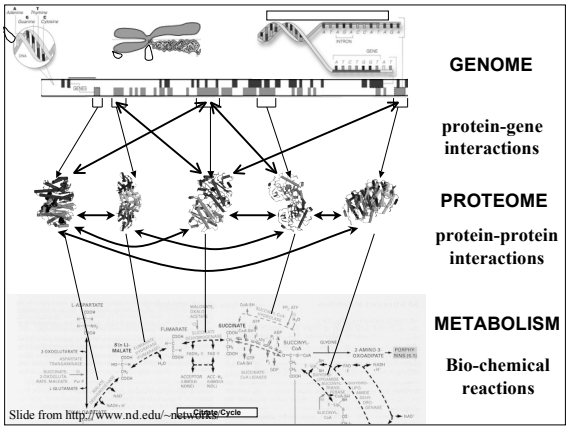
Bioinformatics 2

Proteomics



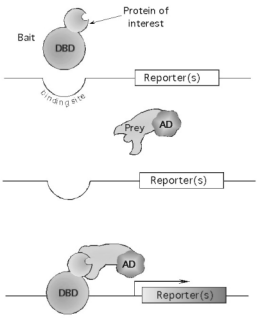
Slide from http://www.nd.edu/~networks

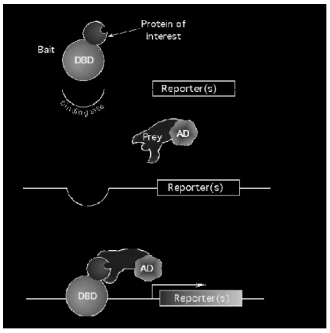
Proteomics

- What is it?
  - Reveal protein interactions
  - Protein profiling in a sample
- Yeast two hybrid screening
- High throughput 2D PAGE
- Automatic analysis of 2D Page

Yeast two hybrid

- Use two mating strains of yeast
- In one strain fuse one set of genes to a transcription factor DNA binding site
- In the other strain fuse the other set of genes to a transcriptional activating domain
- Where the two proteins bind, you get a functional transcription factor.





Data obtained

- Depending on sample, you get a profile of potential protein-protein interactions that can be used to predict functional protein complexes.
- False positives are frequent.
- Can be confirmed by affinity purification etc.

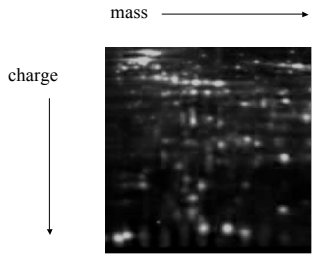
Armstrong, 2006

Proteomics - PAGE techniques

- Proteins can be run through a poly acrylamide gel (similar to that used to seqparate DNA molecules).
- Can be separated based on charge or mass.
- 2D Page separates a protein extract in two dimensions.

Armstrong, 2006

2D Page



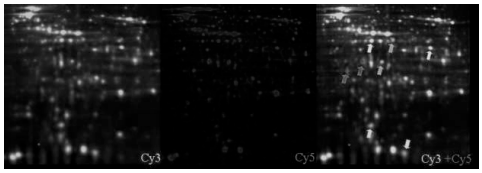
Armstrong, 2006

DiGE

- We want to compare two protein extracts in the way we can compare two mRNA extracts from two paired samples
- Differential Gel Electrophoresis
- Take two protein extracts, label one green and one red (Cy3 and Cy5)

Armstrong, 2006

DiGE



- The ratio of green:red shows the ratio of the protein across the samples.

