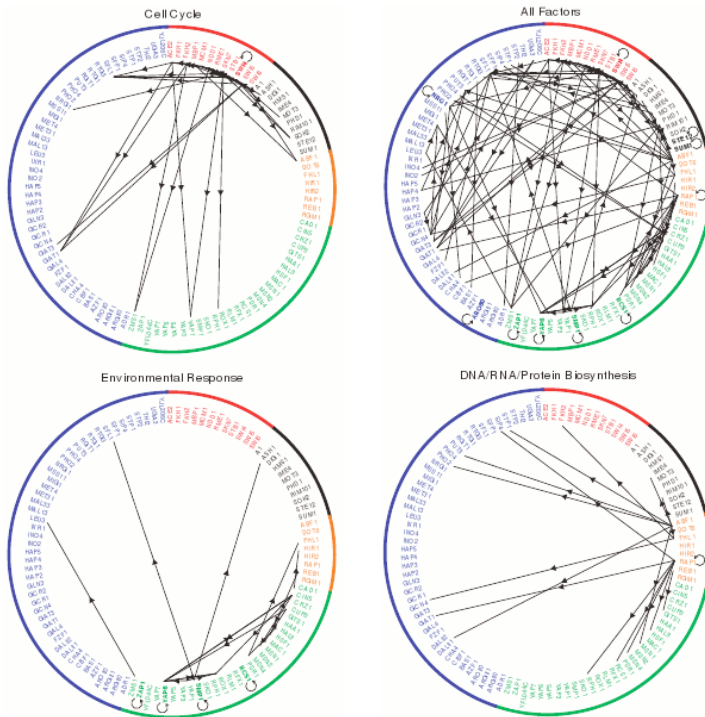


Network Evolution

in the light of biology



A brief and biased exploration by

Daniel Daminieli

António Miranda

José Lourenço

Network Organization

Patterns:

- Motifs
- Modules
- Topology

Network Evolution

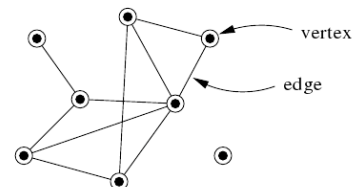
Concept of network

A network can be defined having:

- a set of items: generally called nodes or vertices
- a set of connections: generally called edges or arcs

A huge range of systems take the form of networks:

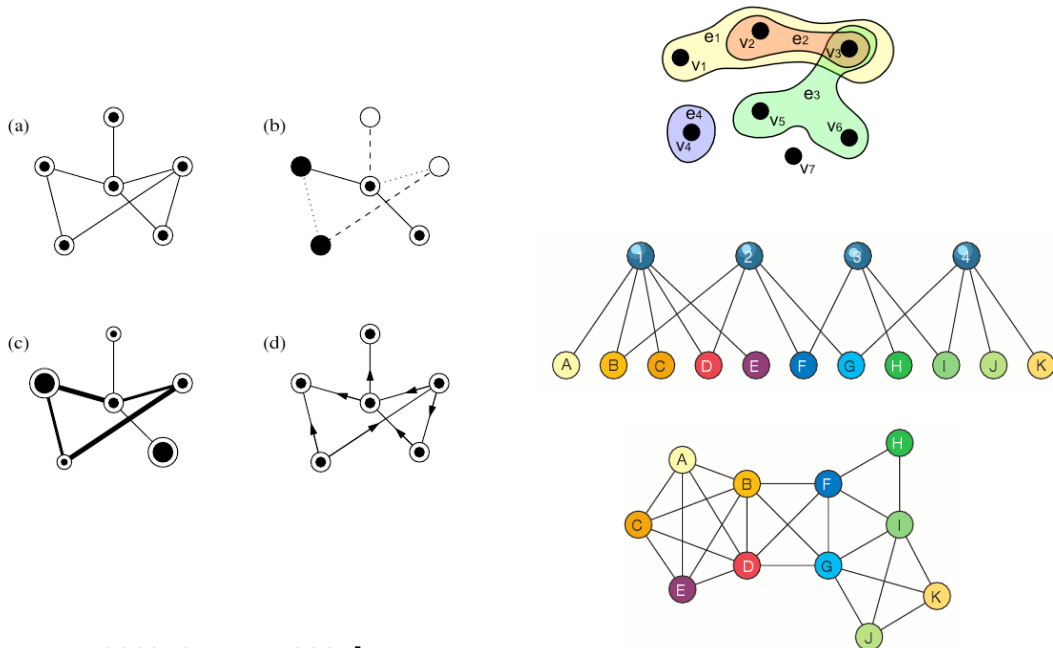
- World Wide Web (WWW)
- Social contacts between humans
- Business systems
- Gene co-expression
- Neural networks
- Human transportation system
- Postal delivery routes
- Citations between papers
- Blood vessels organization
- Metabolic pathways
- Food chains, Etc



Networks are also called Graphs in the mathematical literature.
[Newman 2003]

Representation of network

A set of nodes joined by edges is only the simplest type of network.



[Newman 2003; Strogatz 2001]

Decay of random networks

In the 19th century:

Vilfredo Pareto introduced principle known as the Law of the Vital Few, stating that “for many events, roughly 80% of the effects come from 20% of the causes”

During the 20th century:

Many examples of such events started to be discovered

For ex.: cities by size, publications by scientists, word frequency in texts

Last years of 20th century:

A few publications in major science journals reported the discovery that the WWW followed the Law of the Vital Few principle. Reports on systems that were structured in networks following this principle rapidly invaded the literature.

[Keller 2005, Barabási 1999]

Classes of networks

Empirical and theoretical results show that complex networks can be divided into 2 major classes based on their connectivity distribution $P(k)$

Exponential:

Where $P(k)$ peaks at an average $\langle k \rangle$ and decays exponentially for large k .

Leading to a fairly homogeneous network with each node having the number of connections $\sim \langle k \rangle$

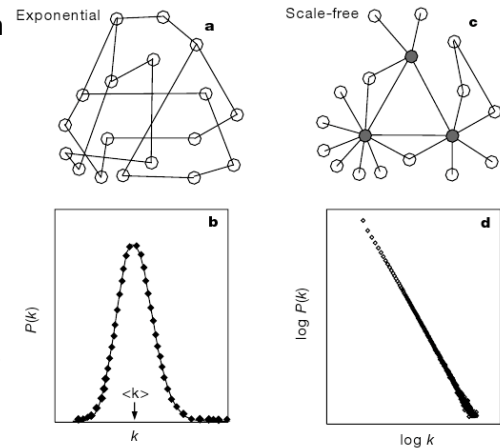
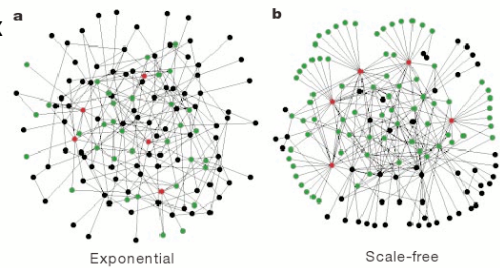
Includes Erdős random graph model and Watts & Strogatz small-world networks.

Scale-free:

Where $P(k)$ decays as a power-law.

Leading to an heterogeneous network where highly connected nodes are statistically significant.

[Keller 2003, Barabási 1998, 2000]
Includes H. Simon model and Barabási's scale-free networks



Scale-free networks

The **interconnectedness of a network (diameter d)** is defined as the average length of the shortest paths between any two nodes in the network.

d characterizes the ability of two nodes to communicate with each other.

The malfunctioning of a node (failure) in general increases the distance between the remaining nodes.

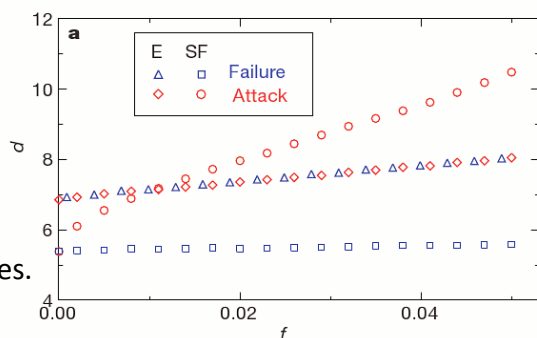
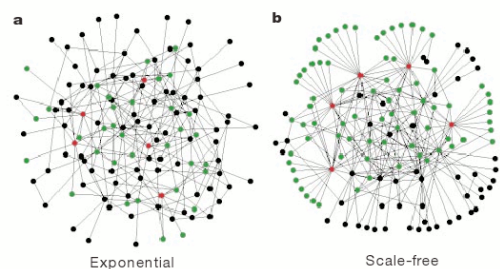
The malfunctioning of a set of nodes (attack) in general disrupts the network structure.

Owing to homogeneity, in exponential networks there is no difference between attacks and failures.

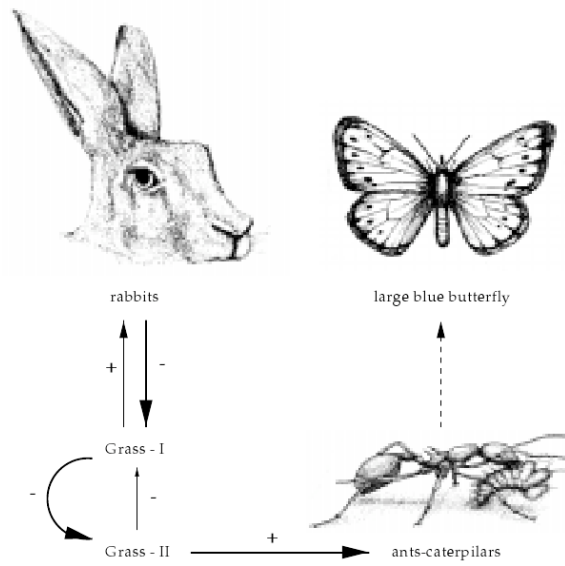
In contrast, due to heterogeneity, scale-free networks are insensible to failures but highly sensible to attacks.

