

Complex Networks in Systems Biology

Network Topology – Structure – Dynamics

Costas Bouyioukos

UMR7216, Paris Epigénétique et Destin Cellulaire
Université Paris Diderot

M2-BI, November 2017



Overview of the course

Introduction

Contents

Global structure measures and degree distributions

Global Structure Measures

Degree Distributions

Network local structure measures

Clustering coefficient

Examples from the yeast GRN

Functional Motifs and Dynamics

Introduction and terminology

Search Methods

Motifs on Biological Networks

Motifs Dynamics

Introduction

Contents

Global Structure

Global Structure Measures

Distributions

Local Structure

Clustering

Examples

Functional Motifs

Terminology

Finding

Network Motifs

Dynamics

Module II

Structure and function of complex biological networks

1. Global structure measures, degree distributions and organisational principles.
2. Local topological measures and properties
3. Functional motifs and dynamics

Bonus :

Introduction and familiarisation with an open source python library for network analysis `networkX`. For more look at :

<https://networkx.github.io/>

`NetworkX` can be easily installed by :

```
pip install -user networkx
```

Introduction

Contents

Global Structure

Global Structure Measures

Distributions

Local Structure

Clustering

Examples

Functional Motifs

Terminology

Finding

Network Motifs

Dynamics

Introduction

Network Comparisons

- It is crucial to develop ways to quantitatively compare networks, so that we will be able to :
 - To discover common organisational principles.
 - To reveal common (or similar) topologies.
- We need to find such topological measures that :
 - Will be simple in implementation and interpretation
 - Will demonstrate the general network organisation.
 - Will be potentially “statistically significant”.

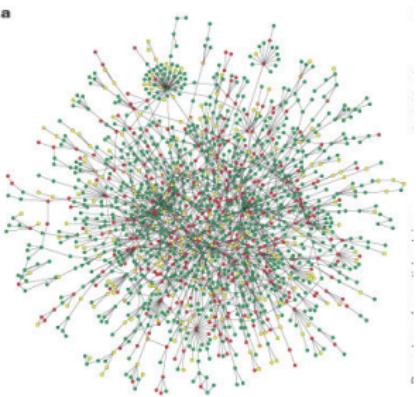
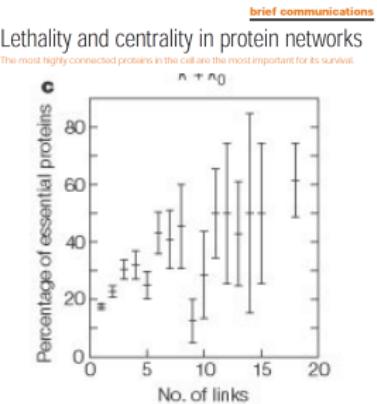
Descriptive Statistics

We will study and discover statistical properties of networks which satisfies the above criteria.

For comparison, a similar “global” measure from genomics is the GC content.

Revisiting the degree

- **Degree** : It is the number of edges (or the number of neighbours) of a node.
 - In PPI networks, degree is associated with :
 - The evolutionary rate of a protein.
 - The essentiality of the coding gene.
 - The probability to be implicated in a disease



More on degrees

Directed networks

Introduction

Contents

Global Structure

Global Structure Measures

Distributions

Local Structure

Clustering

Examples

Functional Motifs

Terminology

Finding

Network Motifs

Dynamics

- The Outdegree :
The number of edges
outgoing (departing) from
a node.

- The Indegree :
The number of edges
incoming (ending) into a
node.

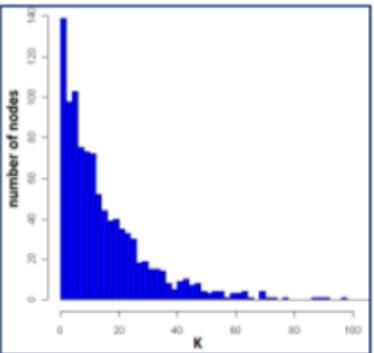
Outdegree
:measures of
expensiveness



Indegree
:measures of
receptivity, or popularity

Degree Distributions

- Degree Distribution :
The probability $P(k)$ that a node has exactly degree k .



- The most common degree distributions :

Poisson

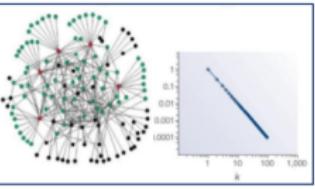
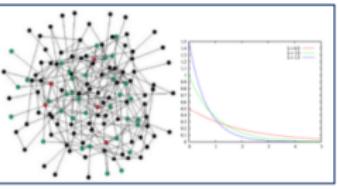
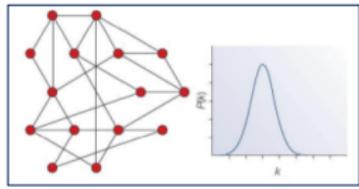
$$P(k) = \frac{e^{-d} d^k}{k!}$$

Exponential

$$P(k) \propto e^{-k/d}$$

Power Law

$$P(k) \propto k^{-c}, k \neq 0, c > 1$$



Introduction

Contents

Global Structure

Global Structure Measures

Distributions

Local Structure

Clustering

Examples

Functional Motifs

Terminology

Finding

Network Motifs

Dynamics

Introduction

Contents

Global Structure

Global Structure Measures

Distributions

Local Structure

Clustering

Examples

Functional Motifs

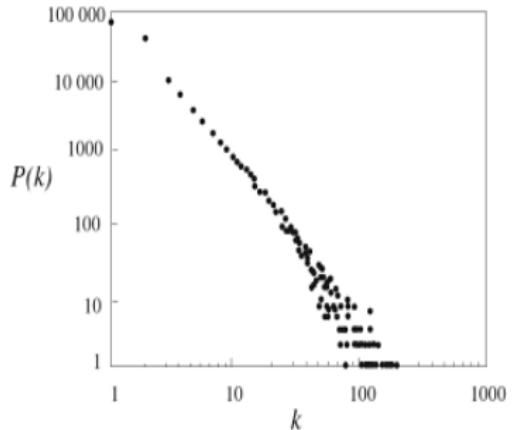
Terminology

Finding

Network Motifs

Dynamics

The internet¹

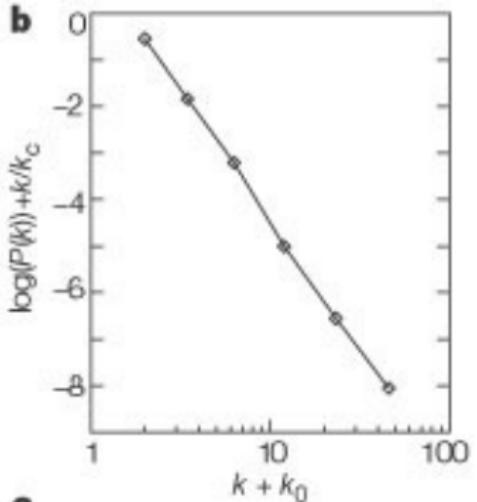


- Nodes : ~150.000 routers.
- Edges, physical connections (with wires, or satellites)
- $P(k) \sim k^{2.3}$

1. Govindan and Tangmunarunkit, 2000

[Introduction](#)[Contents](#)[Global Structure](#)[Global Structure Measures](#)[Distributions](#)[Local Structure](#)[Clustering](#)[Examples](#)[Functional Motifs](#)[Terminology](#)[Finding](#)[Network Motifs](#)[Dynamics](#)

Networks of protein interactions in yeast²



- Nodes : ~10.000 proteins from yeast.
- Edges : P-P Interactions, physical. Stable or transient.
- $P(k) \sim k^{2.5}$

