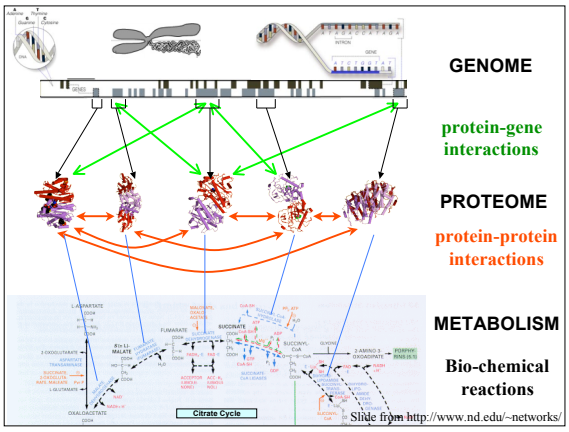
Bioinformatics 2

From genomics & proteomics to biological networks

Aims

- Briefly review functional genomics
- Biological Networks in general
- Genetic Networks
- Briefly review proteomics
- Protein Networks



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

Biological Profiling

- Microarrays
  - cDNA arrays
  - oligonucleotide arrays
  - whole genome arrays
- Proteomics
  - yeast two hybrid
  - PAGE techniques

Why microarrays?

- What genes are expressed in a tissue and how does that tissue respond to one of a number of factors:
  - change in physical environment
  - experience
  - pharmacological manipulation
  - influence of specific mutations

What do we actually get?

- A snap-shot of the mRNA profile in a biological sample
- With the correct experimental conditions we can compare two situations
- Not all biological processes are regulated through mRNA expression levels

What can we learn?

- Identify functionally related genes
- Find promoter regions (common regulation)
- Predict genetic interactions
- If we change one variable a network of gene responses should compensate
- Homeostasis is a fundamental principle of biology - almost all biological systems exist in a controlled state of negative feedback.

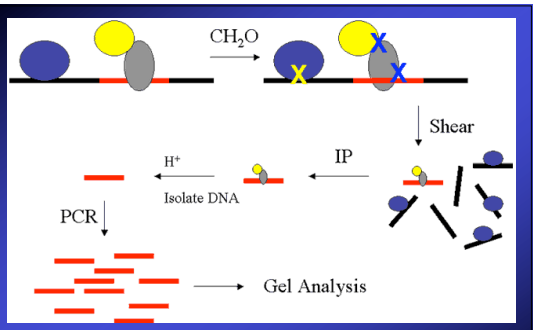
The Transcriptome

- Microarrays work by revealing DNA-DNA binding.
- Transcriptional activators also bind DNA
- Spot genomic DNA onto glass slides
- Label protein extracts
- Hybridise to the genomic probes
- Reveals domains that include promoter regions

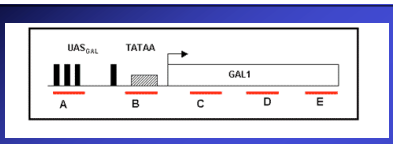
ChIP to Chip

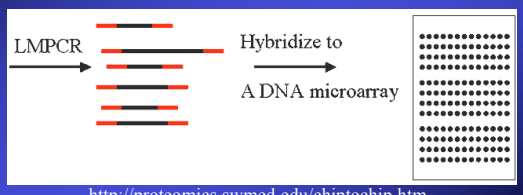
Chromatin Immunoprecipitation to Microarray (Chip)

Protein-DNA interactions
de-novo prediction has many false positives
Which DNA sites do actually bind a specific TF?
Requires an antibody to the protein



http://proteomics.swmed.edu/chiptochip.htm





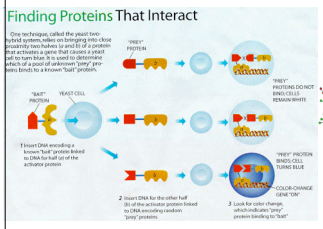
http://proteomics.swmed.edu/chiptochip.htm