Armstrong, 2006

Identifying a protein 'blob'

- Unlike DNA microarrays, we do not normally know the identify of each 'spot' or blob on a protein gel.
- We do know two things about the proteins that comprise a blob:
  - mass
  - charge

Armstrong, 2006

Identifying a protein ‘blob’

- Mass and Charge are themselves insufficient for positive identification.
- Recover from selected blobs the protein (this can be automated)
- Trypsin digest the proteins extracted from the blob (chops into small pieces)

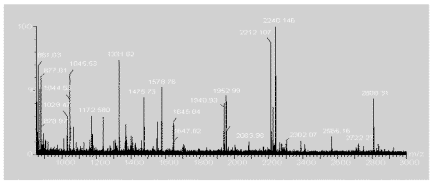
Armstrong, 2006

Identifying a protein ‘blob’

- Take the small pieces and run through a mass spectrometer. This gives an accurate measurement of the weight of each.
- The total weight and mass of trypsin digested fragments is often enough to identify a protein.
- The mass spec is known as a MALDI-TOFF

Armstrong, 2006

Identifying a protein ‘blob’



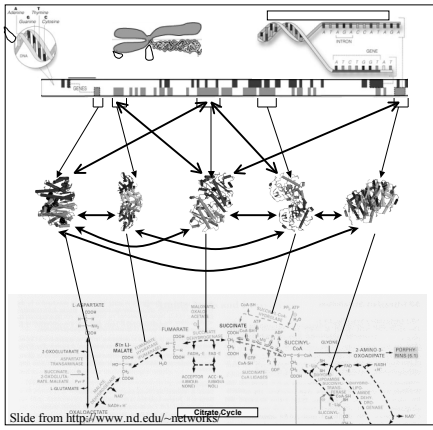
MALDI-TOFF output from myosin
Good for rapid identification of single proteins.
Does not work well with protein mixtures.

Armstrong, 2006

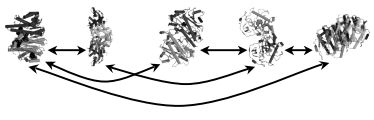
Identifying a protein ‘blob’

- When MALDI derived information is insufficient. Need peptide sequence:
- Q-TOF allows short fragments of peptide sequences to be obtained.
- We now have a total mass for the protein, an exact mass for each trypsin fragment and some partial amino acid sequence for these fragments.

Armstrong, 2006



Slide from http://www.nd.edu/~networks/

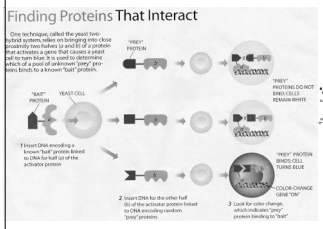


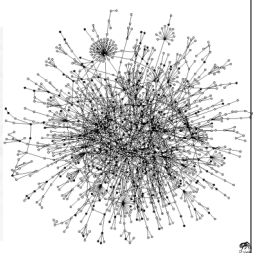
Armstrong, 2006
Slide from http://www.nd.edu/~networks/

Yeast protein network

Nodes: proteins

Links: physical interactions (binding)





P. Uetz, *et al.* *Nature* **403**, 623-7 (2000).

Slide from http://www.nd.edu/~networks/

Building networks…

- Biological Networks
  - Random networks
  - Scale free networks
  - Small worlds
- Metabolic Networks
- Proteomic Networks
- The Mammalian Synapse
- Other synapse models?

Armstrong, 2006

Biological Networks

- Genes - act in cascades
- Proteins - form functional complexes
- Metabolism - formed from enzymes and substrates
- The CNS - neurons act in functional networks
- Epidemiology - mechanics of disease spread
- Social networks - interactions between individuals in a population
- Food Chains

Armstrong, 2006

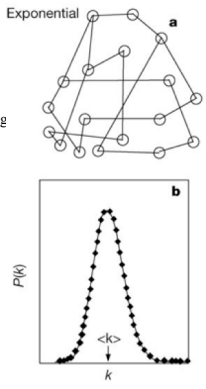
Protein Interactions

- Individual Proteins form functional complexes
- These complexes are semi-redundant
- The individual proteins are sparsely connected
- The networks can be represented and analysed as an undirected graph

Armstrong, 2006

Large scale organisation

- Networks in biology generally modeled using classic random network theory.
- Each pair of nodes is connected with probability p
- Results in model where most nodes have the same number of links $<k>$
- The probability of any number of links per node is $P(k) \simeq e^{-k}$