[Keller 2003, Barabási 1998, 2000]

What is the network class that best fits each biological system is a question under great debate

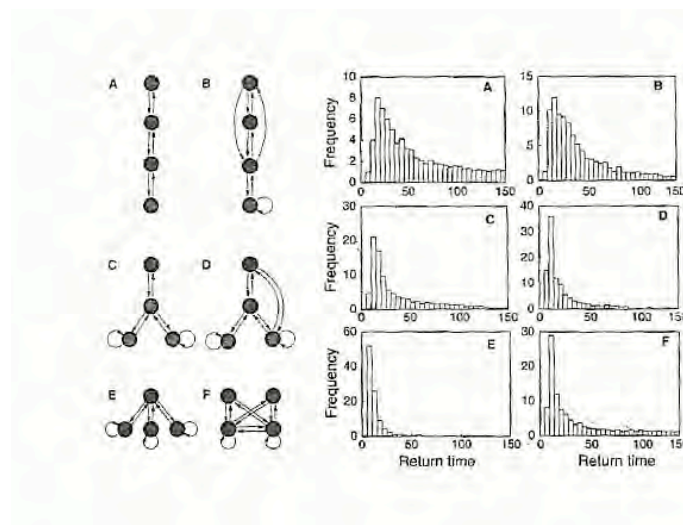


Ecology



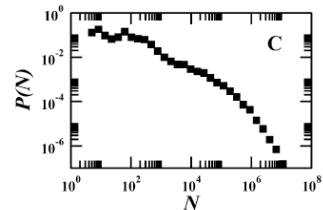
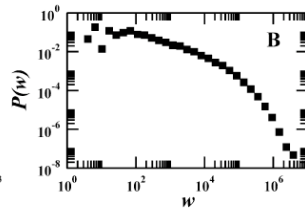
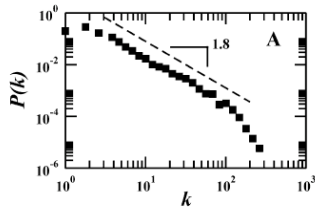
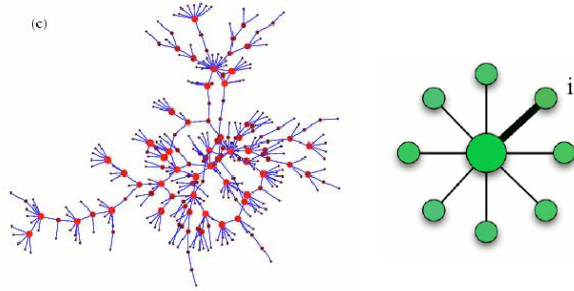
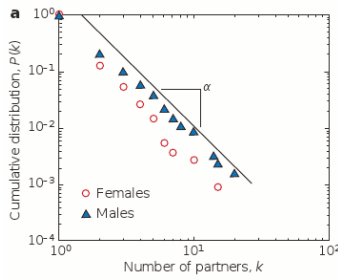
[Solé 2000]

Ecology



[Solé, Jordi 2006]

Social & Epidemiology



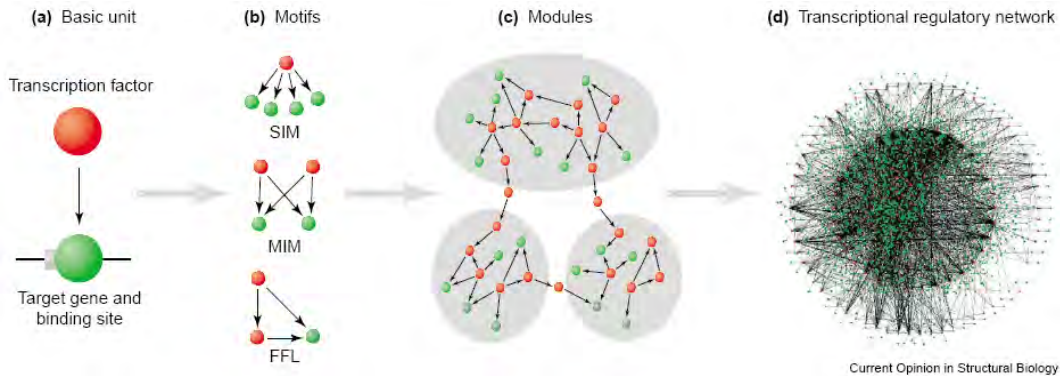
number of connections of each airport

number of passengers travelling on a given connection between a pair of airports

population size of the urban area surrounding each airport

[Liljeros 2001, Newman 2003, Colizza 2007]

Patterns



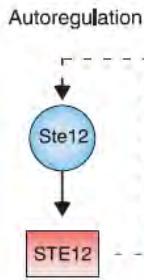
(a) Transcription factor, its target gene with DNA recognition site and the regulatory interaction between them

(b) Recurrent patterns of interconnections called network motifs, which appear frequently throughout the network

(c) Motifs cluster into semi-independent transcriptional units called modules

(d) The entire assembly of regulatory interactions constitutes the 'transcriptional regulatory network'

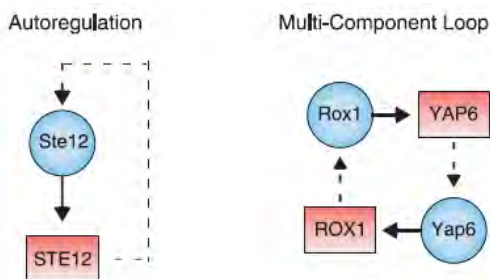
Motifs: Regulatory Networks



Several selective growth advantages, including reduced response time to environmental stimuli, decreased biosynthetic cost of regulation, and increased stability of gene expression

Lee et al (2003)

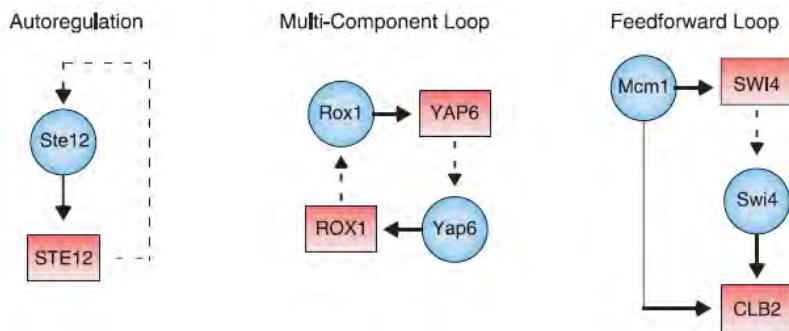
Motifs: Regulatory Networks



The closed-loop structure provides the capacity for feedback control and offers the potential to produce bistable systems that can switch between two alternative states

Lee et al (2003)

Motifs: Regulatory Networks

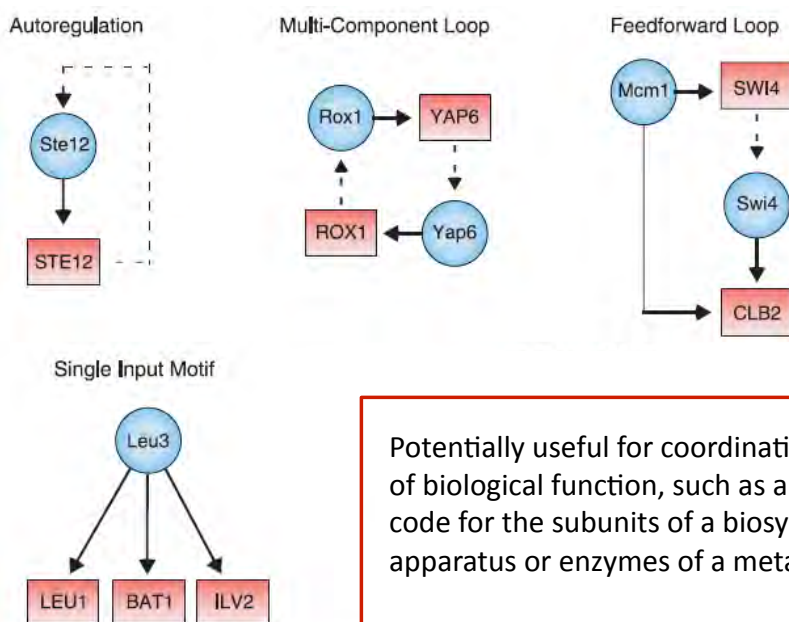


Feedforward loop can provide several features to a regulatory circuit:

- Act as a switch that is designed to be sensitive to sustained rather than transient inputs;
- Provide temporal control of a process;
- Provide a form of multistep ultrasensitivity.

Lee et al (2003)

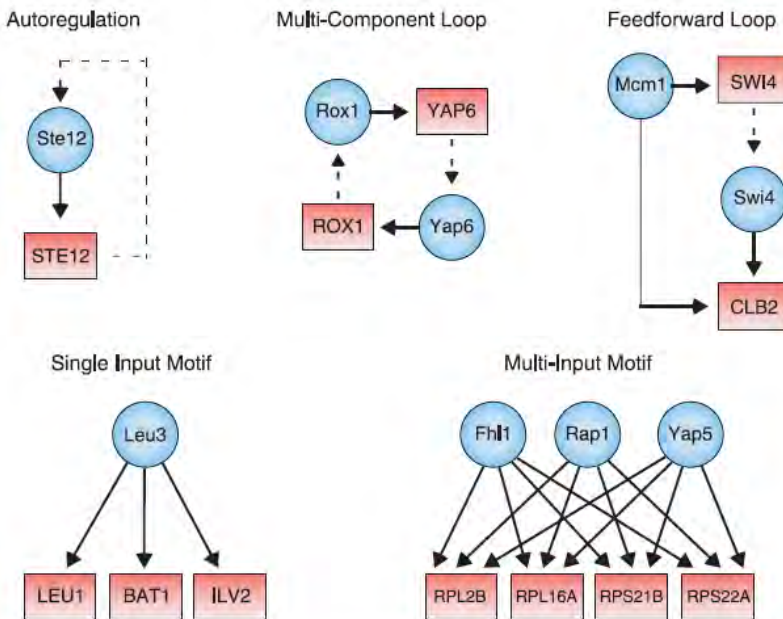
Motifs: Regulatory Networks



Potentially useful for coordinating a discrete unit of biological function, such as a set of genes that code for the subunits of a biosynthetic apparatus or enzymes of a metabolic pathway

