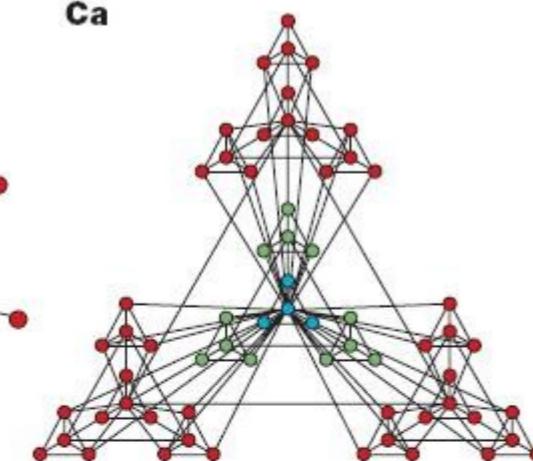
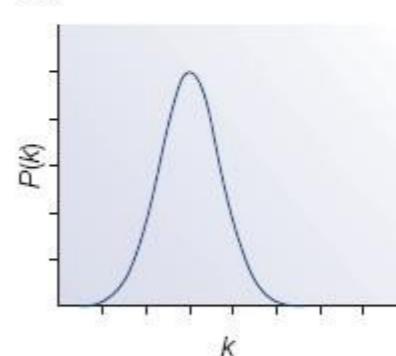
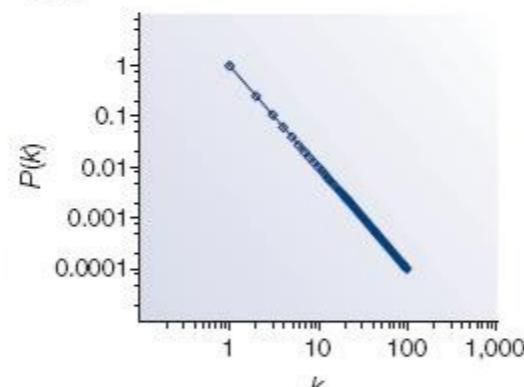
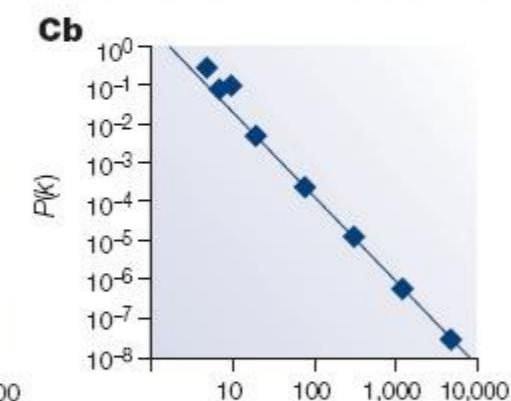
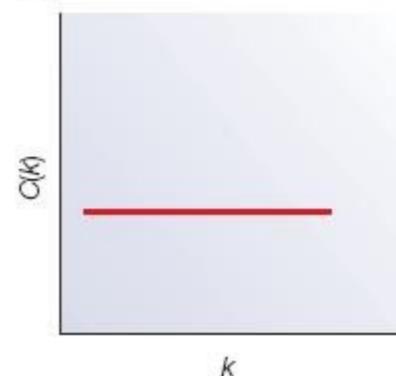
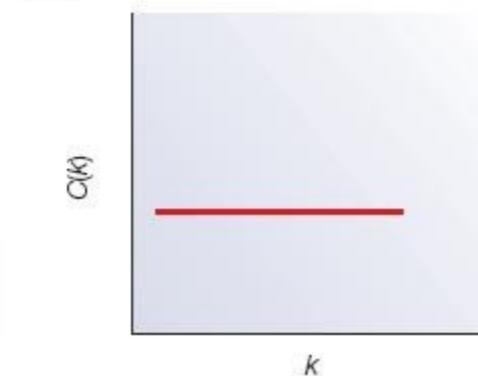
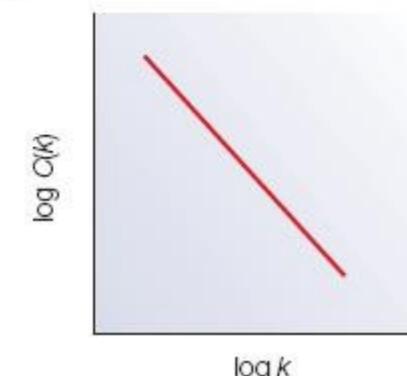
- On a per-node basis:

$$C_i(g) = \frac{\text{number of connected pairs at } i}{\text{number of pairs at } i}$$