Degree Distribution of “real” Networks

Introduction

Contents

Global Structure

Global Structure Measures

Distributions

Local Structure

Clustering

Examples

Functional Motifs

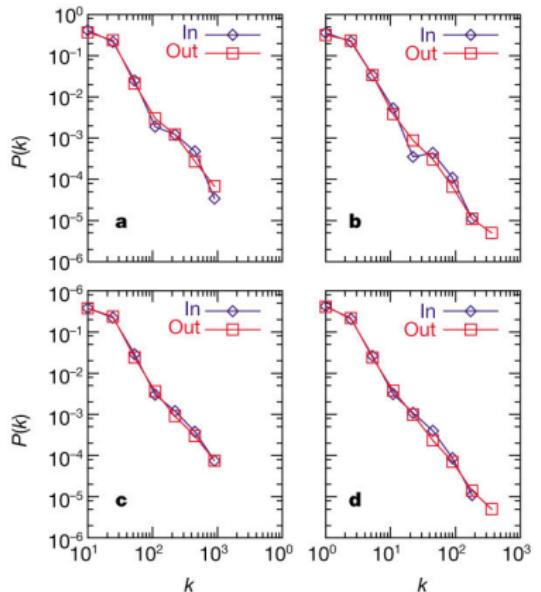
Terminology

Finding

Network Motifs

Dynamics

Metabolic Networks³



- Nodes : Metabolites (primary and secondary)
- Edges : Biochemical reactions
- $P(k) \sim k^{2.2 \pm 2}$

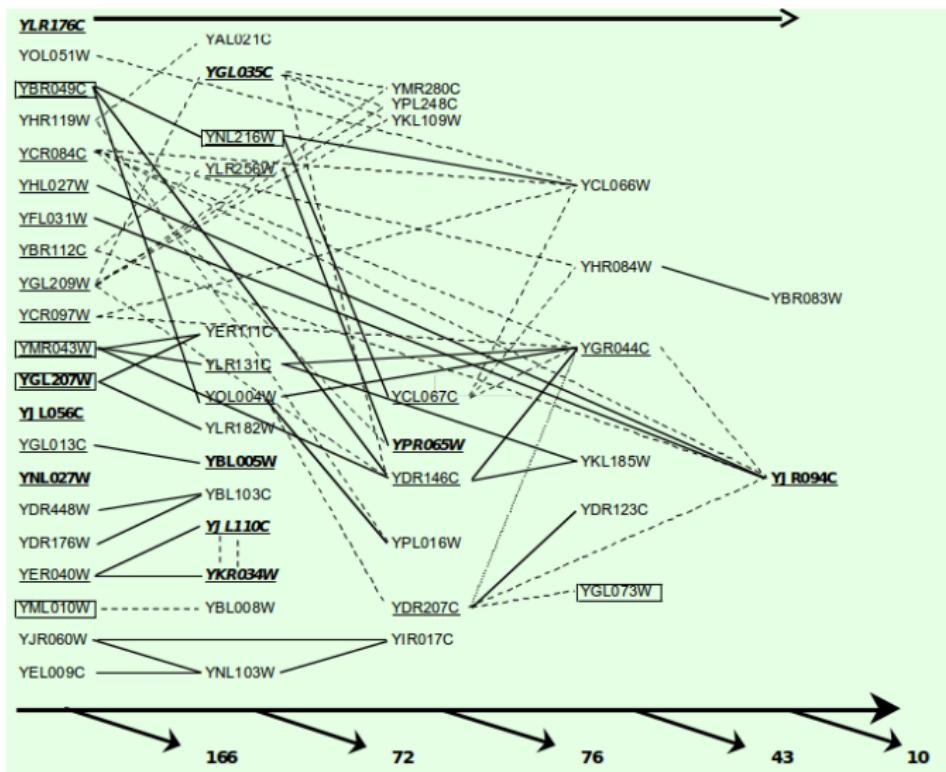
Metabolic networks in all phyla exhibit the *scale free* property.

A : *A. fulgidus*, B : *E. coli*, C : *C. elegans*, D : average

SysBio Complex Networks

Example :

TF regulatory network in yeast

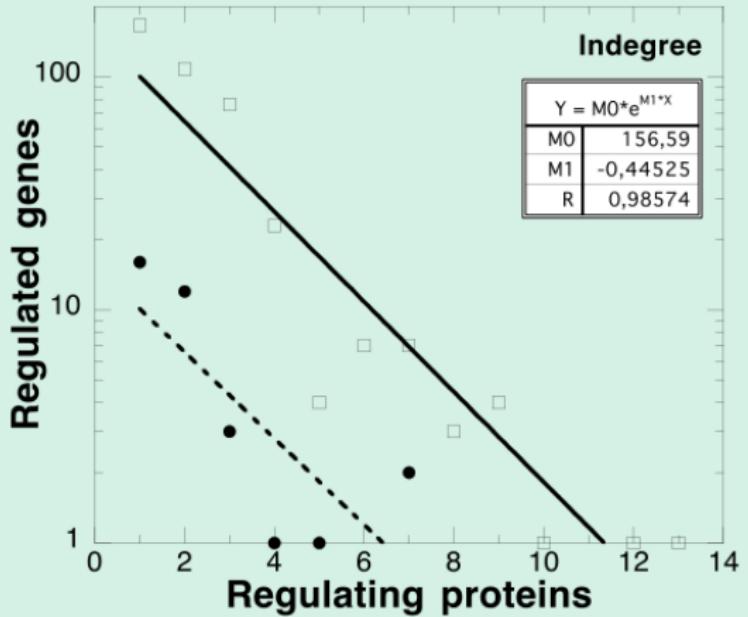


TF regulatory network in yeast II

In-degrees

Μέτρο Συνολικής τοπολογίας

Εκθετική Κατανομή
Μέσος όρος βαθμού: 2.3



Introduction

Contents

Global Structure

Global Structure Measures

Distributions

Local Structure

Clustering

Examples

Functional Motifs

Terminology

Finding

Network Motifs

Dynamics

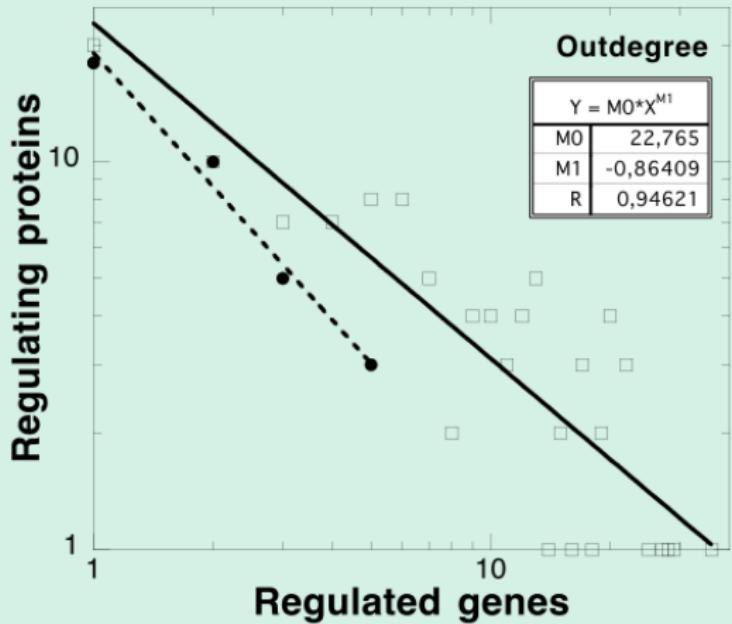
TF regulatory network in yeast III

Out-degrees

Μέτρο Συνολικής Τοπολογίας

Κατανομή Νομού Δύναμης

Μέσος όρος βαθμών: 8.3



Introduction

Contents

Global Structure

Global Structure Measures

Distributions

Local Structure

Clustering

Examples

Functional Motifs

Terminology

Finding

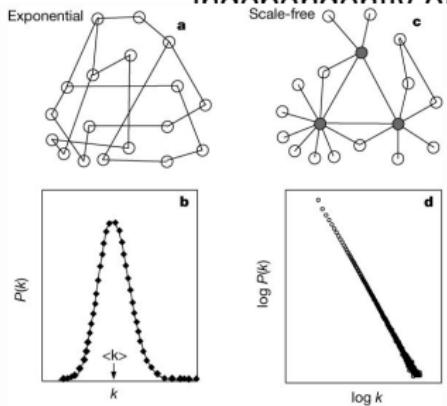
Network Motifs

Dynamics

Power Law Distribution and Scale-Free Networks

- Power law distributions are characterised by their “fat tail”