Lee et al (2003)

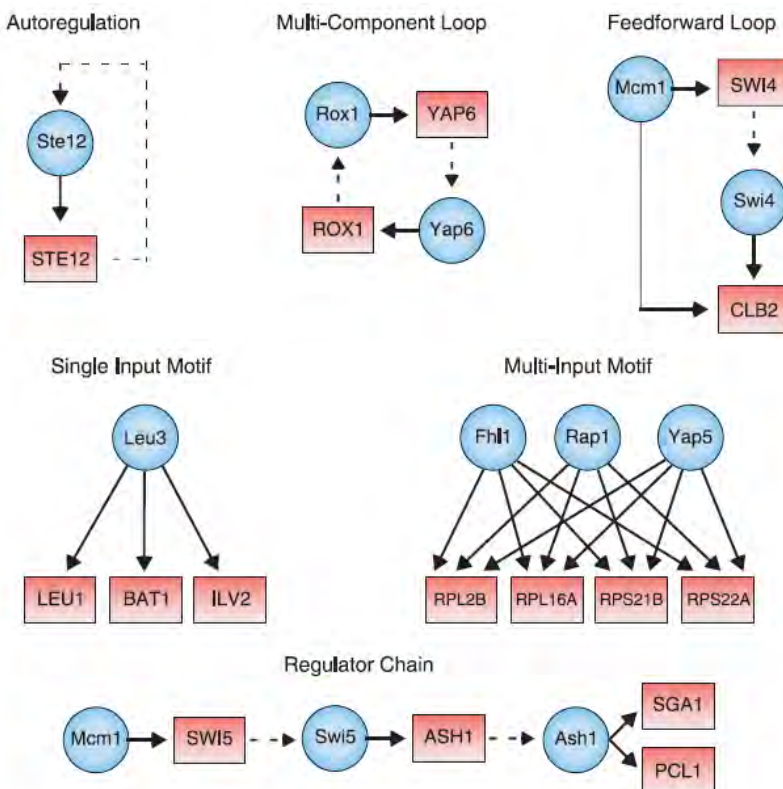
Motifs: Regulatory Networks



This motif offers the potential for coordinating gene expression across a wide variety of growth conditions

Lee et al (2003)

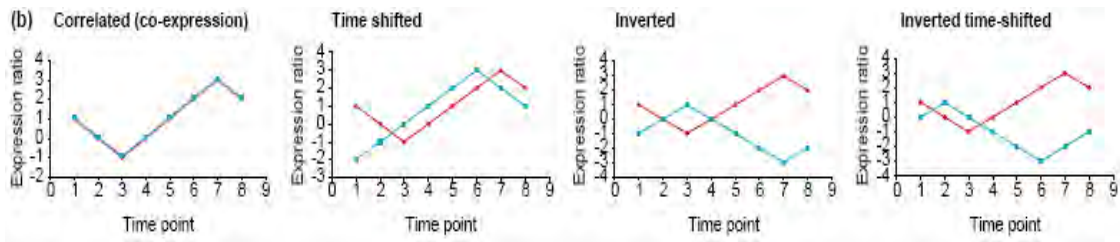
Motifs: Regulatory Networks



The chain represents the simplest circuit logic for ordering transcriptional events in a temporal sequence

Lee et al (2003)

Motifs: Regulatory Networks



Gene expression relationships:

Correlated (genes have similar profiles)

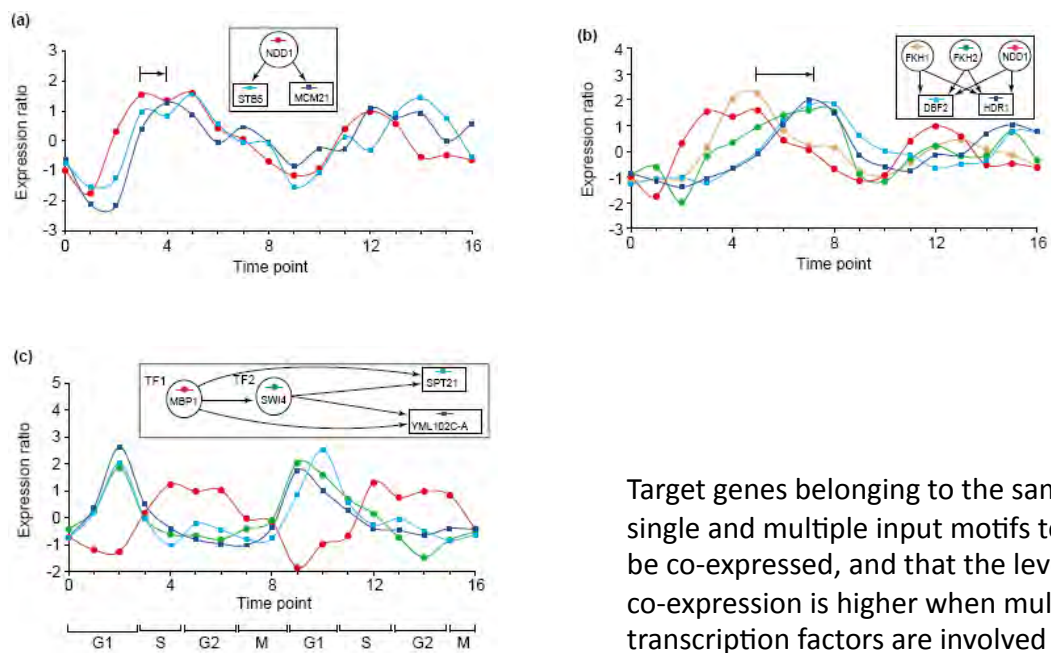
Time-shifted (one gene is delayed with respect to the other)

Inverted (genes have opposing profiles)

Inverted time-shifted

Yu et al (2003)

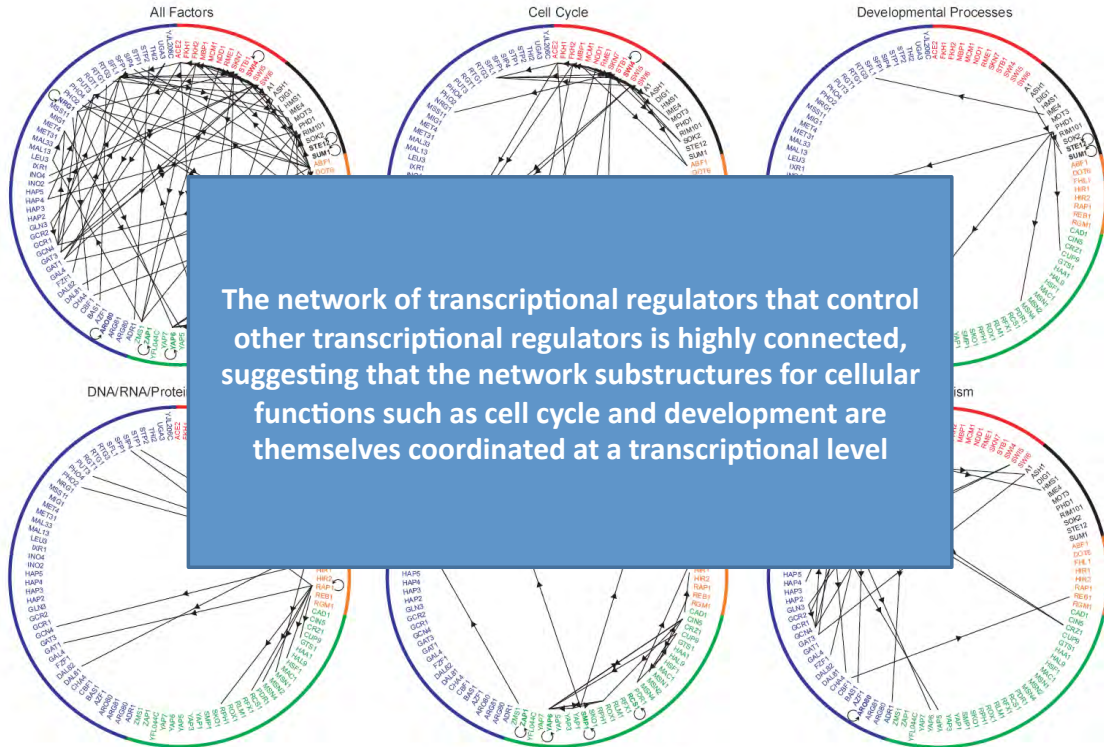
Motifs: Regulatory Networks



Target genes belonging to the same single and multiple input motifs tend to be co-expressed, and that the level of co-expression is higher when multiple transcription factors are involved

Yu et al (2003)

Motifs: Regulatory Networks

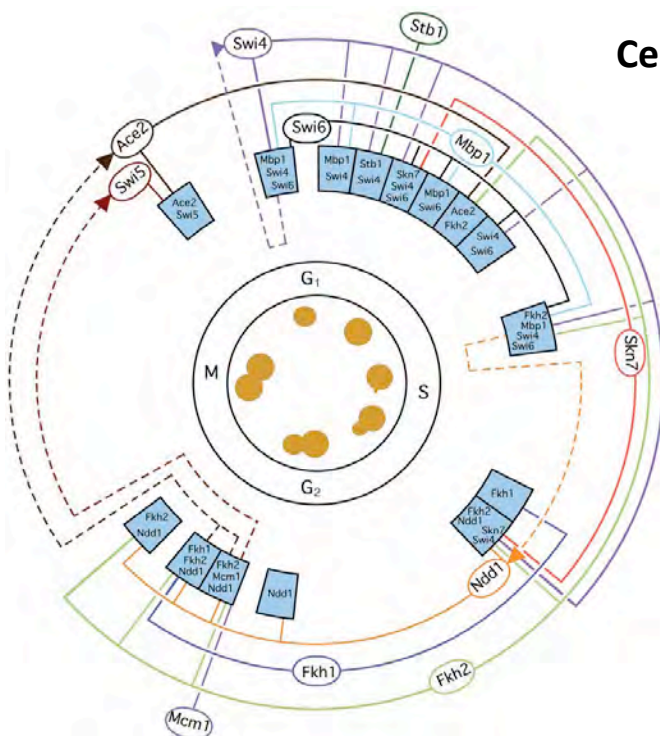


■ Cell Cycle ■ Developmental Processes ■ DNA/RNA/Protein Biosynthesis ■ Environmental Response ■ Metabolism