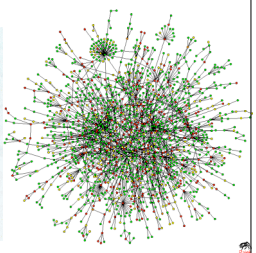
Biological networks

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, 2007

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, 2007

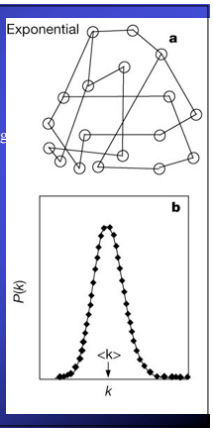
Protein Interactions

- 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

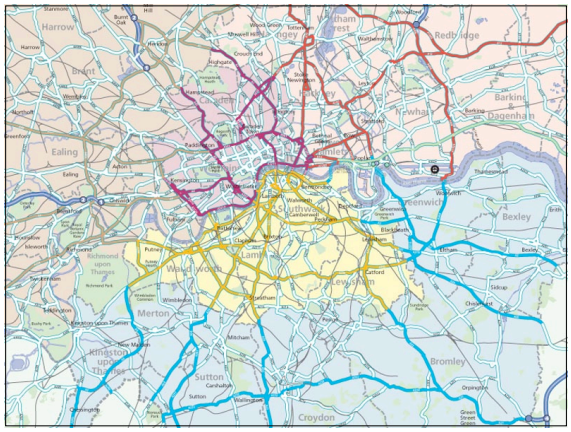
Armstrong, 2007

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 $P(k) \approx e^{-k}$

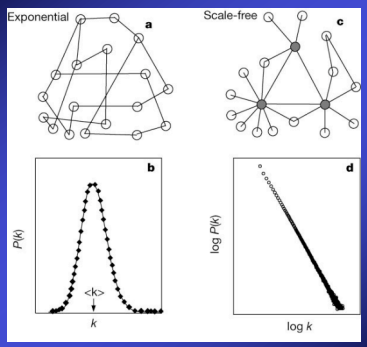


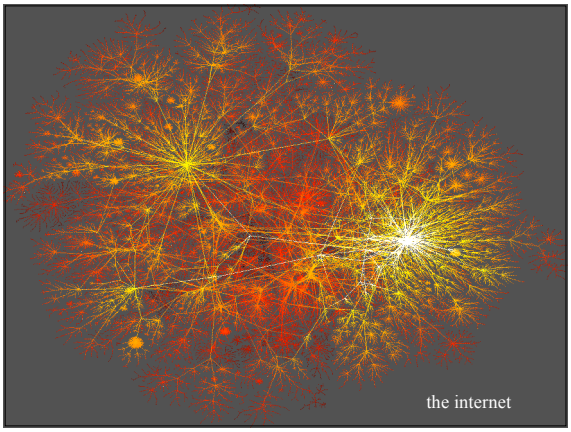
Armstrong, 2007



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

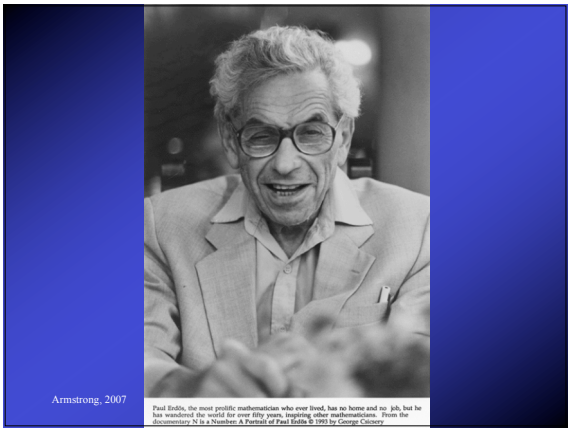




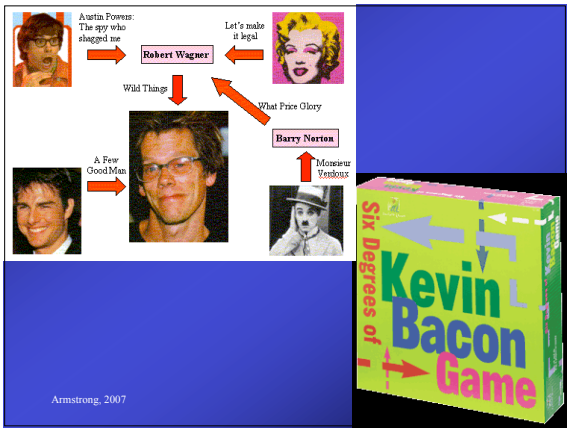
the internet

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.



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

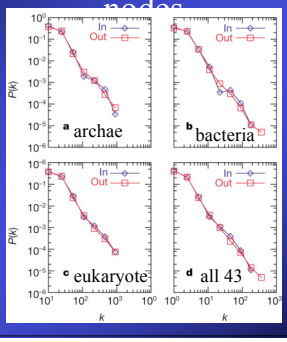


Biological organisation

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

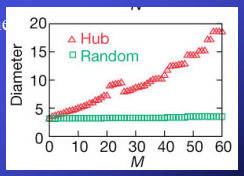
Using metabolic substrates as nodes



=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



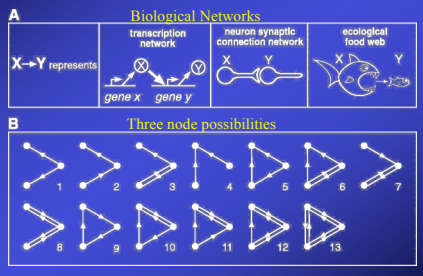
Consequences for scale free networks

- 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)