- Contain a lot of nodes with very few edges and few nodes with a lot of edges (hubs or central nodes).
- The networks are called scale-free because independently of the density of the edges :



1. There is not a characteristic value (i.e. average)
2. The degrees of their nodes follow the following distribution :

$$P(k) \propto k^{-\gamma}$$

Introduction

Contents

Global Structure

Global Structure Measures

Distributions

Local Structure

Clustering

Examples

Functional Motifs

Terminology

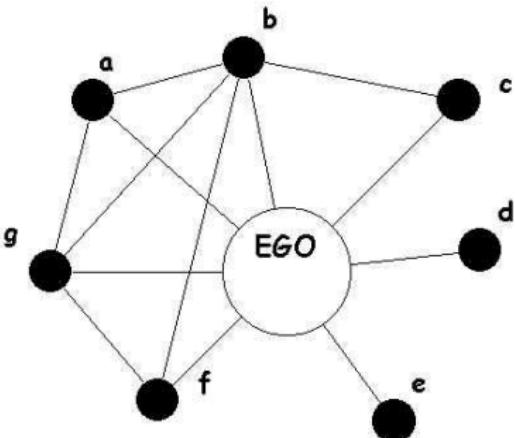
Finding

Network Motifs

Dynamics

EGO Networks

- An EGO network is the sub-network of the 1st order neighbours of a node.
- Studying EGO networks topology is useful for :
 - The development of new genetic markers.
 - Their usage in studies of associations disease networks vs. traits.



Introduction

Contents

Global Structure

Global Structure Measures

Distributions

Local Structure

Clustering

Examples

Functional Motifs

Terminology

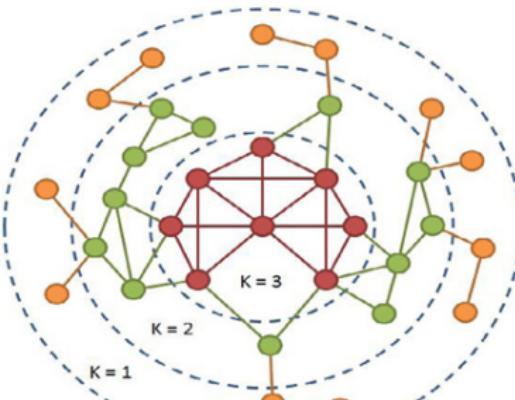
Finding

Network Motifs

Dynamics

***k*-core graphs**

- k -core graphs or degenerated graphs, are the subgraphs which **remain** after the removal of all the nodes with degree smaller than k .
 - It is a global network structure property which allows as to :
 - Compare networks directly ; by comparing higher order k -core graphs
 - Analysing the dynamics of a network ; by computing its highest k -core.



Clustering coefficient

- Clustering Coefficient (C_i) (according to Watts & Strogatz 2000)
 - Informally :
How many of my “friends” know each other (clique) The term clique is a formal term in social networks science.
 - More formally, is a measure of the density of triangles in networks.

Definition

$$C_i = \frac{\text{\# of the edges between the neighbours of } i}{\text{Maximum \# of edges between the neighbours of } i}$$

Local measure, refers to each node. Global measure (average)

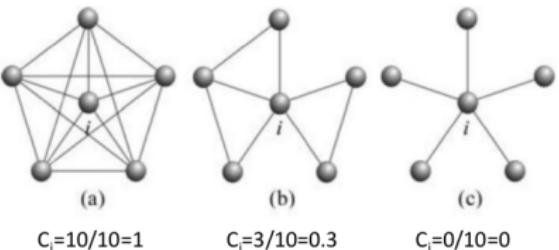
$$C_i = \frac{2E_i}{d_i(d_i - 1)}$$

$$\langle C \rangle = \frac{1}{N} \sum_{i=1}^N C_i$$

Clustering coefficient

Example :

- It is always in the range of $[0, 1]$.
 - For cliques is 1, $C_i = 1$.
 - For graphs without any triangle is $C_i = 0$.



- Important global network measure as it characterises “small world” networks.

Collective dynamics of ‘small-world’ networks

Duncan J. Watts* & Steven H. Strogatz

Department of Theoretical and Applied Mechanics, Kimball Hall,
Cornell University, Ithaca, New York 14853, USA

Introduction

Contents

Global Structure

Global Structure Measures

Distributions

Local Structure

Clustering

Examples

Functional Motifs

Terminology

Finding

Network Motifs

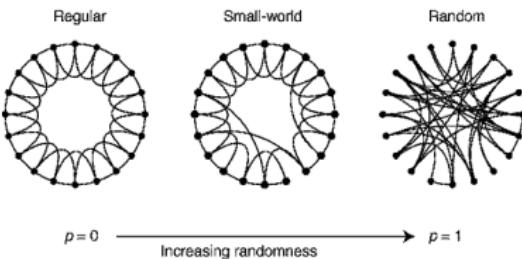
Dynamics

Average distance

Distance and “small world” networks

- **Distance** : The length of the **shortest** path between two nodes.
- Average distance

$$\langle l \rangle = \frac{2}{n(n+1)} \sum_{i \geq j} d_{ij}$$



Characteristics of “small world” networks

- Average clustering coefficient much higher w.r.t. “random networks”.
- Average distance much shorter w.r.t. “random networks”.

...It's a small world after all.