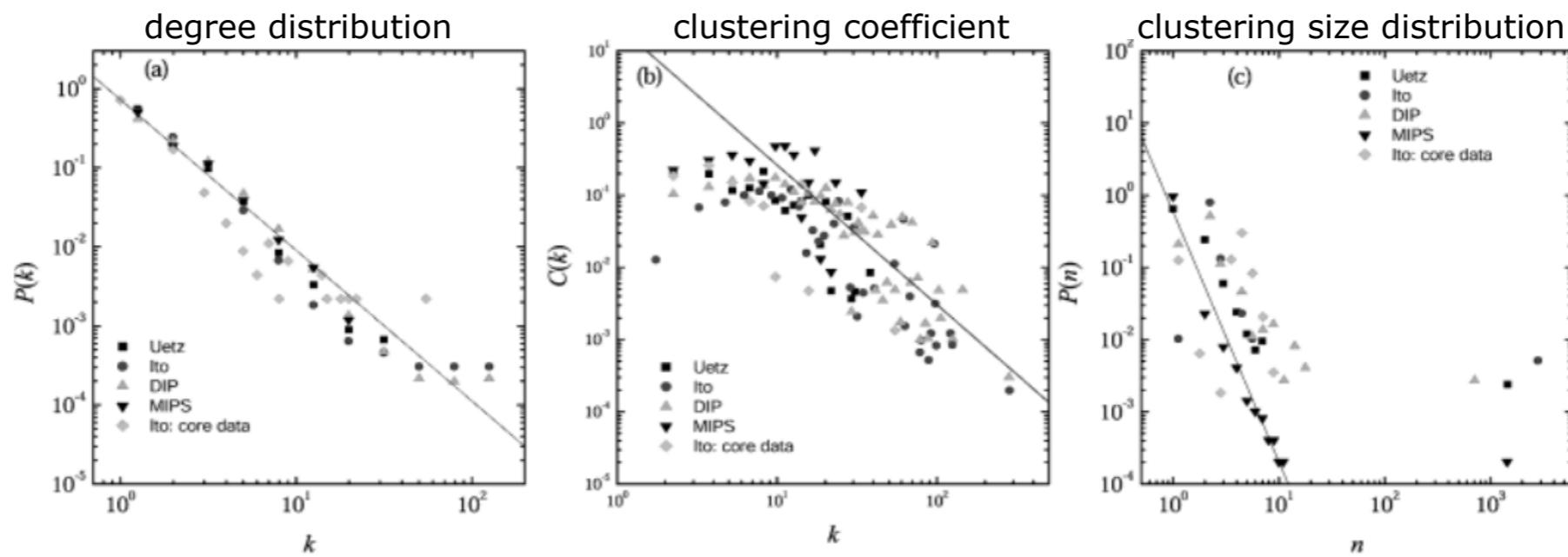
- Average clustering coefficient:

$$\bar{C}_i = \frac{1}{n} \sum_i C_i(g)$$



A Random network**Aa****B Scale-free network****Ba****C Hierarchical network****Ca****Ab****Bb****Cb****Ac****Bc****Cc**