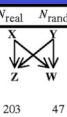
Network Motifs

- Do all types of connections exist in networks?
- Milo et al studied the transcriptional regulatory networks in yeast and E.Coli.
- Calculated all the three and four gene combinations possible and looked at their frequency

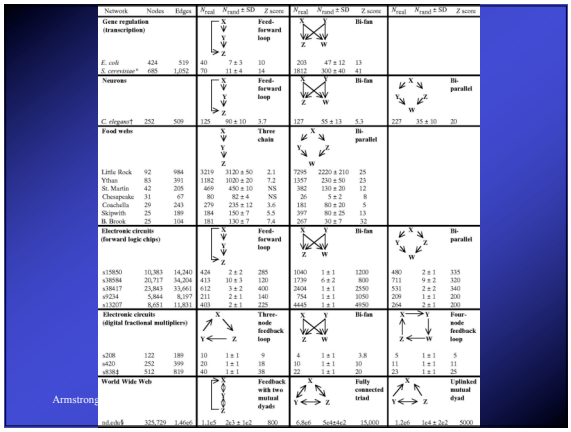
Milo et al. 2002 Network Motifs: Simple Building Blocks of Complex Networks. Science 298: 824-827



Gene sub networks

Network	Nodes	Edges	N_{real}	$N_{rand} \pm SD$	Z score	N_{real}	$N_{rand} \pm SD$	Z score
Gene regulation (transcription)					**Feed-forward loop**			**Bi-fan**
E. coli	424	519	40	7 ± 3	10	203	47 ± 12	13
*S. cerevisiae**	685	1,052	70	11 ± 4	14	1812	300 ± 40	41

Heavy bias in both yeast and E.coli towards these two sub network architectures

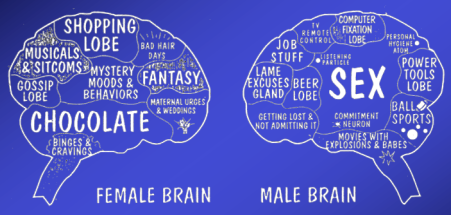


Gene networks and network inference

What is a gene network

- Genes do not act alone.
- Gene products interact with other genes
  - Inhibitors
  - Promoters
- The nature of genetic interactions in complex
  - Takes time
  - Can be binary, linear, stochastic etc
  - Can involve many different genes

What makes boys boys and girls girls?



