

Quick review

- The question of how scale-free topologies arise can be addressed by network assembly and evolution.
- Barabási-Albert (BA) model
 - Growth - random increase of number of nodes and edges over time
 - Preferential attachment - nodes with higher degree have greater probability of acquiring new edge
 - Degree distribution – power-law, constant clustering-degree function, small clustering co-efficient, smaller path lengths
- Other models – non-linear attachment, initial attractiveness, accelerated growth, aging, fitness, node loss

Biological network models

Approaches used to understand the mechanisms behind the evolution of cellular scale-free networks

- Growth/ preferential attachment models: duplication of genes, metabolic reactions, recombination, development
- Comparative genomics/ pathway comparison across phylogenetic tree
- Understanding principles from existing information e.g. probability of existing a link between nodes.
- Study evolution of motifs and cliques

Evolving protein interaction networks

Genes and interactions among genes/gene modules have been conserved through evolution.

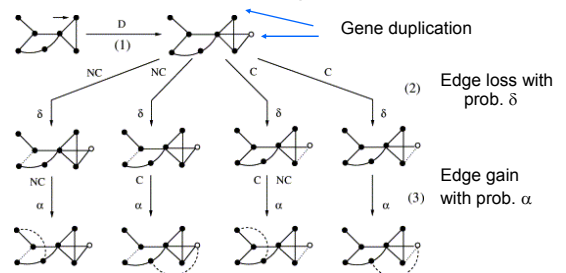
We can consider the topology of protein interaction networks as a result of a network evolution process.

One can formulate evolving network models for protein interaction networks.

Driving forces behind the formation of links:

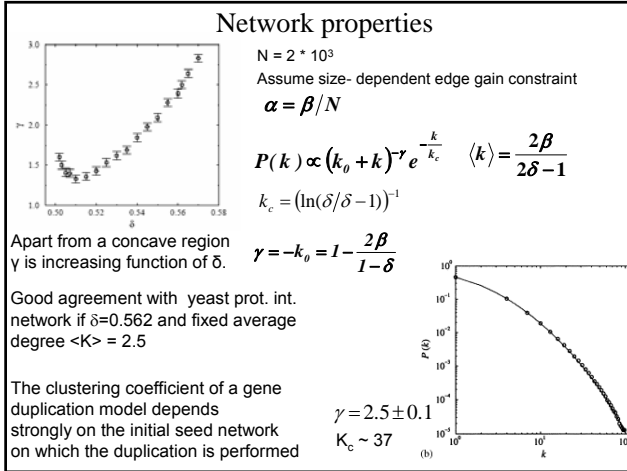
- gene duplication and mutation
- functional coupling
- protein attractiveness (affinity)
 - determined by protein (domain) structure

Duplication-divergence models



correlated connections (C): only the duplicated gene loses/ gains edge
 Uncorrelated connections (NC): edge can be added or removed between any pair of nodes in the network.

Pastor-Satorras et al., *Journ. Theor. Biol.* 22, 199 (2003)



A more detailed model of biological network evolution

Basic principles:

1. Every node has different identity based on its physical properties defining its type and an associated concentrations that changes in time.
 1. Instances are described by a bit string of size N .
 2. a predefined set of T types (nodes).

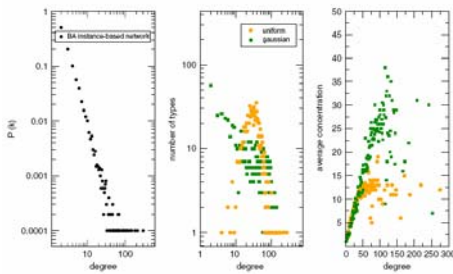
2. The network is type-based rather than instance based

3. Links are established based on mutual attractiveness (affinity).
 1. Links can be established only between nodes of different types
 2. Each type has an associated bit-string that decides whether the instances of some other type can connect to it.
 3. **Affinity** is defined as intersection of bit strings. Link is established if affinity is larger than threshold t .

Bersini et al., *Journ. Theor. Biol* 241, 488 (2006)

Type network vs. instance network

does the level (instance or type) of description change the properties of the network??



Scale-free instance based network transformed into type-based network.

Type-based model cont.

4. Addition of new node depends on the dynamics of the existing nodes.
 - Exogenous – New elements are added from outside
 - Endogenous – elements of the system produces/ leads to new instances.

5. Three approaches to node dynamics.
 - Random introduction of nodes from an outside source - new instances introduced uniformly (Similar with growth with no preferential attachment model).
 - clones of existing instances (similar to duplication - New instances are possibly mutated (bit-flipped) clones of existing instances.
 - collision - combination of two instances to make a third
 - New instances are combinations of $N/2$ bits of one existing instance with $N/2$ bits from another instance.