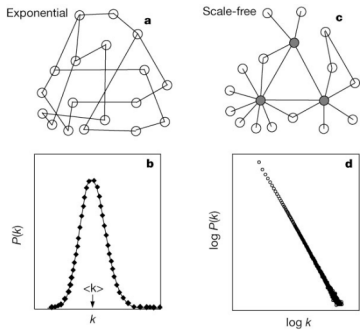


Armstrong, 2006

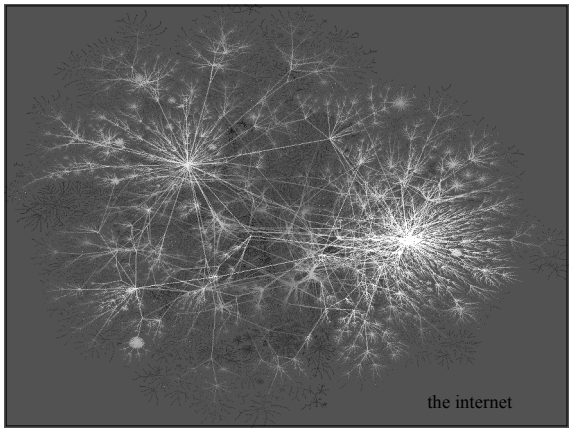
Non-biological networks

- Research into WWW, internet and human social networks observed different network properties
  - 'Scale-free' networks
  - P(k) follows a power law: $P(k) \approx k^{-\gamma}$
  - Network is dominated by a small number of highly connected nodes - hubs
  - These connect the other more sparsely connected nodes

Armstrong, 2006



Armstrong, 2006

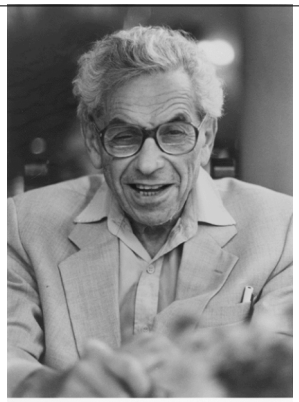


Small worlds

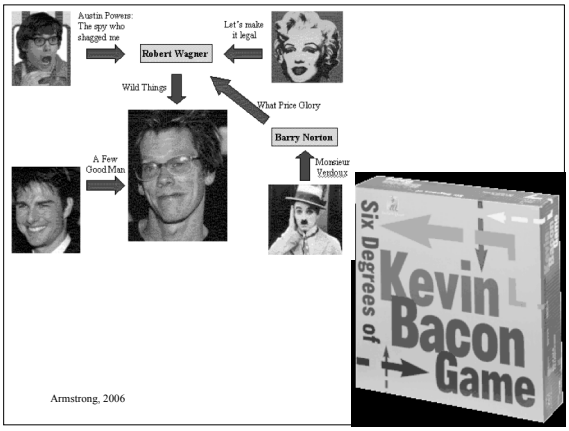
- General feature of scale-free networks
  - any two nodes can be connected by a relatively short path
  - average between any two people is around 6
    - What about SARS???
  - 19 clicks takes you from any page to any other on the internet.

Armstrong, 2006

Armstrong, 2006



Paul Erdős, the most prolific mathematician who ever lived, has no home and no job, but he has wandered the world for over fifty years, inspiring other mathematicians. From the documentary N is a Number: A Portrait of Paul Erdős © 1993 by George Csicsery



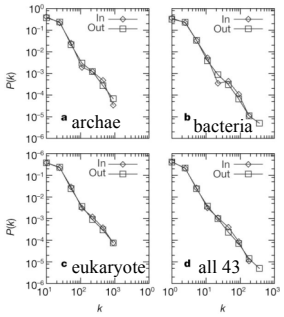
Armstrong, 2006

Biological organisation

Jeong et al., 2000 The large-scale organisation of metabolic networks. Nature 407, 651-654

- Pioneering work by Oltvai and Barabasi
- Systematically examined the metabolic pathways in 43 organisms
- Used the WIT database
  - 'what is there' database
  - http://wit.mcs.anl.gov/WIT2/
  - Genomics of metabolic pathways

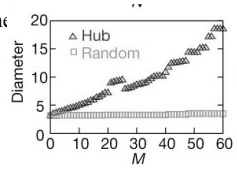
Using metabolic substrates as nodes



=scale free!!!