Sugar, Spice and synthetic Oestrogens?

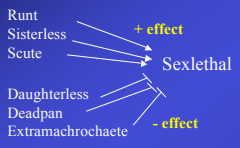
Sex determination: a gene cascade (in flies…)

6 Genes detect X:A ratio

	Females	Males
Runt Sisterless Scute	↑	↓
Daughterless Deadpan Extramachrochaete	↓	↑

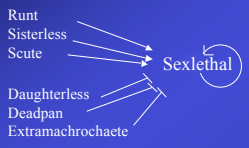
Sex determination
(in flies…)

6 Genes regulate 'Sexlethal'



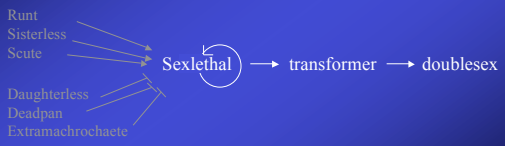
Sex determination
(in flies…)

Sexlethal can then regulate itself...

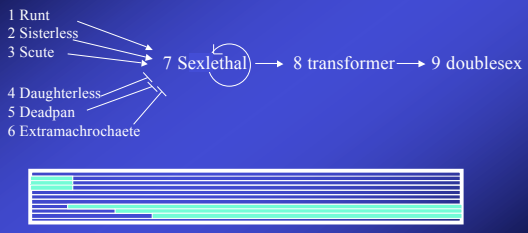


Sex determination
(in flies…)

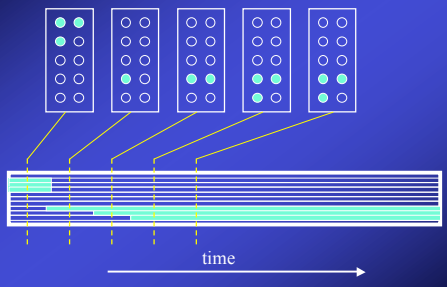
Downstream cascade builds...



Gene expression and time



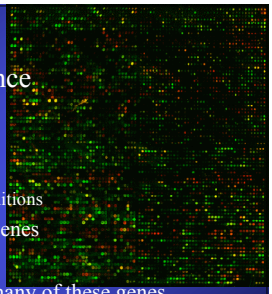
Gene microarrays



Gene Network Inference

- Gene micro-array data
- Learning from micro-array data
- Unsupervised Methods
- Supervised Methods
- Edinburgh Methods

Gene Network Inference



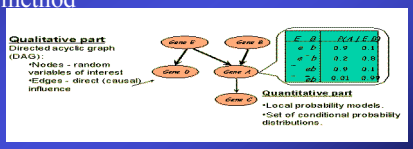
- Gene micro-array data
  - Time Series array data
  - Tests under ranges of conditions
- Unlike example - 1000s genes
- Lots of noise
- Clustering would group many of these genes together
- Aim: To infer as much of the network as possible

Learning from Gene arrays

- Big growth industry but difficult problem
- Initial attempts based on unsupervised methods:
  - Basic clustering analysis - related genes
  - Principal Component Analysis
  - Self Organising Maps
  - Bayesian Networks

Bayesian 'gene' networks

- Developed by Nir Friedman and Dana Pe'er
- Can be easily adapted to a supervised method



Learning Gene Networks

- The field is generally moving towards more supervised methods:
  - Bayesian networks can use priors
  - Support Vector machines
  - Neural Networks
  - Decision Trees

Can we combine network knowledge with gene inference?

- Scale free architecture
  - Chance of new edges is proportional to existing ones
  - Highly connected nodes may well be known to be lethal
- Network motifs
  - Constrain the types of sub networks
- Prior Knowledge
  - Many sub networks already known

Conclusions

- Gene network analysis is a big growth area
- Several promising fields starting to converge
  - Complex systems analysis
  - Using prior knowledge
  - Application of advance machine learning algorithms
  - AI approaches show promise