Ex. 1

Give a network representation of the reaction $2H + O \rightarrow H_2O$ and describe the dynamics of water molecule.

Ex. 2

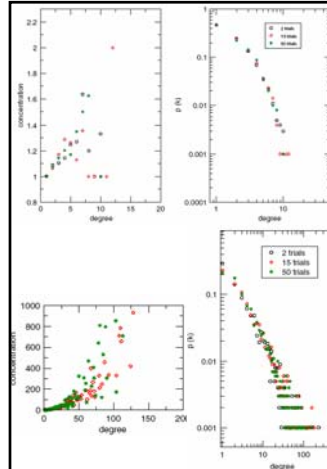
List the cellular processes or factors leading to the addition/removal of nodes/edges.

Ex. 3

Give examples of types of nodes in cellular networks

Ex. 4

In random introduction of nodes what kind of degree distribution do you expect?



Random growth: The degree distribution is an exponential
Clustering coefficient $\sim 10^{-5}$
Disassortative

Cloning:

High-concentration types have a higher chance of acquiring new instances - preferential attachment.

The degree distribution is close to power-law for low trial number.
Maximum degree and clustering coefficient decrease with trial size.
Strong disassortativity.

Variant with below-threshold affinity can produce higher clustering coefficient, decreasing $C(k)$.

Collision model: New instances are combinations of $N/2$ bits of one existing instance with $N/2$ bits from another instance.

The degree distribution is close to power-law for low trial number.

Higher trial number increases degree exponent, decreases assortativity.
For both cloning and collision model the degree and concentration of types are strongly correlated – abundant gets richer.

Combination with random growth leads to increase of degree exponent, approach to exponential degree dist.

Attractiveness threshold: Low attractiveness threshold – more possibilities to connect.

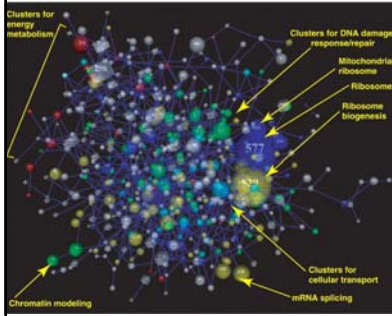
Selection– uniform random in random model; cloned with mutation in cloning model, between two “parent” thresholds in the collision model.

Fit-gets-rich scenario competes with abundant-gets-rich scenario.

Comparison of networks (nodes/orthologs, topology) between different species.

- Assembly of cellular networks:
 - New network can be formed by integrating the information from the literature e.g. co-expression, co-citation, etc.
 - Networks can also be formed by using and scoring the interactions from various databases.
- To ensure the method of network integration, usually new interactions or components (depicted by nodes) are detected and are validated.
- Then the network is used to analyze functionally conserved genes and interactions.
 - Graph theoretical measures
 - Interacting motifs

Functional assessment of yeast genes



Inter-module linkages
 Colors – modules
 Sizes proportional to number of genes.
 Connections – inversely proportional to the fraction of genes linking the clusters.

- Data used: Co-expression, gene fusion, phylogenetic profiles, co-citations, protein interactions.
- Ability to reconstruct the known gene pathways and systems is measured by calculating likelihood of functional linkages

log likelihood score:
 linkages Exp.

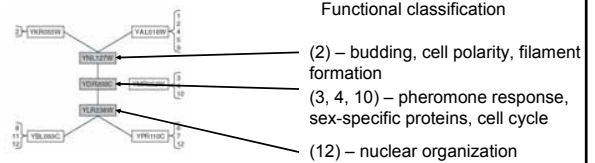
$$LLS = \ln \left(\frac{P(L|E)/\sim P(L|E)}{P(L)/\sim P(L)} \right)$$

Prior expectation

Functionality of proteins

- Approaches to predict functions – clustering of co-regulated genes, phylogenetic profiles, protein complexes.
- Assignment of functional classes on the basis of their network of physical interactions.

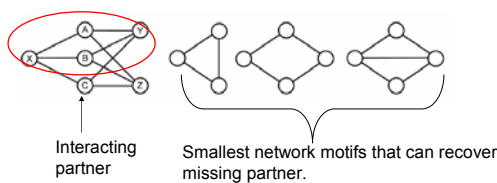
Subgraph of protein interaction network.