Introduction

Contents

Global Structure

Global Structure Measures

Distributions

Local Structure

Clustering

Examples

Functional Motifs

Terminology

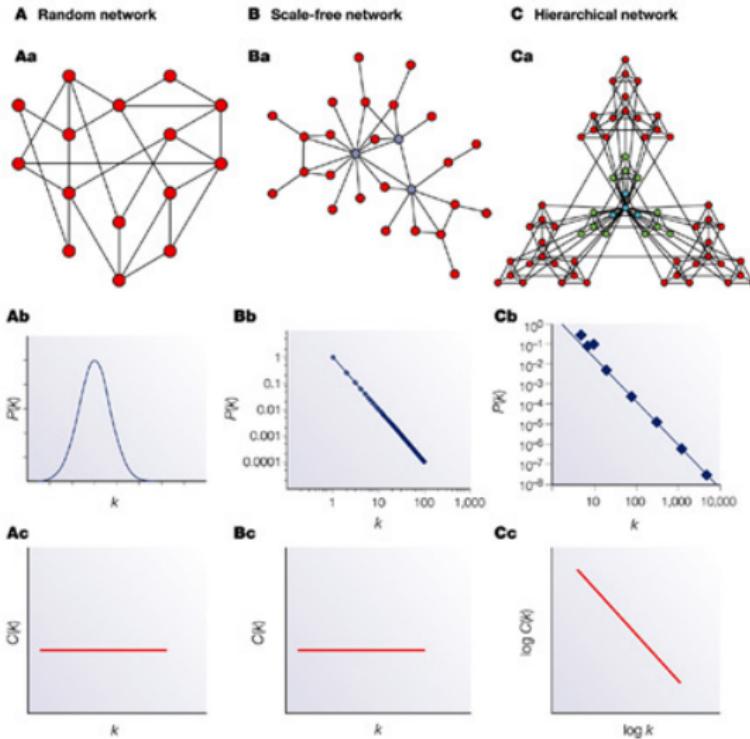
Finding

Network Motifs

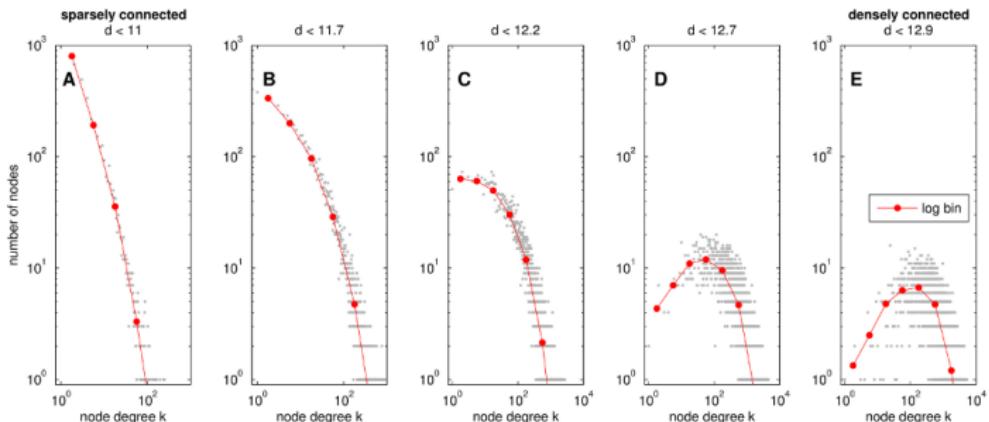
Dynamics

More on power law degree distribution

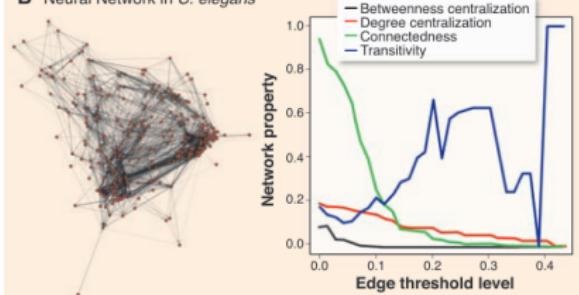
... and a new type of network



Dispute of the power-law degree distribution



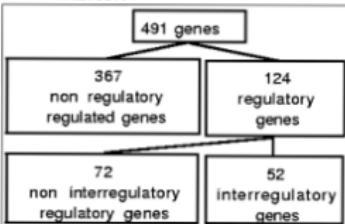
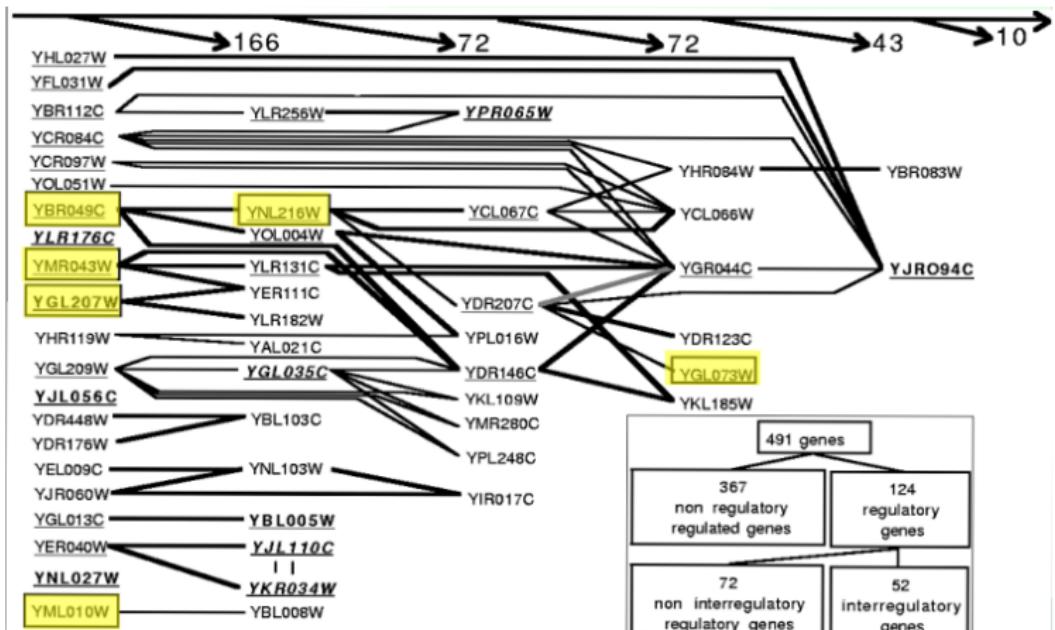
B Neural Network in *C. elegans*



- A** The network density.
- B** The threshold to calculate interactions.

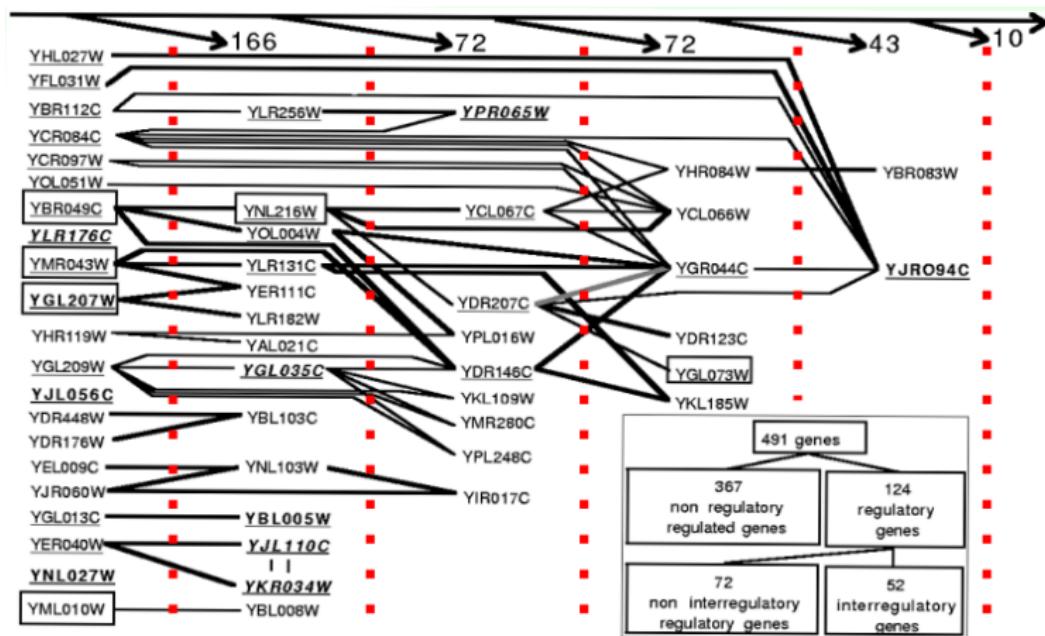
Essential regulatory genes

Substantiality - connectivity



Small World network

Longest distance 5



Additional measures

- Giant Connected Component
- Network modularity
- Centrality measures :
 - Betweenness centrality
 - Eigenvector centrality
 - Bridging centrality
- Information flow
- Spectral density

... but also 100s of more in [Boccaletti S. et al. 2006](#)

Introduction

Contents

Global Structure

Global Structure Measures

Distributions

Local Structure

Clustering

Examples

Functional Motifs

Terminology

Finding

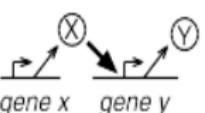
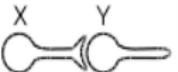
Network Motifs

Dynamics

Motif extraction

Homogeneous Networks

- Why do we need network segmentation ;
 1. Explain better network dynamics.
 2. Better analyse modules.
 3. Discover biological properties underlying the structure.
 4. Compositionality : The principle under which the significance of a complex entity is determined by the significance of its components.