Random mutations in metabolic networks

- Simulate the effect of random mutations or mutations targeted towards hub nodes.
  - Measure network diame
  - Sensitive to hub attack
  - Robust to random



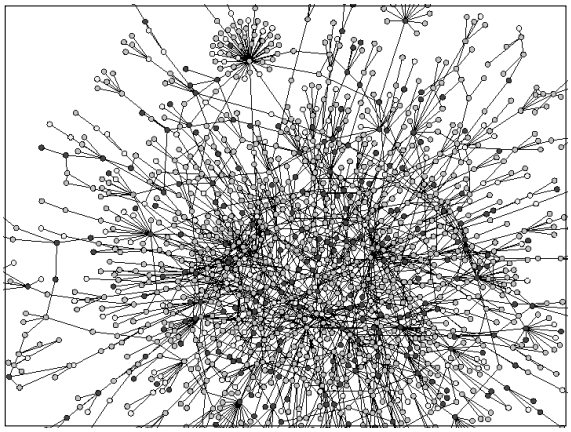
Consequences for scale free networks

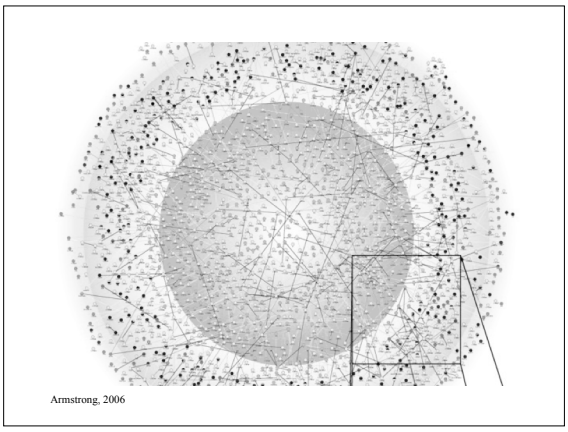
- Removal of highly connected hubs leads to rapid increase in network diameter
  - Rapid degeneration into isolated clusters
  - Isolate clusters = loss of functionality
- Random mutations usually hit non hub nodes
  - therefore robust
- Redundant connectivity (many more paths between nodes)

Protein Networks

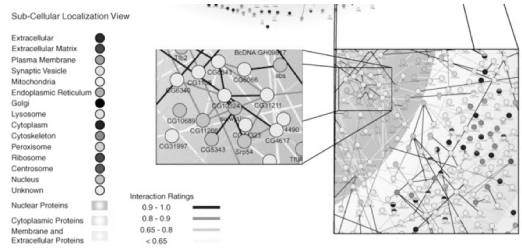
Protein Networks

- Networks derived from high throughput yeast 2 hybrid techniques
  - yeast
  - *Drosophila melanogaster*
  - *C.elegans*
- Predictive value of reconstructed networks
- Sub-clusters and sub-architecture
- Comparison with known sub-networks, pathways and protein complexes