Lee et al (2003)

Modules: Regulatory Networks

Cell Cycle Regulatory Network

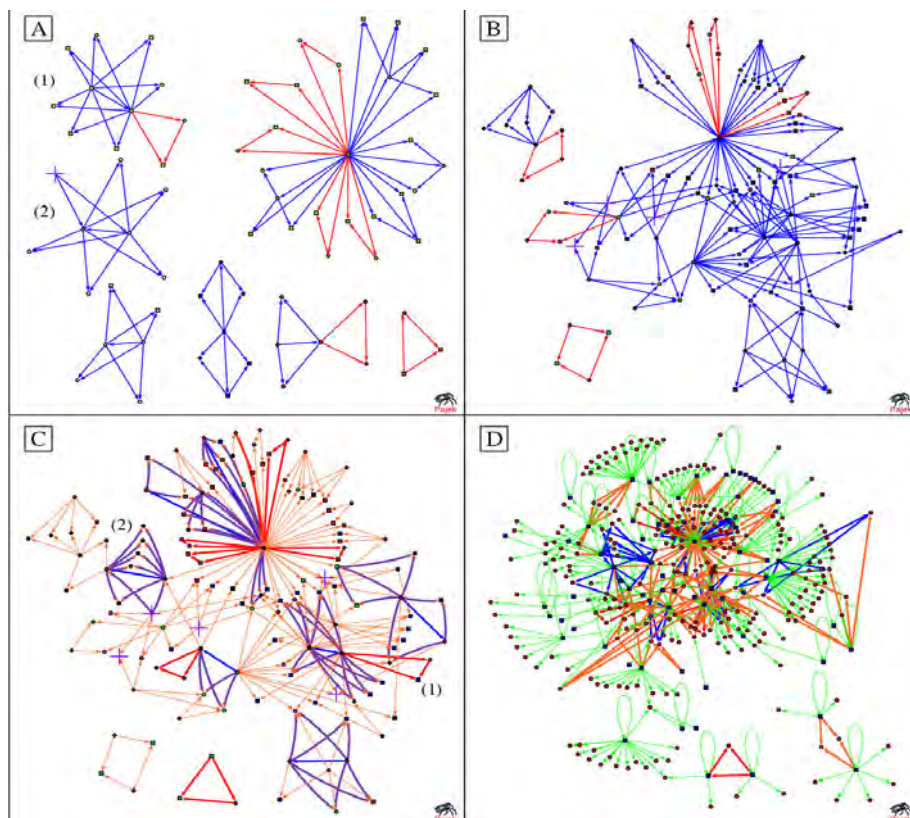


- Binding + Expression data

Reconstruction of the regulatory architecture was automatic and required no prior knowledge of the regulators that control transcription during the cell cycle

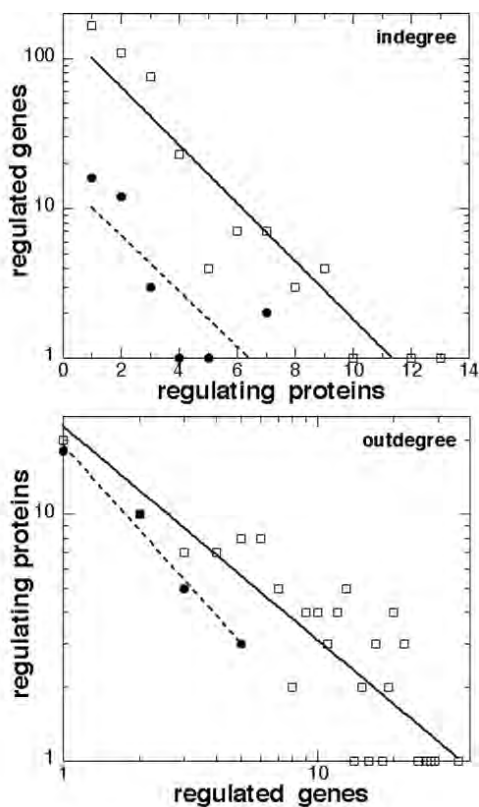
Lee et al (2003)

Modules: Regulatory Networks



Dobrin et al (2004)

Topology: Regulatory Networks

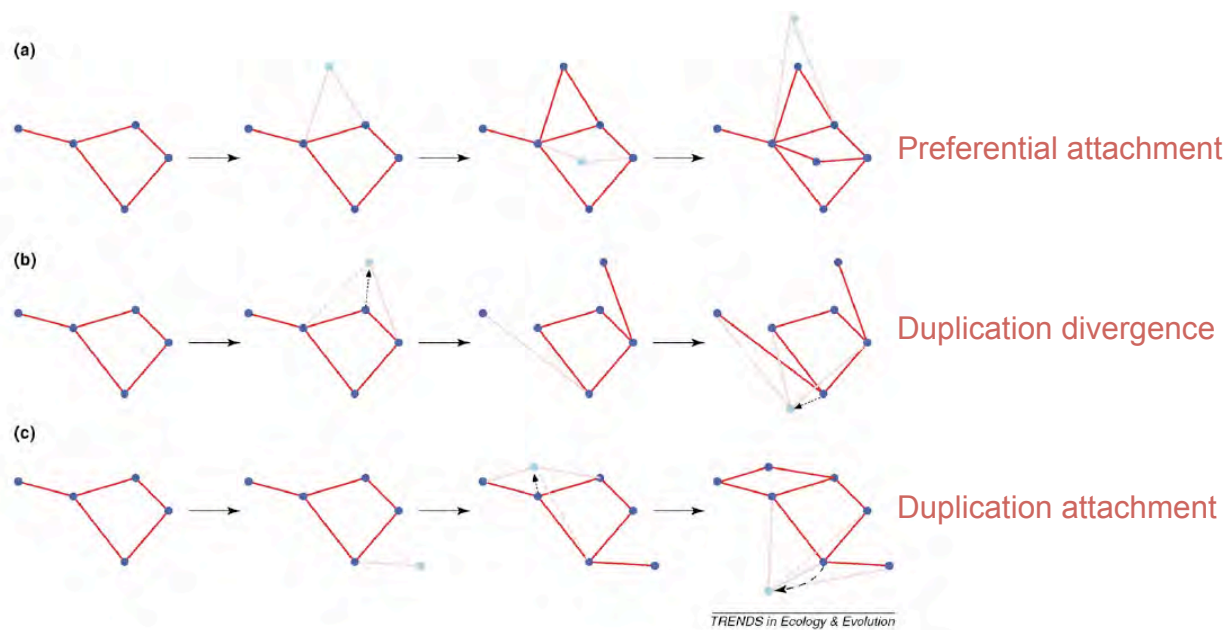


Incoming connectivity: Yeast network follows an exponential distribution, with 93% of the genes being regulated by 1–4 regulating proteins

Outgoing connectivity: is distributed according to a power law, this is indicative of a hub-containing network structure, in which a select few transcription factors participate in the regulation of a disproportionately large number of target genes

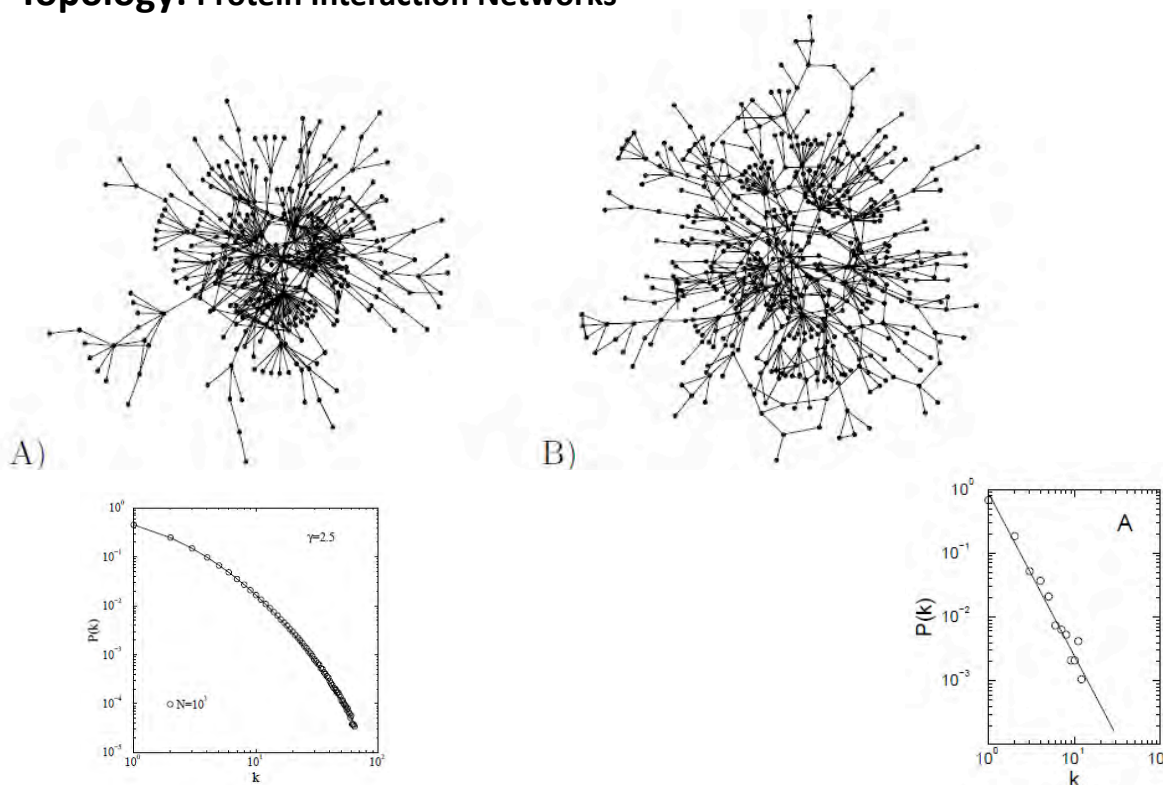
Guelzim et al (2002)