$X \rightarrow Y$ represents	transcription network	neuron synaptic connection network	ecological food web
			

Introduction

Contents

Global Structure

Global Structure Measures

Distributions

Local Structure

Clustering

Examples

Functional Motifs

Terminology

Finding

Network Motifs

Dynamics

What are network motifs

- Repeating patterns of interactions (sub-graphs) which are statistically significantly overrepresented in the network of study compared to a “null” model.

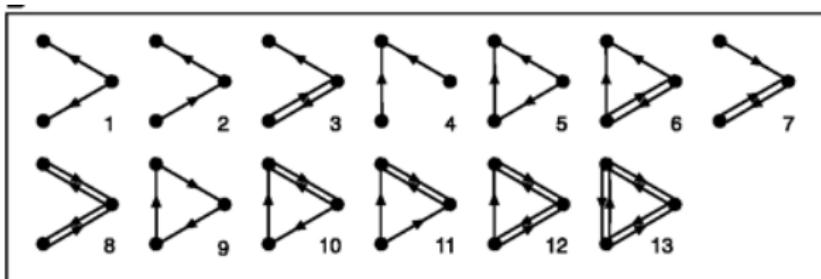


FIGURE – All 13 possible 3-node subgraphs in a directed network.

Introduction

Contents

Global Structure

Global Structure Measures

Distributions

Local Structure

Clustering

Examples

Functional Motifs

Terminology

Finding

Network Motifs

Dynamics

What are network motifs

- Repeating patterns of interactions (sub-graphs) which are statistically significantly overrepresented in the network of study compared to a “null” model.

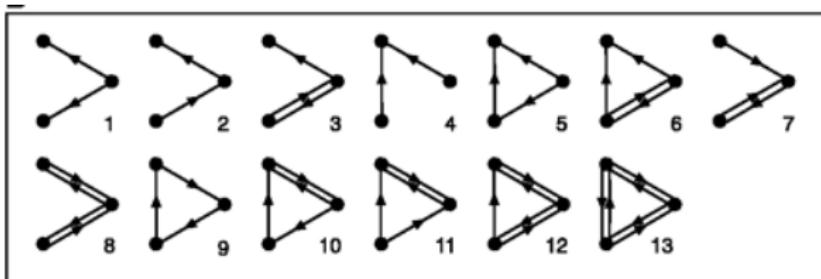


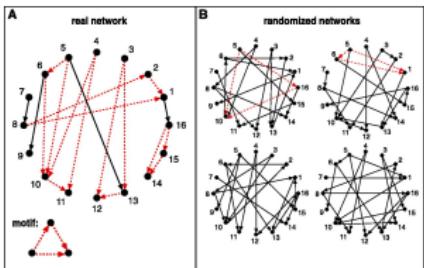
FIGURE – All 13 possible 3-node subgraphs in a directed network.

(199 possible 4-node sub-graphs.)

Searching for network motifs

Algorithm :

1. Generation of a large number of randomised networks.
 2. Scanning of the network of study to count all the different n-node sub-graphs.
 3. Calculating frequency of all n-node sub-graphs.
 4. Repeat steps 2 and 3 for each of the randomised networks.
 5. Calculate empirical p-value and evaluate the significance of the appearance of each of the n-node sub-graphs.



Network Motifs: Simple Building Blocks of Complex Networks

R. Milo,¹ S. Shen-Orr,¹ S. Itzkovitz,¹ N. Kashtan,¹ D. Chklovskii,²
U. Alon^{1*}

Complex networks are studied across many fields of science. To uncover their structural design principles, we defined "network motifs," patterns of interconnections occurring in complex networks at numbers that are significantly higher than those in randomized networks. We found such motifs in networks from biochemistry, neurobiology, ecology, and engineering. The motifs shared by ecological food webs were distinct from the motifs shared by the genetic networks of *Escherichia coli* and *Saccharomyces cerevisiae* or from those found in the World Wide Web. Similar motifs were found in networks that perform information processing, even though they describe elements as different as biomolecules within a cell and synaptic connections between neurons in *C. elegans*. Motifs may thus define universal classes of networks. This approach may uncover the basic building blocks of most networks.

Randomised Networks

Generation of a set of randomised networks.

- What makes a **good** background (or null) model.
- Do we really want totally “random” networks, or we wish to maintain some of the “characteristic” properties.

Randomised networks

Maintain the number of incoming and outgoing edges from each node.

For motifs $k > 3$ nodes we need to keep the frequency of the lower order $k - 1$ -node sub-graphs.

Introduction

Contents

Global Structure

Global Structure Measures

Distributions

Local Structure

Clustering

Examples

Functional Motifs

Terminology

Finding

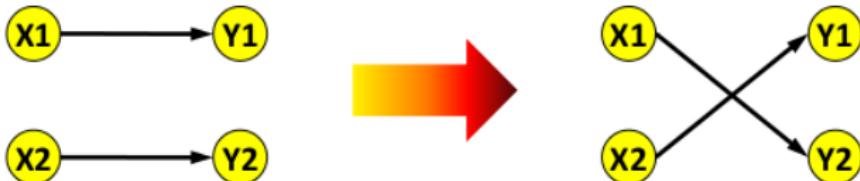
Network Motifs

Dynamics

Randomising Networks

Algorithm :

- Markov chain algorithm steps :
 1. Starting from the “real” network we choose a pair of edges randomly.
 2. We swap the edges according too : $(x_1, y_1), (x_2, y_2)$ will become $(x_1, y_2), (x_2, y_1)$.
 - the random swapping is not accepted if the new, to be formed, (x_1, y_2) or (x_2, y_1) edges do exist in the edge list.
 3. Repeat the steps 2 and 3 for a sufficiently and reasonably large amount of times.



Motif discovery criteria

- A motif is selected as statistically significant if, and only if, satisfies the following criteria :
 1. Its frequency in the network of study is at least 4.
 2. The probability to appear in a random network more times than in the real is at least $p \leq 0.001$.
(yes, this is the empirical p-value ! therefore at least how many random networks we have to construct ?)
 3. The frequency of its appearance in the network of study is “substantially” higher than in the random networks :
$$(N_{real} - N_{rand} > 0.1N_{rand})$$

Introduction

Contents

Global Structure

Global Structure Measures

Distributions

Local Structure

Clustering

Examples

Functional Motifs

Terminology

Finding

Network Motifs

Dynamics

Motif discovery criteria

- A motif is selected as statistically significant if, and only if, satisfies the following criteria :
 1. Its frequency in the network of study is at least 4.
 2. The probability to appear in a random network more times than in the real is at least $p \leq 0.001$.
(yes, this is the empirical p-value ! therefore at least how many random networks we have to construct ?)
 3. The frequency of its appearance in the network of study is “substantially” higher than in the random networks :
$$(N_{real} - N_{rand} > 0.1N_{rand})$$

Introduction

Contents

Global Structure

Global Structure Measures

Distributions

Local Structure

Clustering

Examples

Functional Motifs

Terminology

Finding

Network Motifs

Dynamics

Motif discovery criteria

- A motif is selected as statistically significant if, and only if, satisfies the following criteria :
 1. Its frequency in the network of study is at least 4.
 2. The probability to appear in a random network more times than in the real is at least $p \leq 0.001$.
(yes, this is the empirical p-value ! therefore at least how many random networks we have to construct ?)
 3. The frequency of its appearance in the network of study is “substantially” higher than in the random networks :
$$(N_{real} - N_{rand} > 0.1 N_{rand})$$