Functional classification

Lee, et. al. 2004 Science (306) 1555-58
 Vazquez, et al. 2003 Nat. Biotech. (21) 697-700

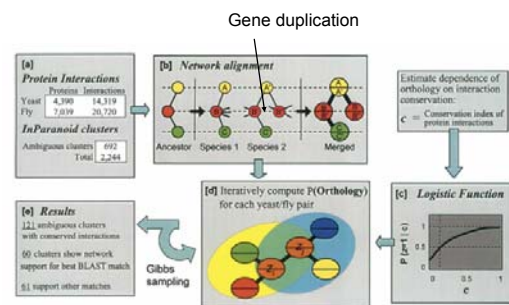
Prediction of interacting partner based on conserved network motifs



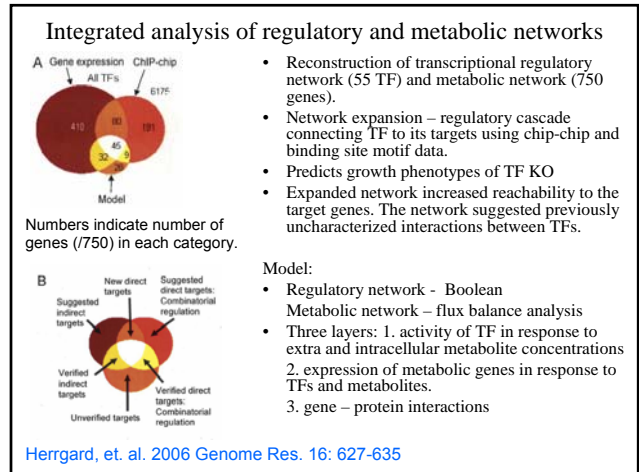
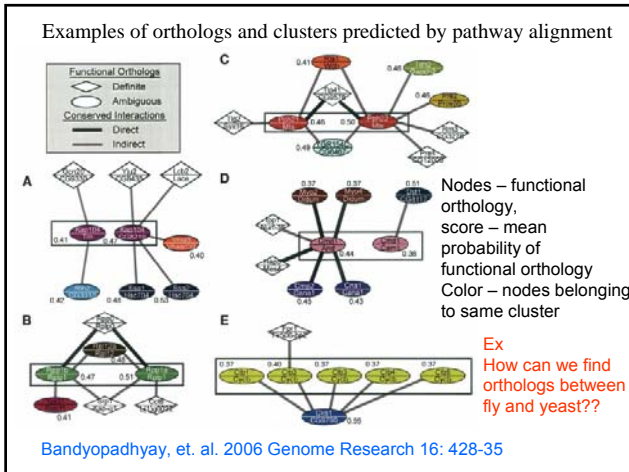
- The method used in *S. cerevisiae* protein-protein interaction map.
- Similar method can be used in aligned pathways.

Albert et al. 2004 Bioinformatics (18) 3346-52

Network alignment of yeast and fly PPI to find functional orthologs



Only conserved interactions are used.
 Logistic function is used to calculate the probability of functional orthology.
 The network supports different hypothesis for 61 interactions.



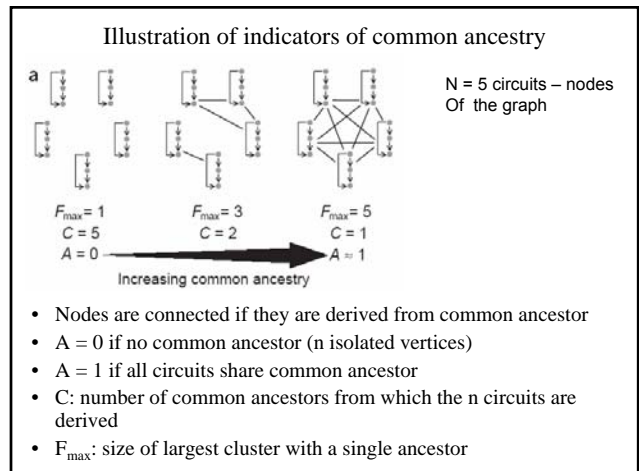
Evolutionary origins of cellular network motifs

- The motifs may have evolved through random duplication and subsequent diversification of ancestral circuits
- Convergent evolution - These motifs can arise from unrelated genes

Ex:
 What does convergent evolution indicate?

Indicators of evolutionary origin of circuits:

- Common circuit ancestry:
 $A = 1 - (C/n)$
 C = number of connected components in the graph
 n = number of motifs
 The greater A , the greater is the fraction of circuits sharing common ancestry.
- F_{max} = size of largest family of circuits with common ancestry.
 Conant, et. al. 2003 Nat. Genet. 34: 264-266



Optimal circuit design

- Comparison between motifs in *S. cerevisiae* and *E. coli*.
 - *E. coli* – Bi-fans, feed forward loops
 - *S. cerevisiae* – Bi-fans, feed forward loops, multi-input motifs, regulatory chains.