Topology: Network growth models



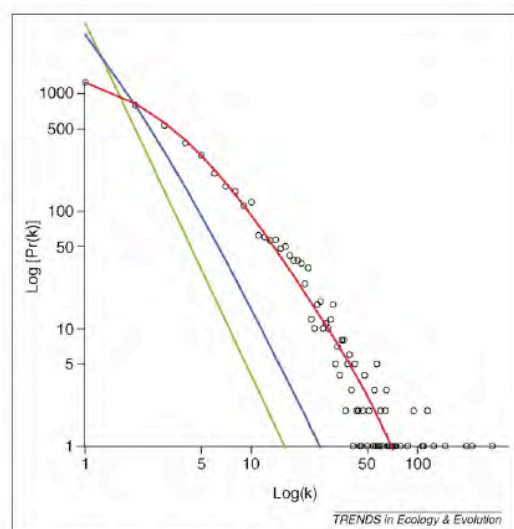
Stumpf et al. 2007 TRENDS in Ecology and Evolution

Topology: Protein Interaction Networks



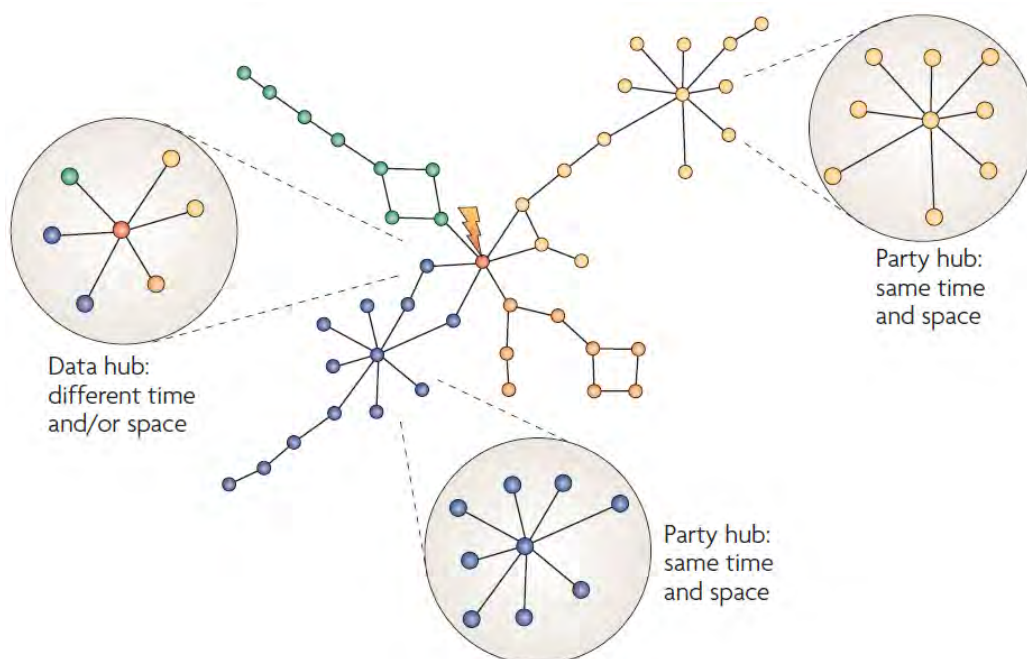
Solé et al. 2008 Advances in Complex Systems

Topology: Protein Interaction Networks



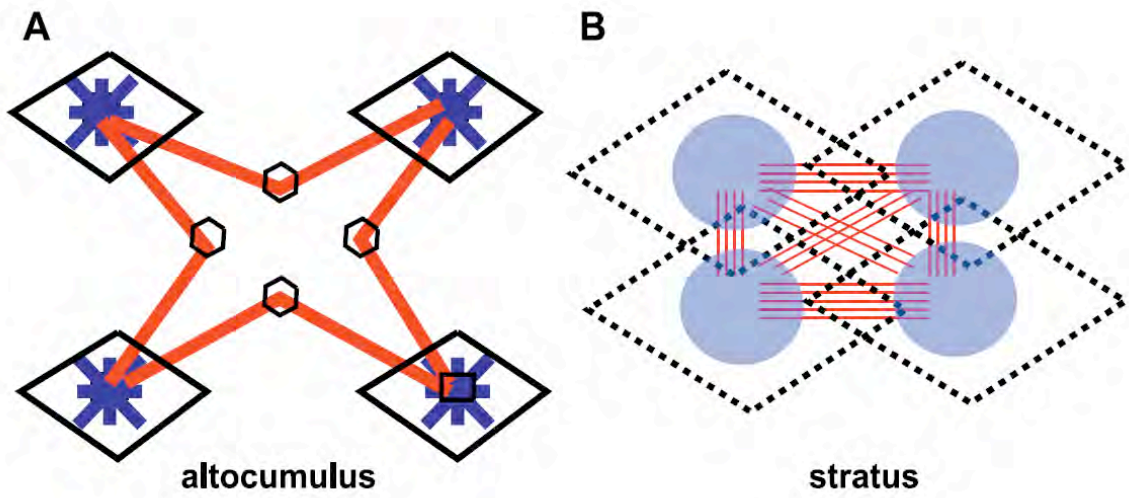
Stumpf et al. 2007 *TRENDS in Ecology and Evolution*

Topology: Hubs



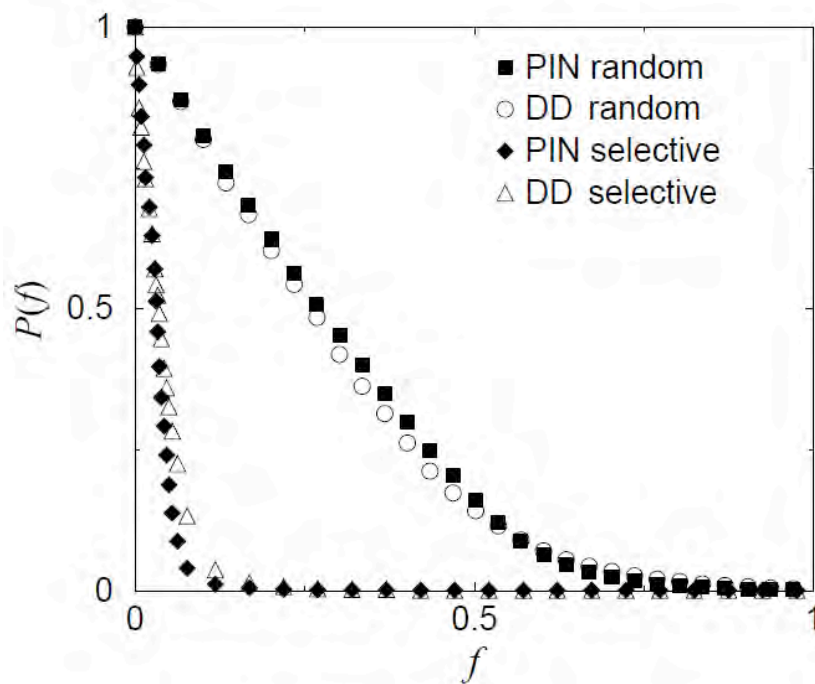
Wagner et al. 2007 *Nature Genetics*

Topology: Hubs & Modules



Batada et al. 2006 PLOS Biology

Robustness: Attack tolerance



Vazquez et al. 2001 Cond-Math. Stat. Mech.

Comparative Biology & Networks

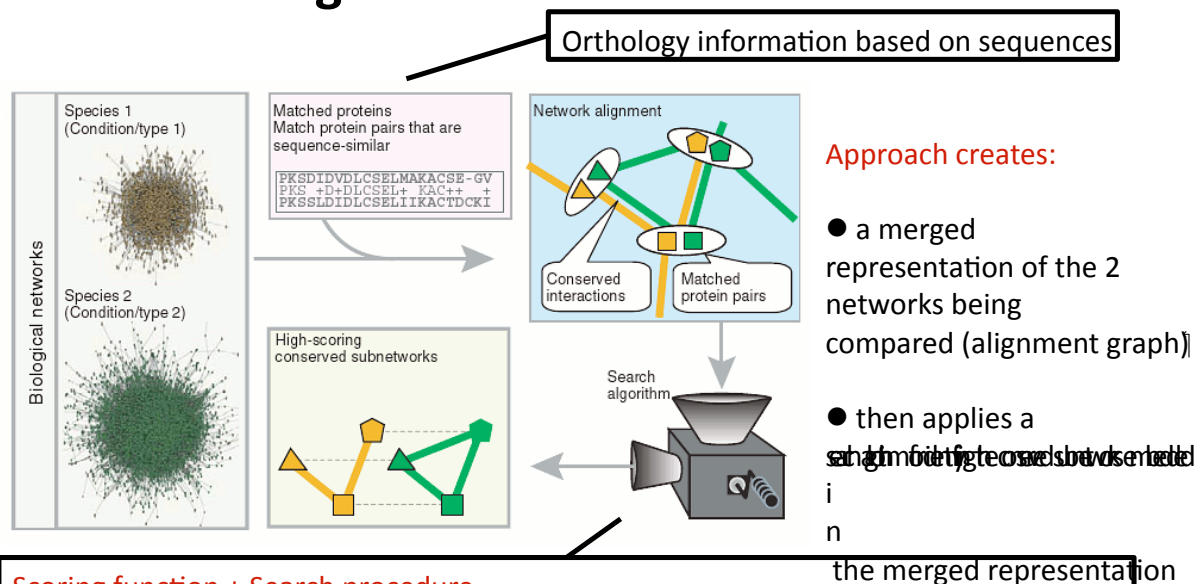
Molecular networks represent the backbone of molecular activity within the cell.