Introduction

Contents

Global Structure

Global Structure Measures

Distributions

Local Structure

Clustering

Examples

Functional Motifs

Terminology

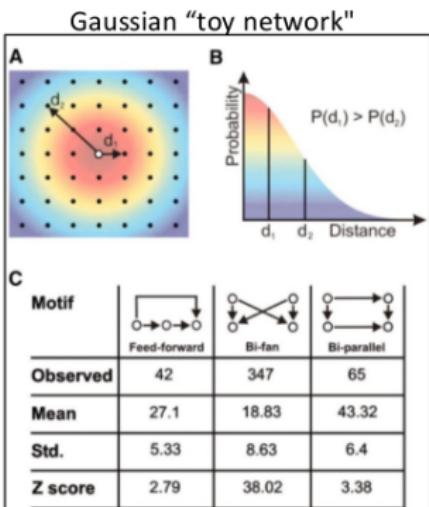
Finding

Network Motifs

Dynamics

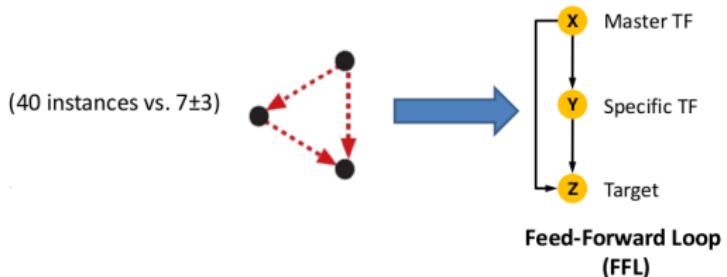
Criticism on the randomisation process

- Is the background model complete ?
- Incompatibility of random edges with regulation or metabolism.
- Local clustering effect :
 - e.g. neurons which are physically closer have higher probability to be connected. (A and B)
- Functional motifs appear significant in networks without rules.
 - In a simple Gaussian model (C)
 - In differential attachment networks.



Y. Artzy-Randrup et al. Comment on "Network motifs: simple building blocks of complex networks".

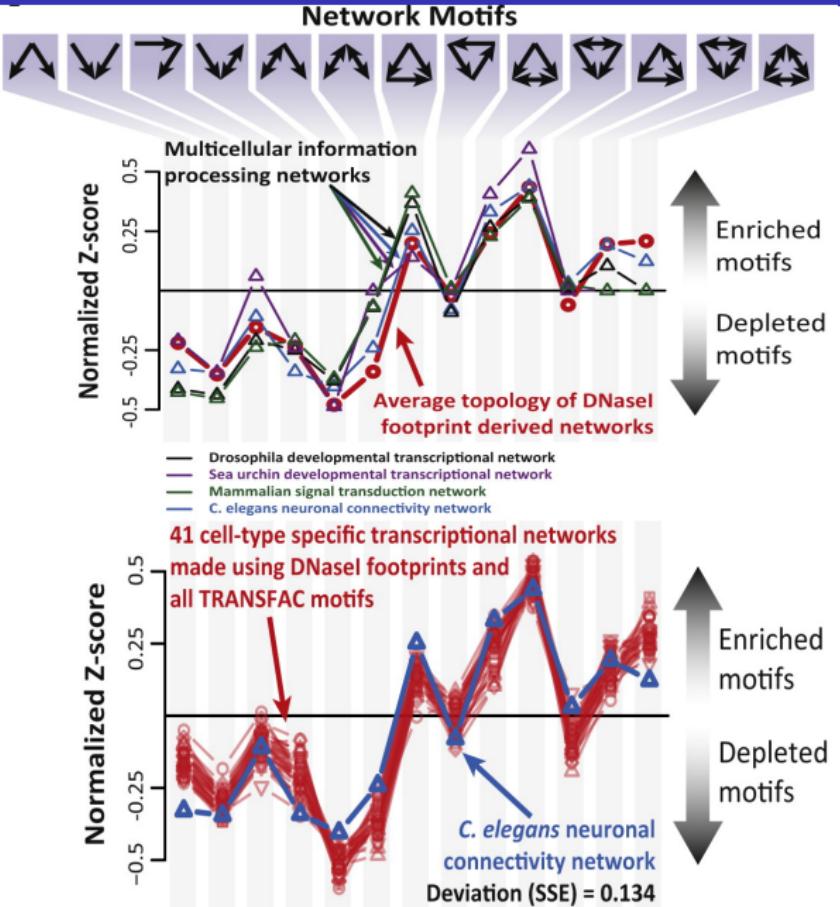
- *E.coli* genetic regulatory network.
 - 442 operons, 116 TFs.
 - 577 interactions (known at that time).
 - Statistically significant enrichment of motif No.5



Coherent Feed Forward Loop, FFL

- The action of X on Y has the same sign as the action of X on Z.
- More than 85% of all FFLs are coherent.

Motif from the GRN of *H.sapiens*



Introduction

Contents

Global Structure

Global Structure Measures

Distributions

Local Structure

Clustering

Examples

Functional Motifs

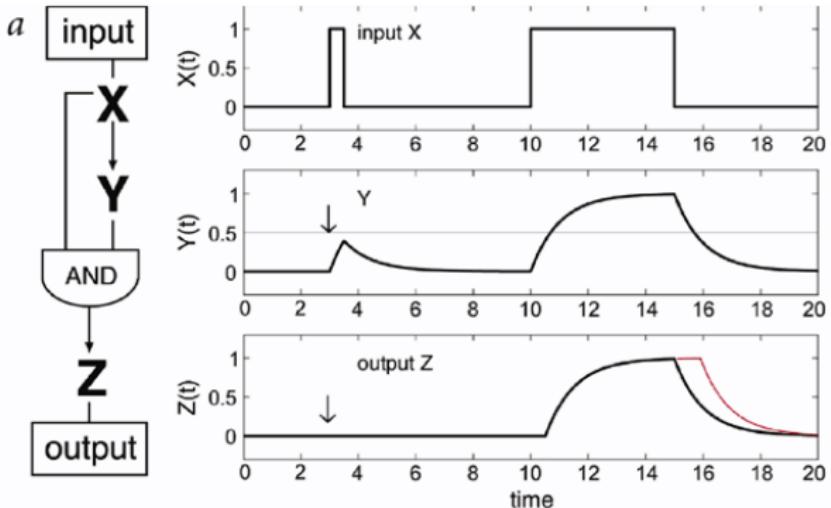
Terminology

Finding

Network Motifs

Dynamics

Coherent FFL dynamics



A coherent feed forward loop is operating as a circuit which filters the transient transduction signals, reacts only in persisting signals and allows a swift shut down of the circuit after the end of the stimuli.

Introduction

Contents

Global Structure

Global Structure Measures

Distributions

Local Structure

Clustering

Examples

Functional Motifs

Terminology

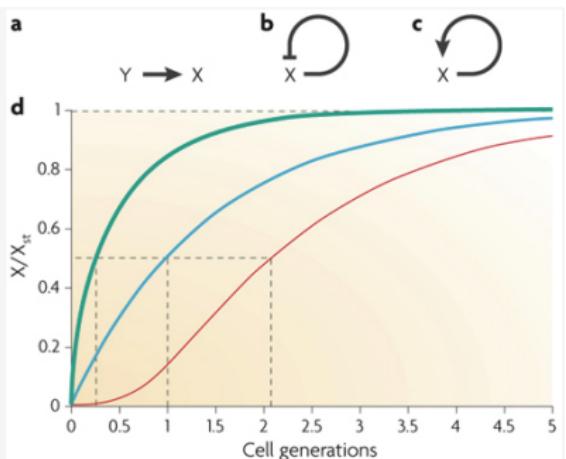
Finding

Network Motifs

Dynamics

Feedback loop dynamics

1. Positive feedback (b) and negative feedback (c)
2. Positive feedback increases the time of reaction to the signal and the time to reach equilibrium.
3. Negative feedback has got much sharper response and reaches equilibrium faster.