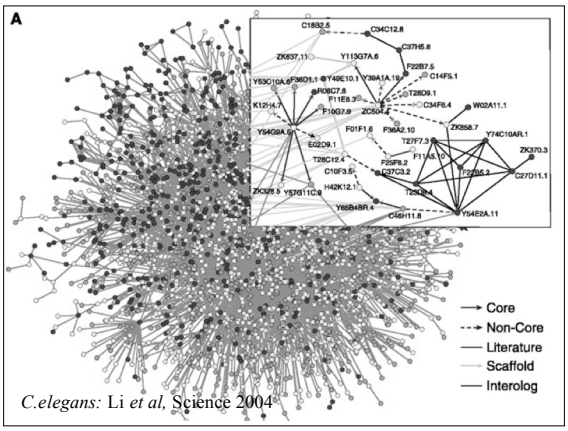


Armstrong, 2006



Giot *et al*, Science 2003

Armstrong, 2006



C.elegans: Li *et al,* Science 2004

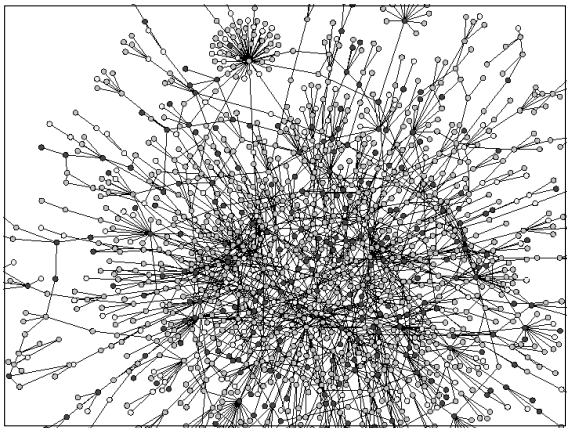
Predictive value of networks

Jeong et al., (2001) Lethality and Centrality in protein networks. Nature 411 p41

- In the yeast genome, the essential vs. unessential genes are known.
- Rank the most connected genes
- Compare known lethal genes with rank order

k	fraction	%lethal
<6	93%	21%
>15	0.7%	62%

Armstrong, 2006

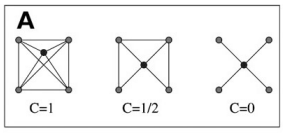


What about known complexes?

- OK, scale free networks are neat but how do all the different functional complexes fit into a scale free proteome arrangement?
  - e.g. ion channels, ribosome complexes etc?
- Is there substructure within scale free networks?
  - Examine the clustering co-efficient for each node.

Clustering co-efficients and networks.

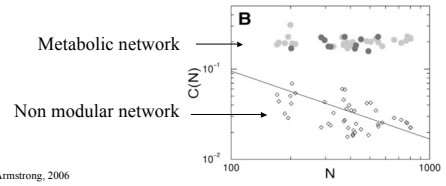
- $C_i = 2n/k_i(k_i-1)$
- n is the number of direct links connecting the k_i nearest neighbours of node *i*
- A node at the centre of a fully connected cluster has a C of 1

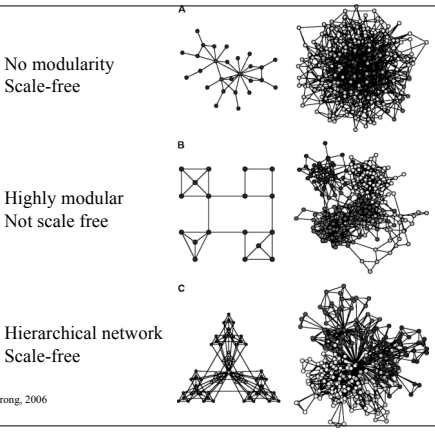


Clustering co-efficients and networks.

Ravasz et al.,(2002) Hierarchical Organisation of Modularity in Metabolic Networks. Science 297, 1551-1555

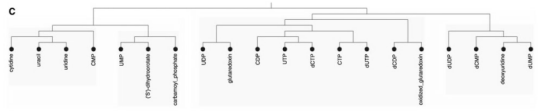
- The modularity (ave C) of the metabolic networks is an order of magnitude higher than for truly scale free networks.





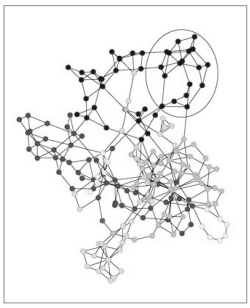
Clustering on C

- Clustering on the basis of C allows us to rebuild the sub-domains of the network



- Producing a tree can predict functional clustered arrangements.

Cluster analysis on the network



Genes 2 Cognition

Armstrong, 2006