Comparison of networks aims to answer fundamental biological questions. Such questions motivated the existence of 3 types of methods:

Mode	Common application	Main goals
Alignment	At least two networks of the same type across species	Identification of functional (conserved) protein modules; study of network evolution; interaction prediction
Integration	At least two networks of different types for the same species	Identification of modules (supported by several networks); study of interrelations between data types; interaction prediction
Querying	Subnetwork module versus a network	Identification of duplicated/conserved instances of the module; knowledge transfer

[Sharan 2006]

Network alignment



Scoring function + Search procedure

Scoring function ex.: likelihood (fit of substructure versus if 1 subnetwork was random)

Search procedure ex.: greedy algorithm with local search based on iterative modifications (additions and deletions) that maximizes the score

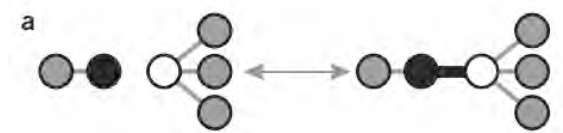
[Sharan 2006]

Network evolution status

Understanding how networks evolve is a fundamental issue, which affects each of the above methods of comparative biology.

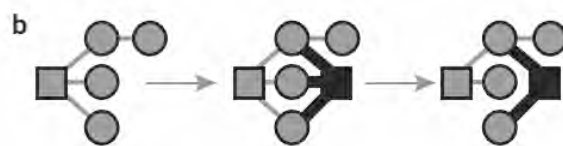
Two main processes have been invoked and highly defended in the literature:

Sequence mutations in a gene



- Easily observable in sequences

Gene duplication



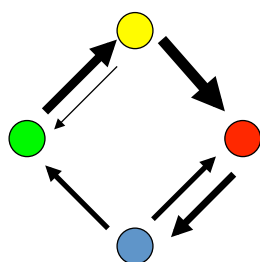
- In accordance with the growth model that predicts that molecules that appeared early in the network are the most connected (as in scale-free)
- Hard to observe since duplication is followed by differentiation

[Sharan 2006]

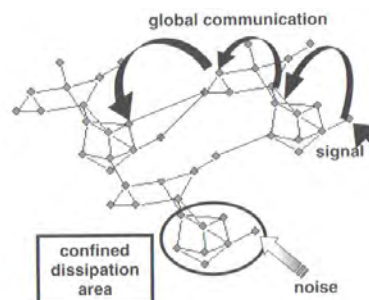
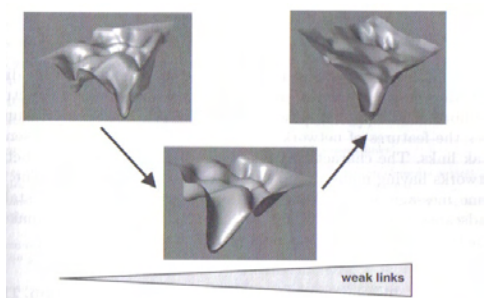
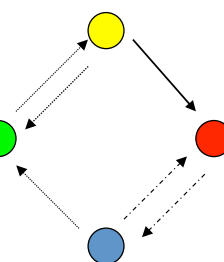
Robustness & stability:

Network connectance, weak links, modularity...

$\uparrow C_g$

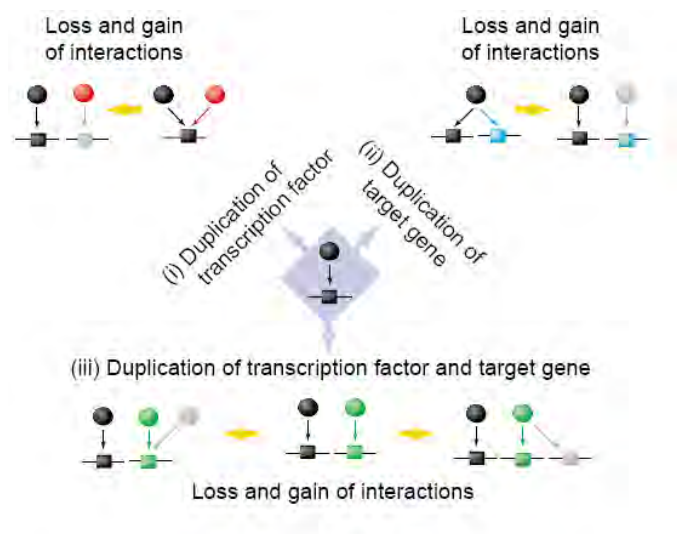


$\downarrow C_g$



Effects of Gene Duplication

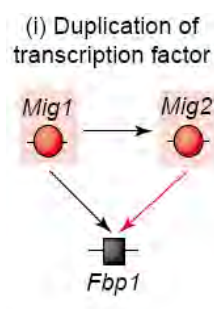
- (i) Both copies regulating the same gene
- (ii) Both copies being regulated by the same trans. factor



Babu et al (2004)

Effects of Gene Duplication

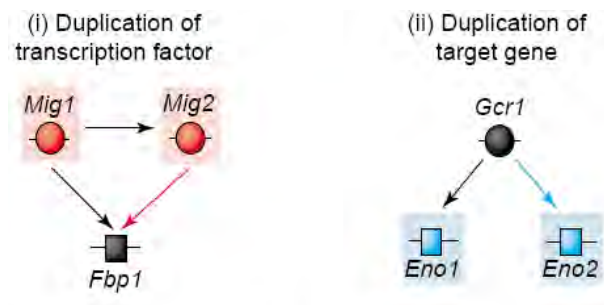
- (i) *Mig1* and *Mig2* are repressors of glucose metabolism, both recognise the same DNA binding sites, suggesting that there is redundancy in their regulatory roles



Babu et al (2004)

Effects of Gene Duplication

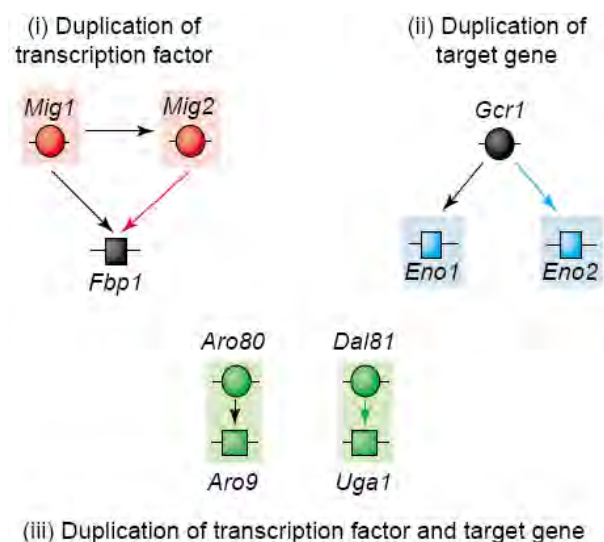
(ii) *Gcr1* is a global regulator that controls the expression of two Genes. Both targets probably inherited the regulator during duplication



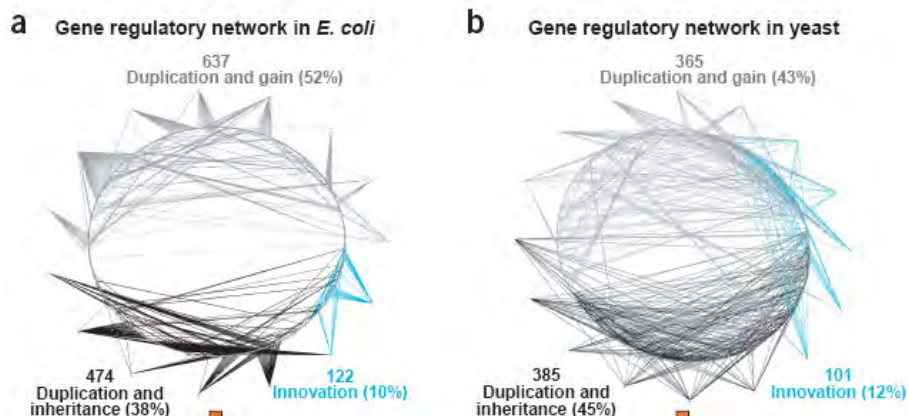
Babu et al (2004)

Effects of Gene Duplication

(iii) *Aro80* and *Dal81* are transcription factors that are homologous to each other, as are their respective targets, *Aro9* and *Uga1*. Homologous pairs of interactions such as these may have evolved by duplication of a single chromosomal region encoding both the transcription factor and target

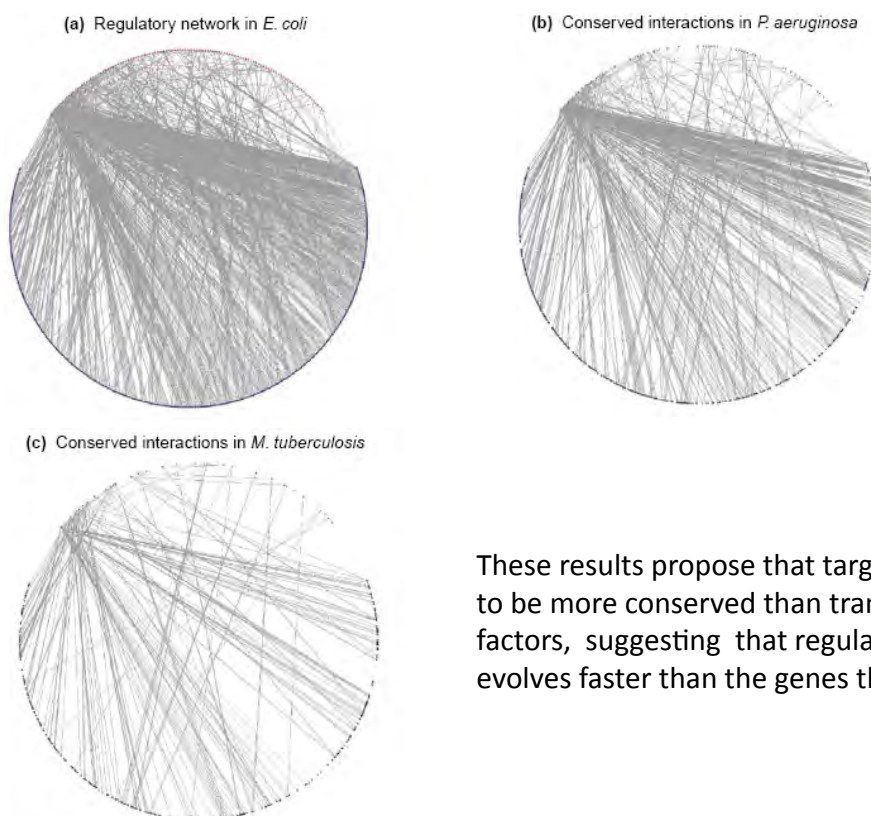


Babu et al (2004)



Gene duplication played a major role in the evolution of the regulatory network and that about 45% of regulatory interactions by duplication with inheritance of interaction

Teichmann et al (2004)



These results propose that target genes tend to be more conserved than transcription factors, suggesting that regulation of genes evolves faster than the genes themselves

Babu et al (2004)

Non-adaptive processes

as generative forces

Emergence and maintenance of:

Complexity

Evolvability

Genome architecture

Motifs

Modularity

Redundancy

Robustness



Unjustified summoning of natural selection and adaptation

Absence of null model: How would neutral evolution take place?