Introduction

Contents

Global Structure

Global Structure Measures

Distributions

Local Structure

Clustering

Examples

Functional Motifs

Terminology

Finding

Network Motifs

Dynamics

Dynamics of forward loops

Circuits

Introduction

Contents

Global Structure

Global Structure Measures

Distributions

Local Structure

Clustering

Examples

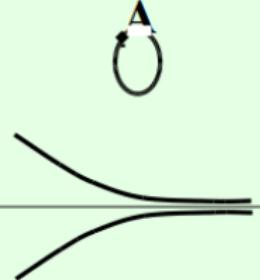
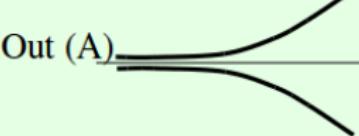
Functional Motifs

Terminology

Finding

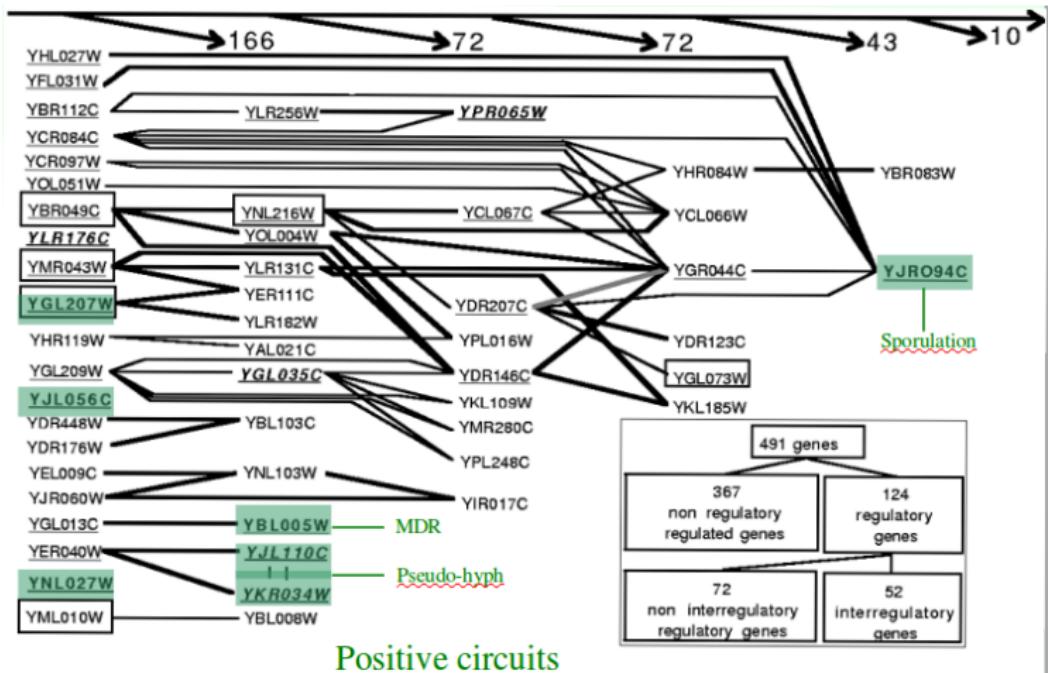
Network Motifs

Dynamics

MODULES	FEEDBACK CIRCUITS	
	POSITIVE	NEGATIVE
SIGN	Even	Odd
# negative interactions		
Dynamic property	Multistationarity	Homeostasis
Biological property	Differentiation	Stable regulation
Topology		
Qualitative dynamics (<i>in vivo</i> and numerical simulations carried out)	Out (A) 	

Dynamics of feed forward loops

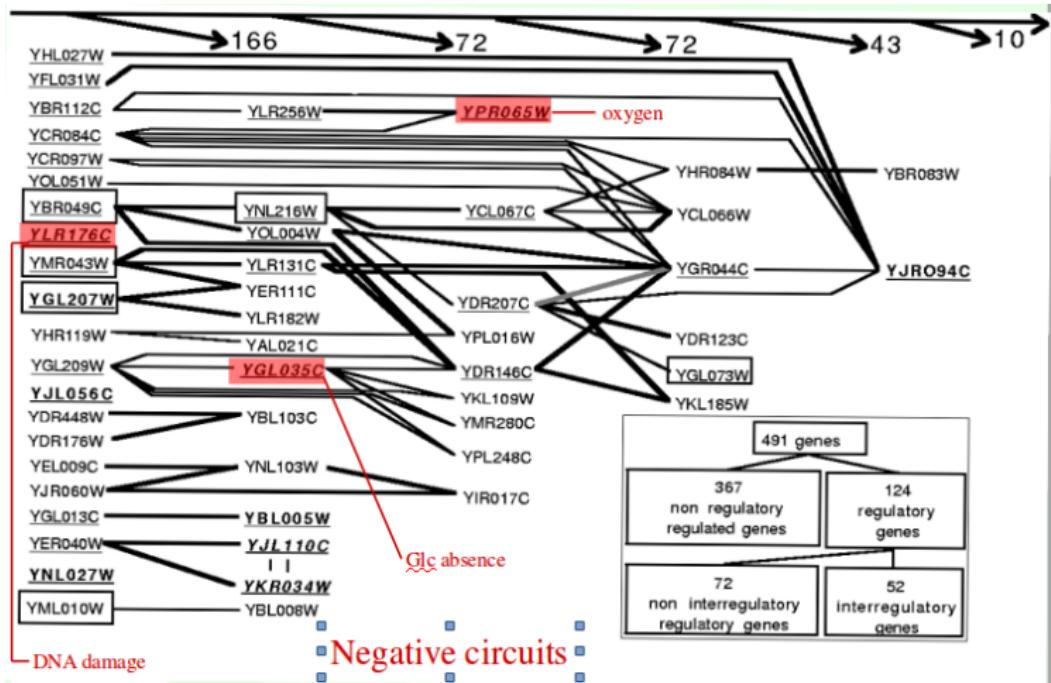
At yeast GRN



Positive circuits

Dynamics of negative feedback loops

At yeast GRN



Introduction

Contents

Global Structure

Global Structure Measures

Distributions

Local Structure

Clustering

Examples

Functional Motifs

Terminology

Finding

Network Motifs

Dynamics

MODULES

TYPE

Dynamic property

CASCADES (Linear set of regulations)

SHORT

Rapidly shoots up

LONG

Long lag before it shoots up

Biological property

Fast response in microbe

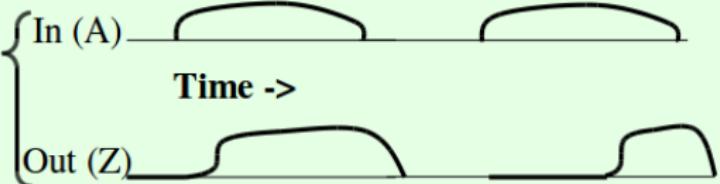
Time counting
in multicellular

Topology

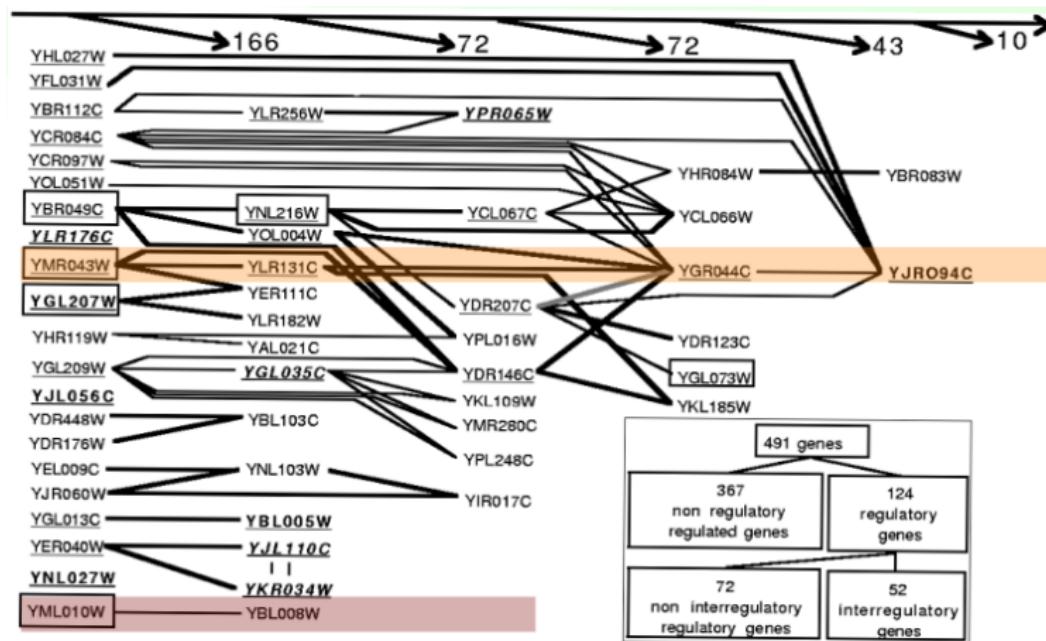


Qualitative dynamics

(Not simulated yet)