Yeast Proteome

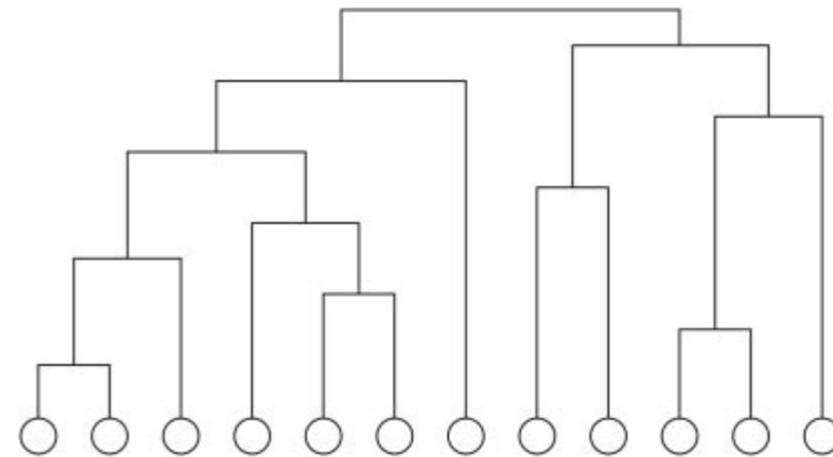
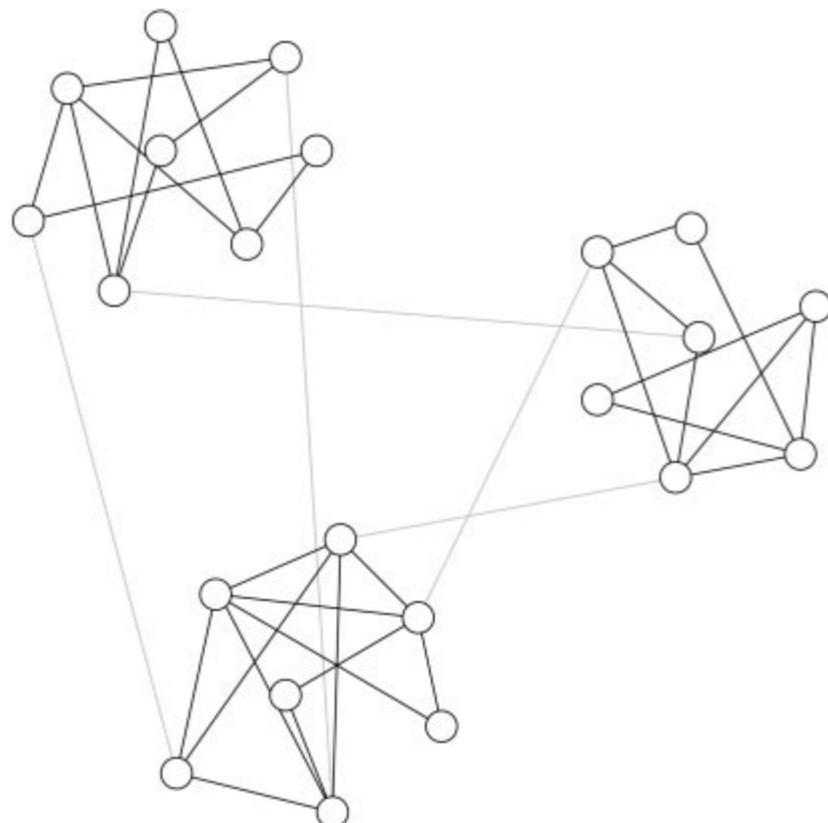


Dataset	Number of proteins in the database (N)	Number of proteins with at least one interaction (N_{int})	Total number of interactions (L)	Largest cluster (LC)		Diameter of the largest cluster	Clustering coefficient (C/C_{rand})	Degree exponent (γ)	Average segregation ($m(1)/m^*$)
				Number of protein in the LC (N_{LC})	Number of interaction in the LC (L_{LC})				
MIPS	6 745	2 043 (118)	5 434	1 441	4 538	7.71	34.91	2.34	3.28
DIP	5 798	5 798 (352)	20 098	4 198	15 892	4.9	117.09	2.50	3.48
Uetz	2 115	1 870 (74)	4 480	1 458	3 941	6.8	54.64	2.32	2.28
Ito	3 280	3 280 (82)	8 868	2 840	8 371	4.9	36.40	2.44	1.49
Ito core	797	797 (52)	1 560	417	1 055	6.2	4.94	2.1	7.06

Yook et al., Proteomics, 2004

Community detection

Communities are groups of densely interconnected nodes, which connect with fewer links to other communities.



Attach weights to edges and add strongest first
(hierarchical clustering)

Community detection

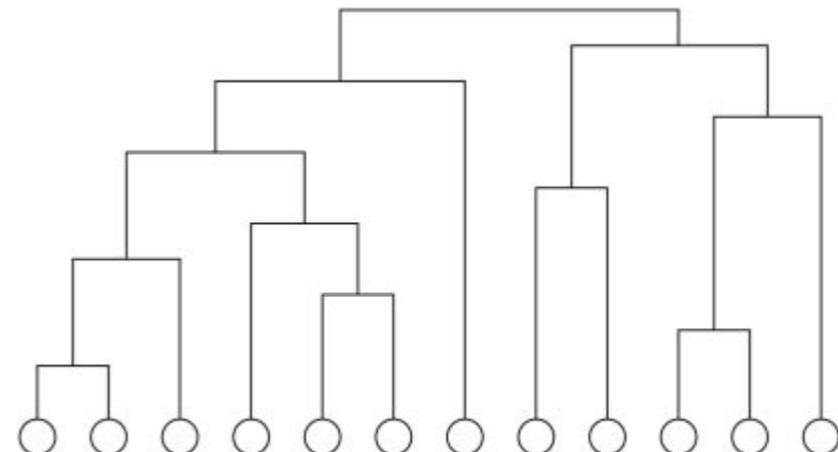
Hierarchical Clustering Algorithms

Add edges progressively according to weight, which can be:

- number of node/edge independent paths
- total number of paths (weighted by length, e.g. a^l)

Problem:

- Isolated nodes are wrongly removed from communities



Community detection

Girvan & Newman Algorithm

Progressively remove least integrated edges according to betweenness centrality:

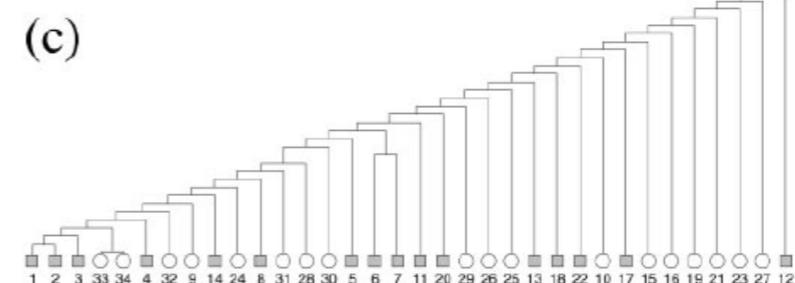
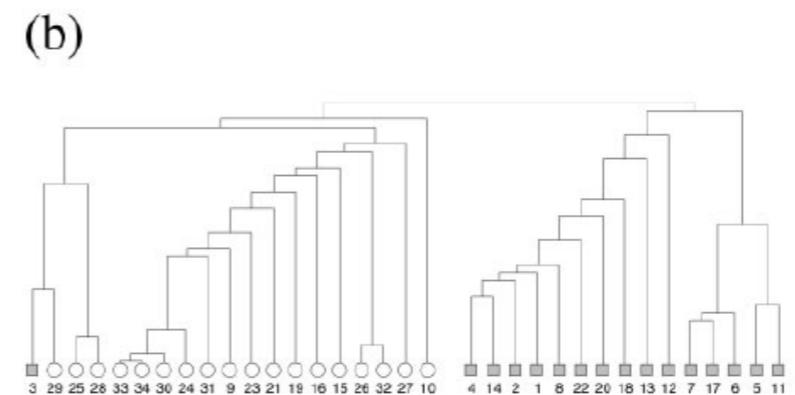
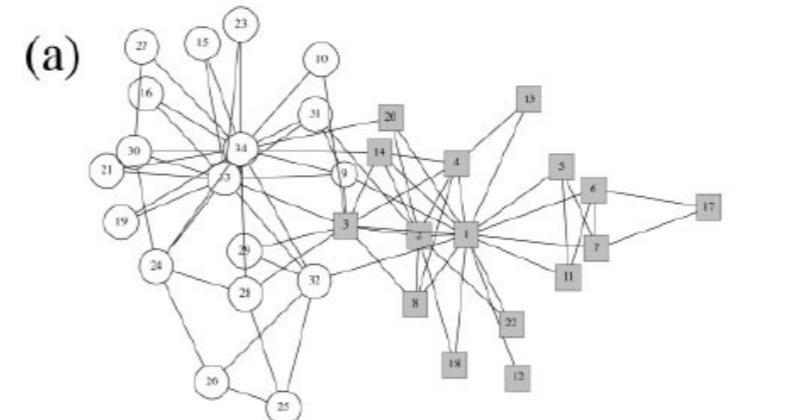
- number of shortest paths running through an edge
- measure of information flow

Procedure:

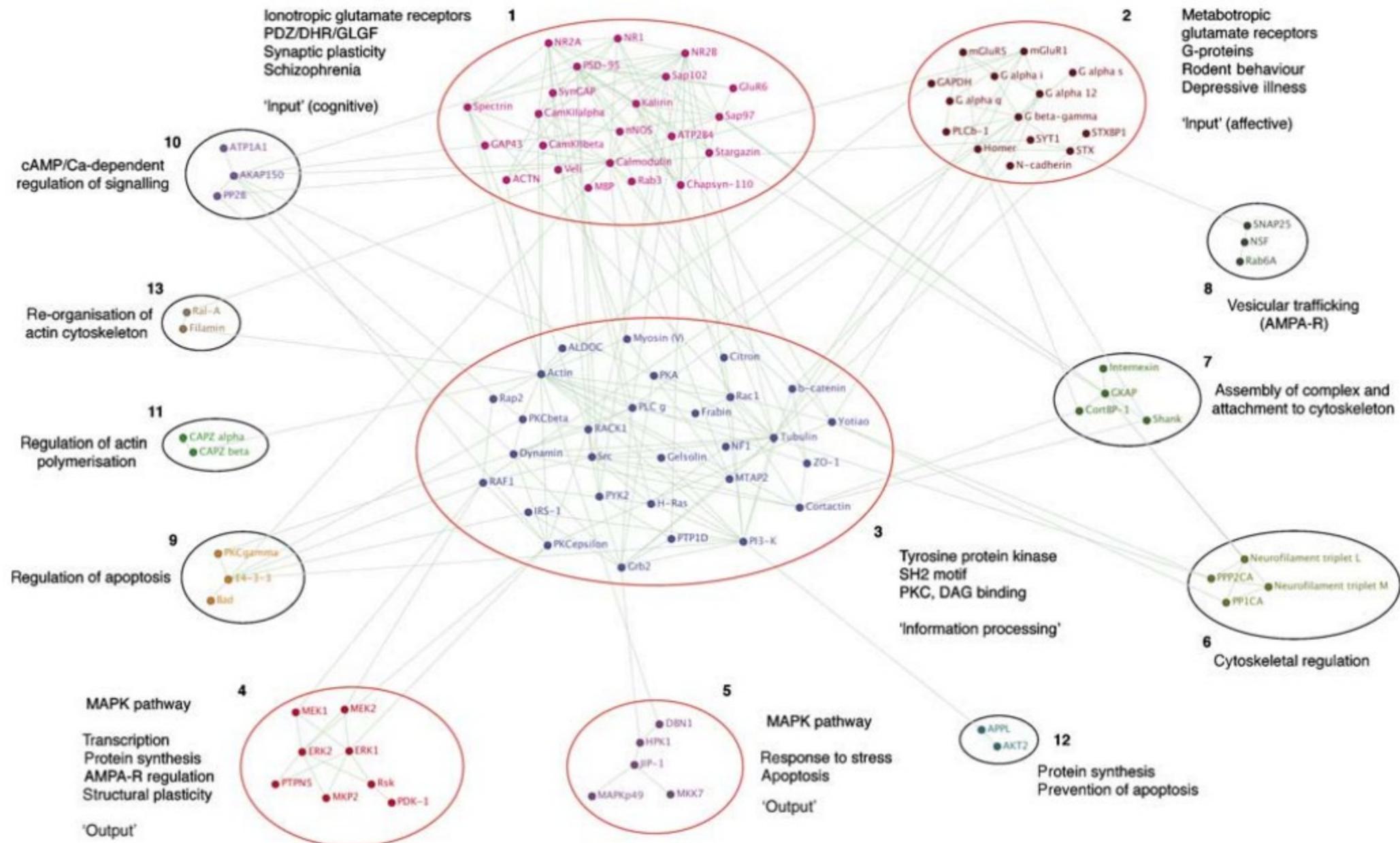
1. Calculate betweenness for all edges
2. Remove edge with highest betweenness
3. Recompute betweenness
4. Goto 2 until all edges removed

... $O(N^3)$

Girvan, Michelle, and Mark EJ Newman. "Community structure in social and biological networks." *Proceedings of the National Academy of Sciences* 99.12 (2002): 7821-7826.



Zachary's Karate Club



Pocklington, Andrew J., et al. "The proteomes of neurotransmitter receptor complexes form modular networks with distributed functionality underlying plasticity and behaviour." Molecular systems biology 2.1 (2006).

G-N Algorithm with edge betweenness

Some further reading

- A few important papers and reviews:

Albert, R., Jeong, H., & Barabási, A. L. (2000). Error and attack tolerance of complex networks. *Nature*, 406(6794), 378-382.

Strogatz, S. H. (2001). Exploring complex networks. *Nature*, 410(6825), 268-276.

Newman, M. E. (2003). The structure and function of complex networks. *SIAM review*, 45(2), 167-256.

Barabási, A. L., & Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, 5(2), 101-113.

Palla, G., Derényi, I., Farkas, I., & Vicsek, T. (2005). Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435(7043), 814-818.

Barabási, A. L., Gulbahce, N., & Loscalzo, J. (2011). Network medicine: a network-based approach to human disease. *Nature Reviews Genetics*, 12(1), 56-68.

- Review of community detection algorithms:

A Lancichinetti, S Fortunato (2010). Community detection algorithms: a comparative analysis. ArXiv:0908.1062, <http://arxiv.org/abs/0908.1062>

- Textbook:

Newman, M. (2009). Networks: an introduction. Oxford University Press.