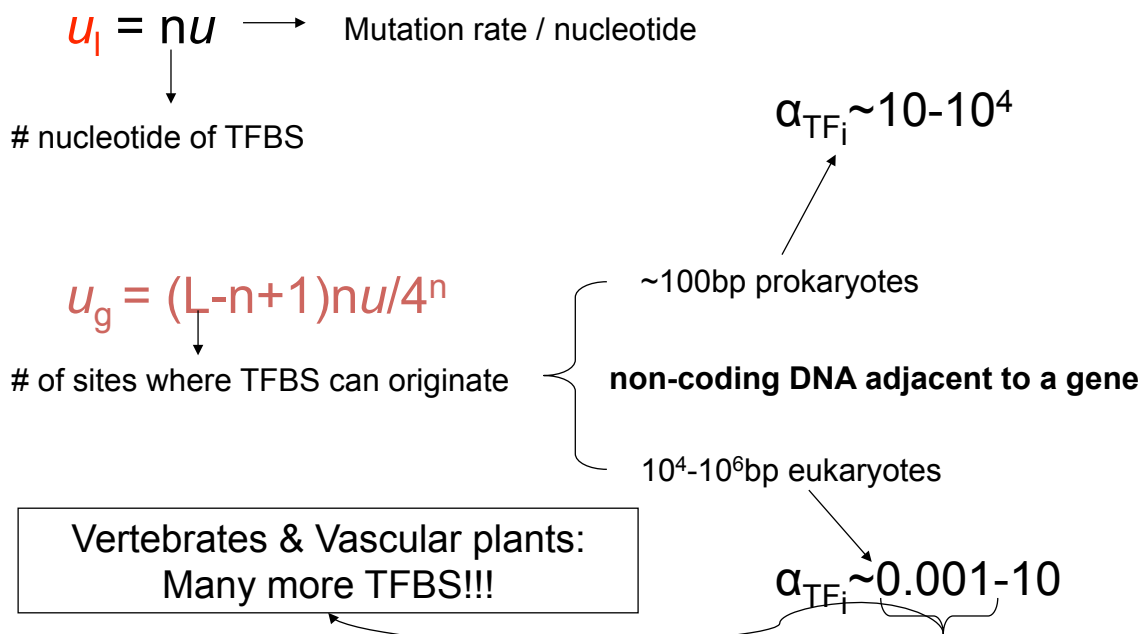
Absence of population-level processes

Consider...

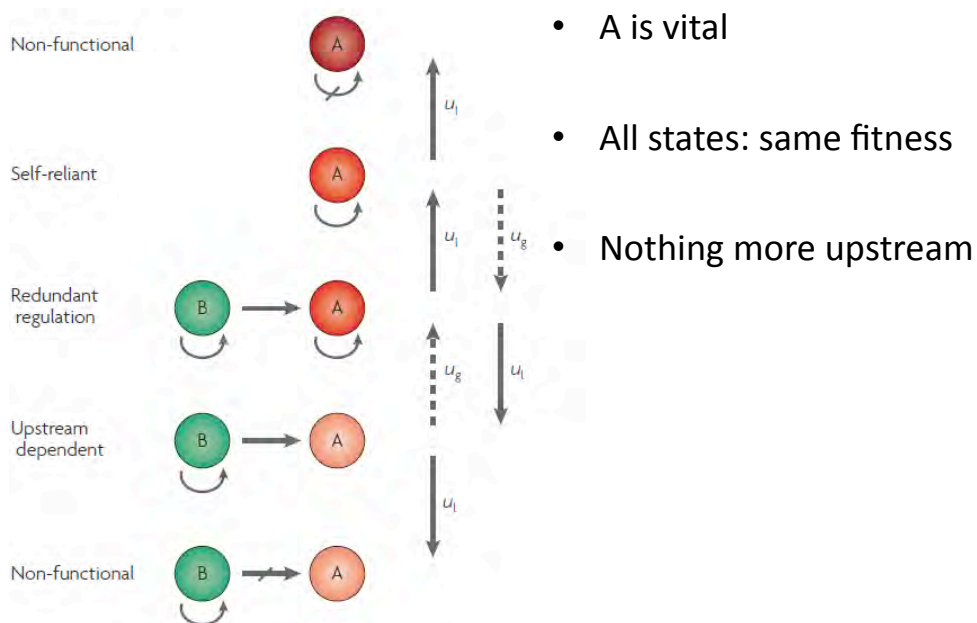
- $D \rightarrow C \rightarrow B \rightarrow A^*$
- Dramatic changes in regulatory machinery without altering phenotype
- Gene duplicate 0.1-1% per N \rightarrow thousands of functional changes at population level in 10^6 generations
- *De novo* origin of TFBS (~5-10bp)
- Evolution of genetic pathways depend:

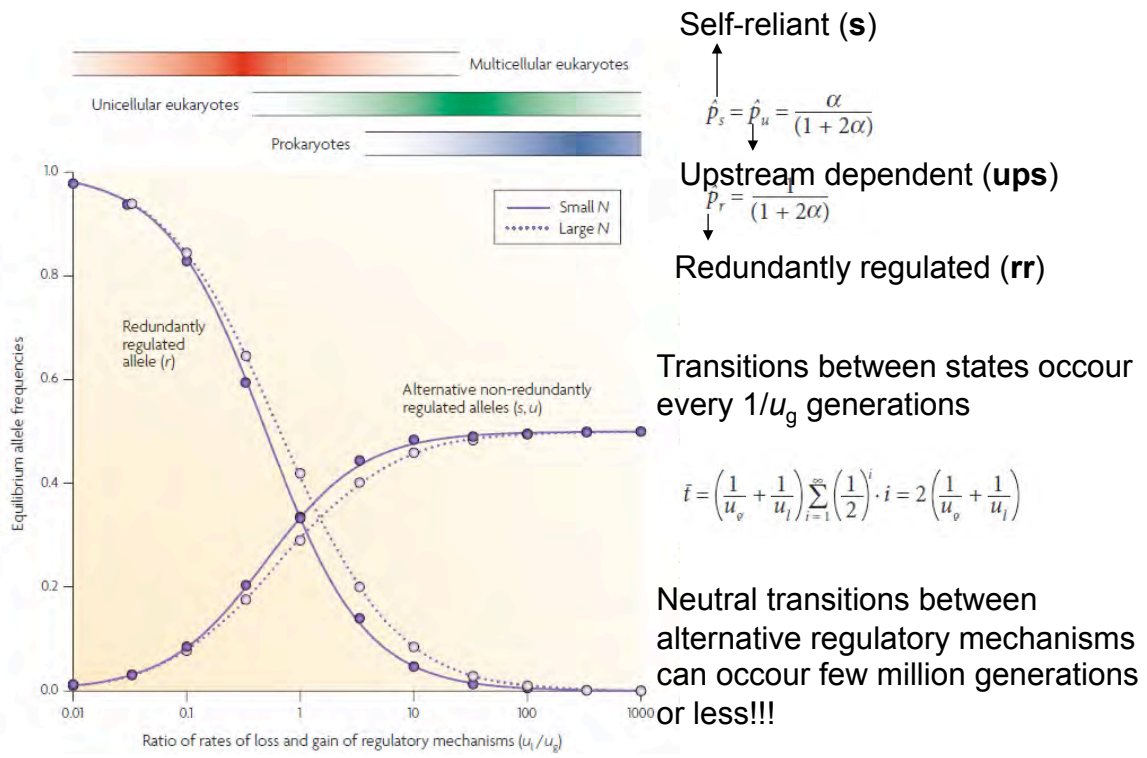
$$\alpha = u_l / u_g$$

Consider... TFBS birth and death process



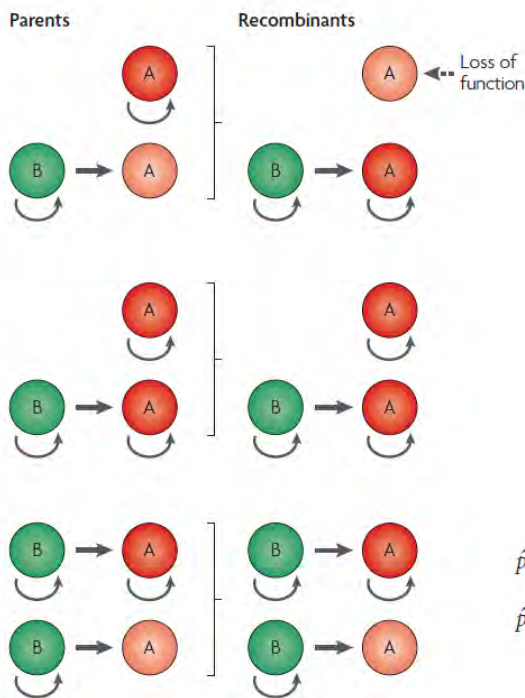
Recruitment of upstream activator





Nearly independent of population size

Recombination – The all mighty



Loss of **ups** A allele → recombinational inactivation
Possible acquisition of latent regulatory elements

All recombinants with at least 1 parental **rr** allele are functional

Indirect recombinational advantage of **rr** allele → increases with population size

Equilibrium frequencies for an effectively infinite population:

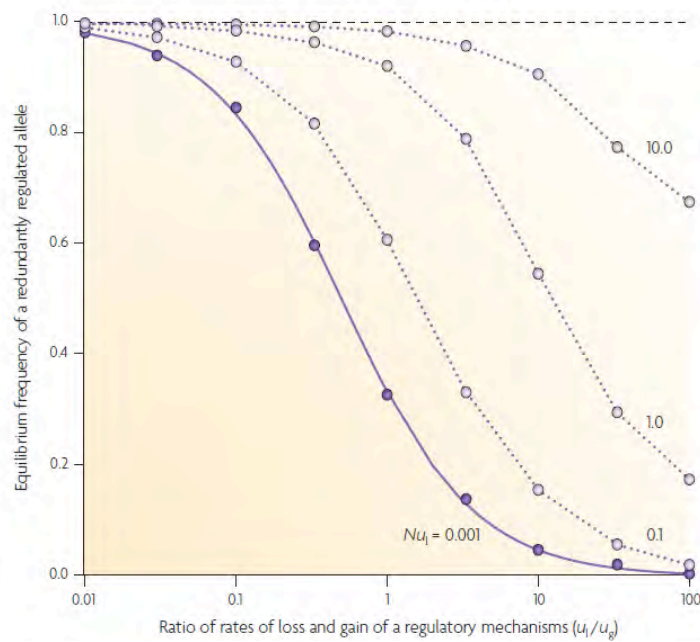
$$\hat{p}_s = \hat{p}_u = - (u_g + 3u_l) + \sqrt{(u_g + 3u_l)^2 + 2u_l} = \sqrt{2u_l}$$

$$\hat{p}_r = 1 - 2\hat{p}_s$$

Nearly all individuals carry **rr**!

$$\sim 1 - 8\text{sqrt}(u)$$

$$\sim <0.001$$

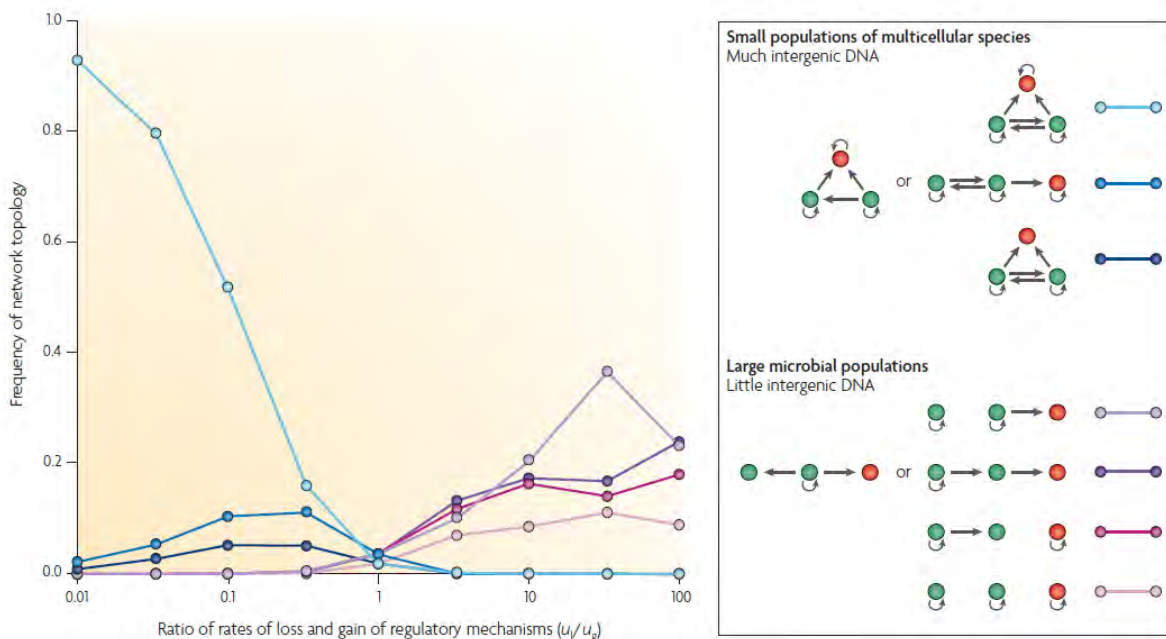


Even in most non-recombinogenic backgrounds, with sufficiently large N , recombination inactivation promotes **rr** alleles if regulatory sites are closer than 100bp

Regulatory redundancy: promoted by at least 2 non-adaptive processes

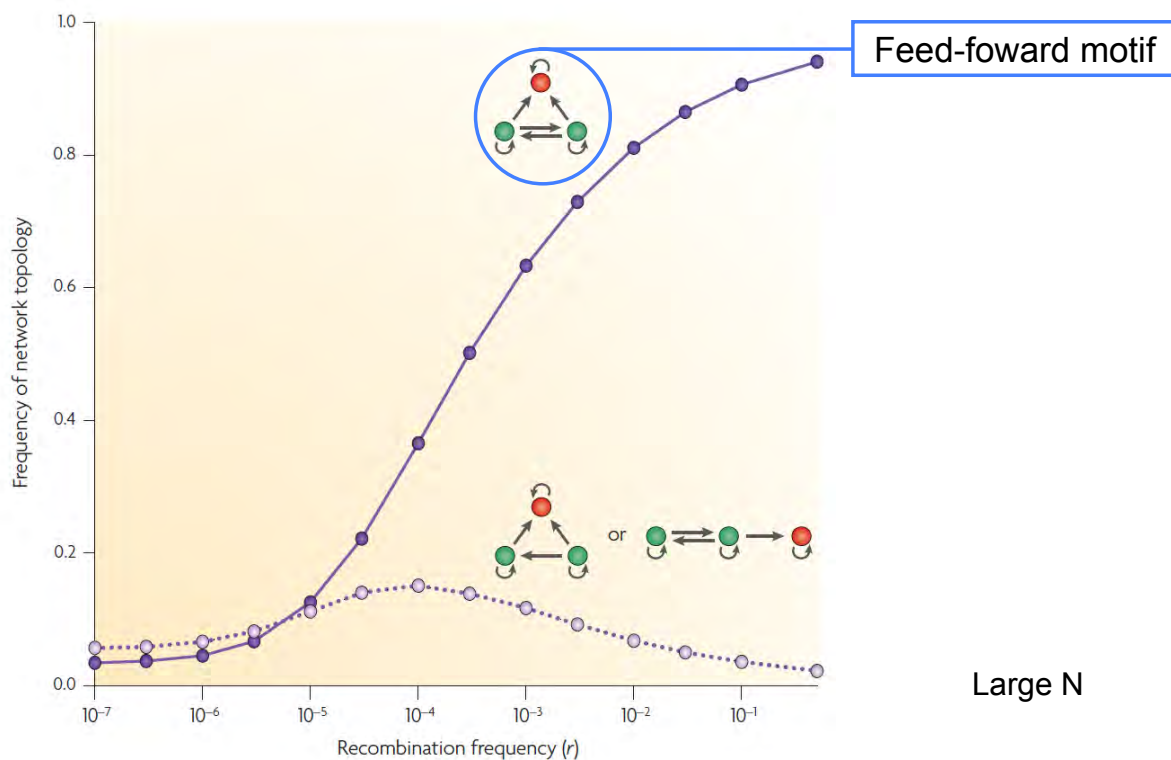
- Selection acts poorly on insertions
- Species with small N accumulate excess DNA → natural susceptibility to TFBS arisal
- High levels of **rr** alleles expected regardless of selective advantage
- Microbial species: little non-coding DNA → few targets for TFBS arisal
- Large N can be sufficient to promote **rr** alleles → recombinational inactivation

Increasing regulatory factors...



No recombination

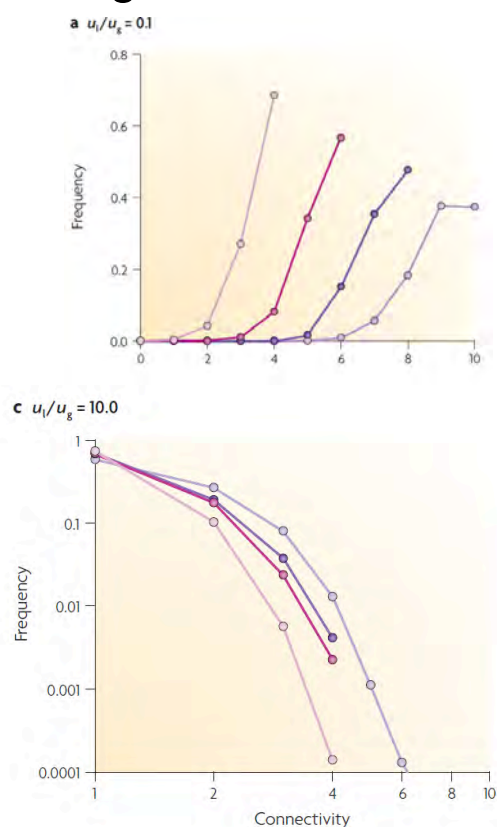
Increasing regulatory factors...With recombination



Larger number of interacting genes

- Increasing number of TFBS (k) increases potential pathways (high computational cost)
- System with low α → mutation pressure primary governing force (multicellular eukariotes)
- Probability distribution of all alternative pathways: approximated by matrix description of pathway architecture
- Stochastic gain/loss of regulatory elements
- Confirm previous results
- Overall level of node connectivity

Larger number of interacting genes



- Power-law as inferred for PINs and TRNs?

- Problematical evolutionary interpretations of node connectivity patterns (non-adaptive origin, time variability, k , α etc.)

Conclusions

- Dependence of average pathway architecture on population-genetics environment
- Considerable variation between expected patterns
- Even with constant average pathway architecture: changes in regulatory network
- Predict allelic variation in regulatory mechanisms in natural populations
(e.g. echinoderm)
- Network complexity not necessarily fixed by selection
- Complexity, modularity, motifs, redundancy and other features cannot be directly attributed to natural selection
- Limits of models of evolution by saltational events: pre-specified rules for stochastic node attachment, gene duplication, TFBS origins with no intermediate states in a populational context, fixation of mutation is not instantaneous
- Inadequacy of null models generated by randomization and saltational changes as neutral models of evolution