Signalling Cascades in Yeast



Introduction

Contents

Global Structure

Global Structure Measures

Distributions

Local Structure

Clustering

Examples

Functional Motifs

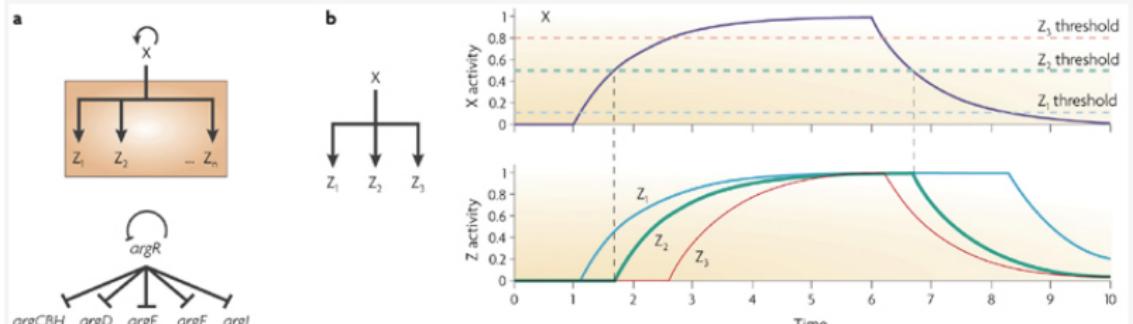
Terminology

Finding

Network Motifs

Single Input Motif (SIM)

- A motif which is based at the regulation of a single factor (the master regulator)
- This motif can control the ordered response to stimuli.
- Different activation thresholds generate repeated activation. The turn and the height of each threshold (i.e. kinetics) control the timely ordered expression.



Introduction

Contents

Global Structure

Global Structure Measures

Distributions

Local Structure

Clustering

Examples

Functional Motifs

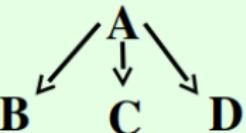
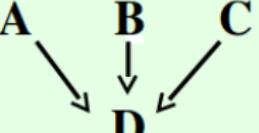
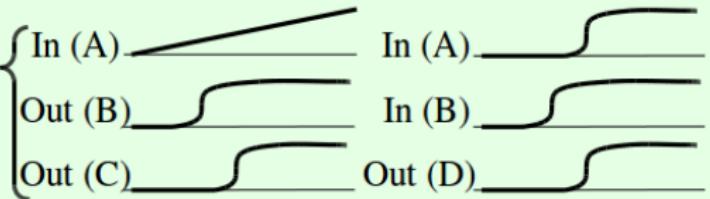
Terminology

Finding

Network Motifs

Dynamics

Single Input Single Output Motifs

MODULES	SIM / SOM	
TYPE	Single-Input Module	Single-Output Module
Dynamic property	Ordered temporal response	' AND ' logical gate
Biological property	Sequential firing based on differential thresholds	Co-regulation
Topology		
Qualitative dynamics (Verified <i>in vivo</i>)		

Introduction

Contents

Global Structure

Global Structure Measures

Distributions

Local Structure

Clustering

Examples

Functional Motifs

Terminology

Finding

Network Motifs

Dynamics

Introduction

Contents

Global Structure

Global Structure Measures

Distributions

Local Structure

Clustering

Examples

Functional Motifs

Terminology

Finding

Network Motifs

Dynamics

MODULES Combination of long cascade and positive circuits

TYPE

LOCK-ON

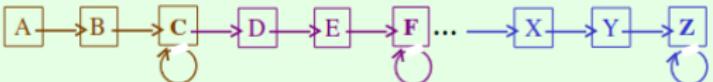
Dynamic property

Ratchet

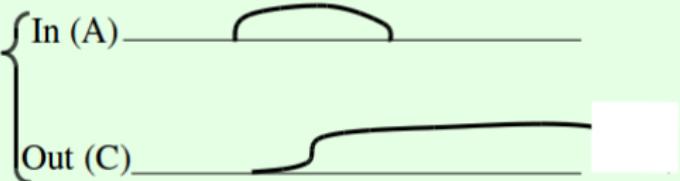
Biological property

Succession of time lags and differentiation events in multicellular development

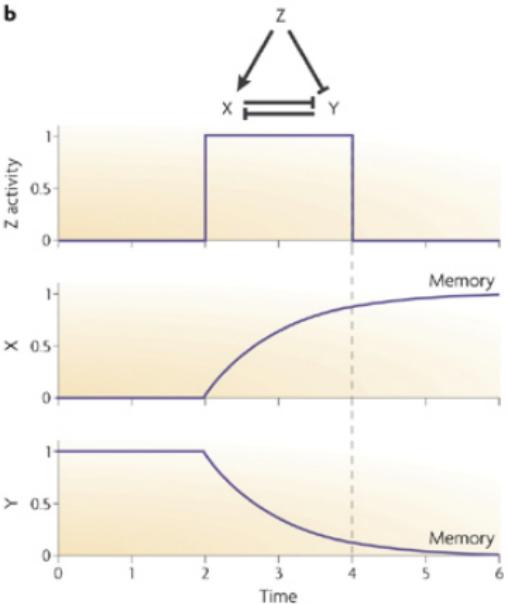
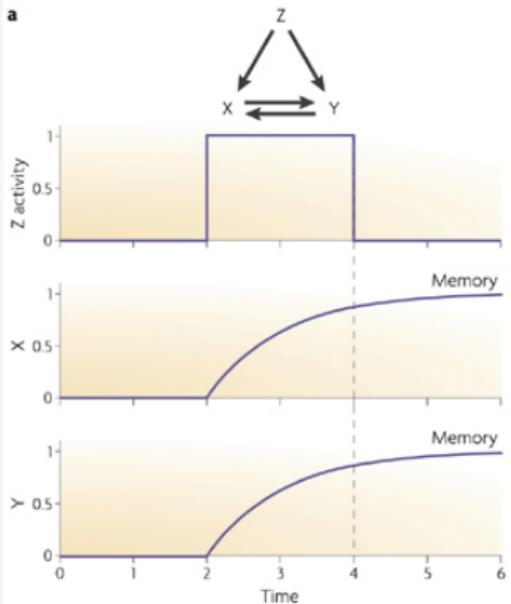
Topology



Qualitative dynamics
for a single-colored
series
(Not simulated yet)



Motif dynamics in Developmental regulatory Networks



- Double **positive** feedback loop (a). Master switch.
- Double **negative** feedback loop (b). Alternator.

Methods/software, publications

Introduction

Contents

Global Structure

Global Structure Measures

Distributions

Local Structure

Clustering

Examples

Functional Motifs

Terminology

Finding

Network Motifs

Dynamics

Software

- Stand-alone software for motif detection [FANMOD](#)
- Cytoscape plug-in for motif detection [CytoKavosh](#)

Further reading

Current innovations and future challenges of network motif detection