- No evidence for common ancestry of *E. coli* motifs.
- No common ancestry of yeast regulatory chains and multi-input motifs for more than 2 regulators.
- Yeast Bi-fan and smaller MIM show common ancestry.

b

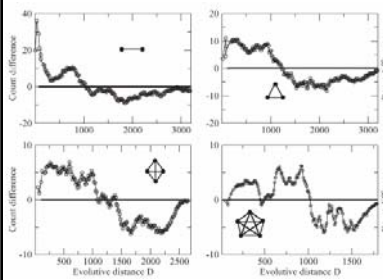
	Circuit type	Number of circuits	Number of families (C)	Index of common ancestry (A)	Largest circuit family (F_{max})
Yeast	Bi-fan	48	44 (46.8 ± 3.9; $P = 0.09$)	0.082 (0.023 ± 0.035; $P = 0.02$)	5 (1.9 ± 1.4; $P = 0.32$)
	Feedforward	542	435 (469.0 ± 37.7; $P = 0.18$)	0.197 (0.135 ± 0.070; $P = 0.18$)	49 (41.0 ± 31.1; $P = 0.33$)
	MIM-2	176	168 (164.5 ± 8.8; $P = 0.60$)	0.045 (0.065 ± 0.050; $P = 0.60$)	5 (7.4 ± 6.2; $P = 0.59$)
<i>E. coli</i>	Bi-fan	33	33	0	1
	Feedforward	11	11	0	1
	Bi-fan	27	27	0	1

(Yeast data was further analyzed for Statistical significance)

Co-operative evolution

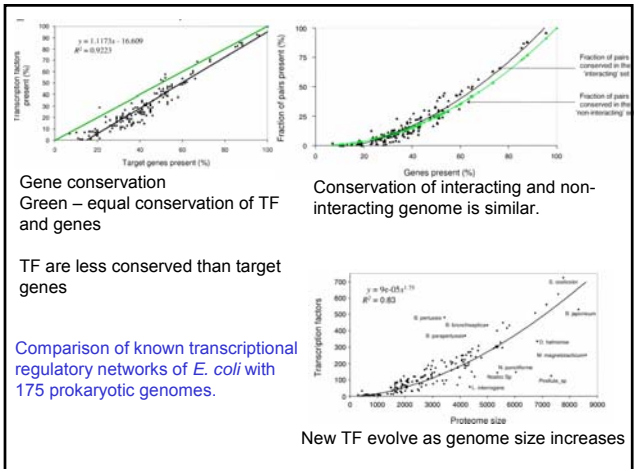
- Frequent concern – comparative analysis is performed with distant relatives of *S. cerevisiae*
- Comparative analysis of four species of the class hemiascomycetes.
- They share functional similarities.
- Data sets – yeast two hybrid, TAP tag data
- Hypothesis: proteins in cliques evolve together suggesting that compensatory mutations take place.
- Evolutionary divergence rates are calculated from the ranks indicating sequence similarity of *S. cerevisiae* proteins participating in the cliques with the other species.

Co-operative co-evolution



- Cliques up to 5 proteins
- Comparison between *S. cerevisiae* and *K. lactis*.
- Curves- difference between the histograms in the real and randomized networks.
- Low values of D – strong co-evolution patterns.
- Note the higher numbers of events with small D.

Vergassola, et. al. 2005 Proteomics 5: 3116-19



Gene conservation
Green – equal conservation of TF and genes

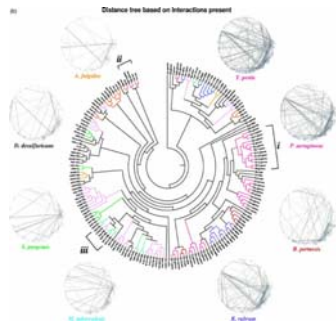
Conservation of interacting and non-interacting genome is similar.

TF are less conserved than target genes

Comparison of known transcriptional regulatory networks of *E. coli* with 175 prokaryotic genomes.

New TF evolve as genome size increases

Conservation of regulatory interactions in different genomes



- i – genomes in same phylogenetic group
- ii – parasite genomes
- iii – genomes with similar life-style but different phylogenetic group

Target genes are embedded in motifs

Different TF evolve independently as hubs – convergent evolution

Madan Babu, et. al. 2006 J. Mol. Bio. 358: 614-633