

Systems Biology Across Scales: A Personal View

III. Networks: Basic concepts

Sitabhra Sinha
IMSc Chennai

Ubiquity of Networks

Networks appear at all scales in biology

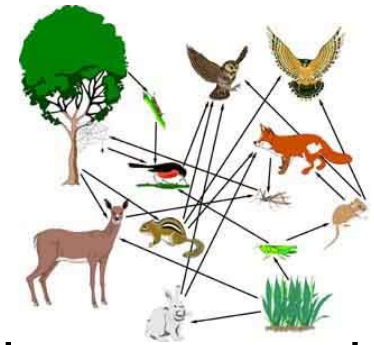
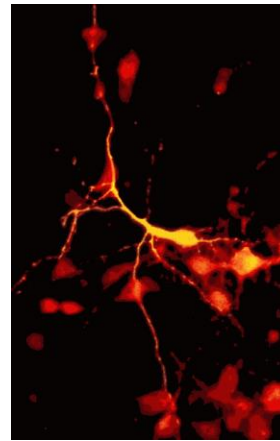
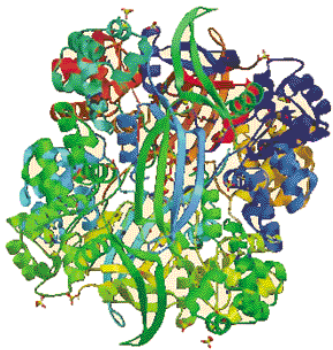
Proteins

Intra-cellular
signalling

Neuronal
communication

Epidemics

Food webs



10^{-9} m

10^{-6} m

10^{-3} m

1 m

10^3 m

10^6 m

Molecules

Cells

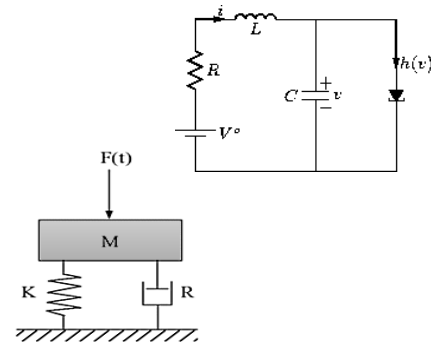
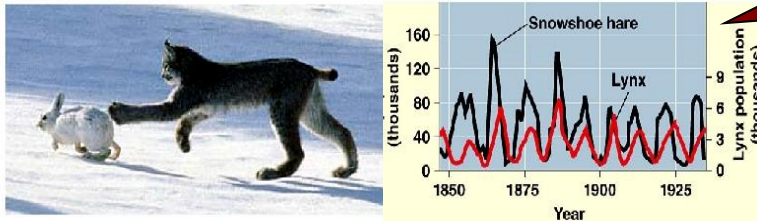
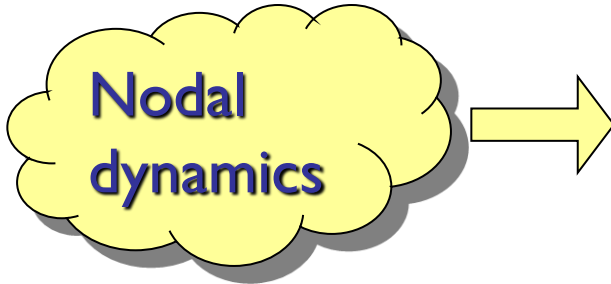
Organisms

Populations

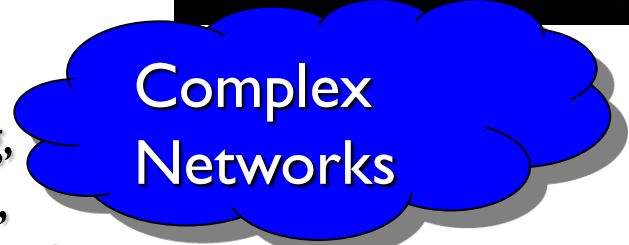
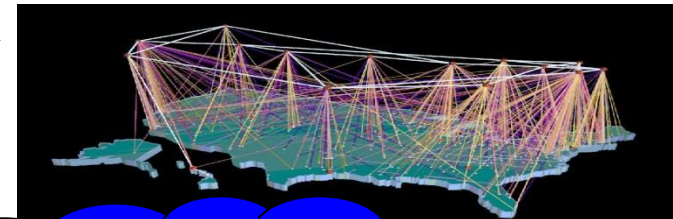
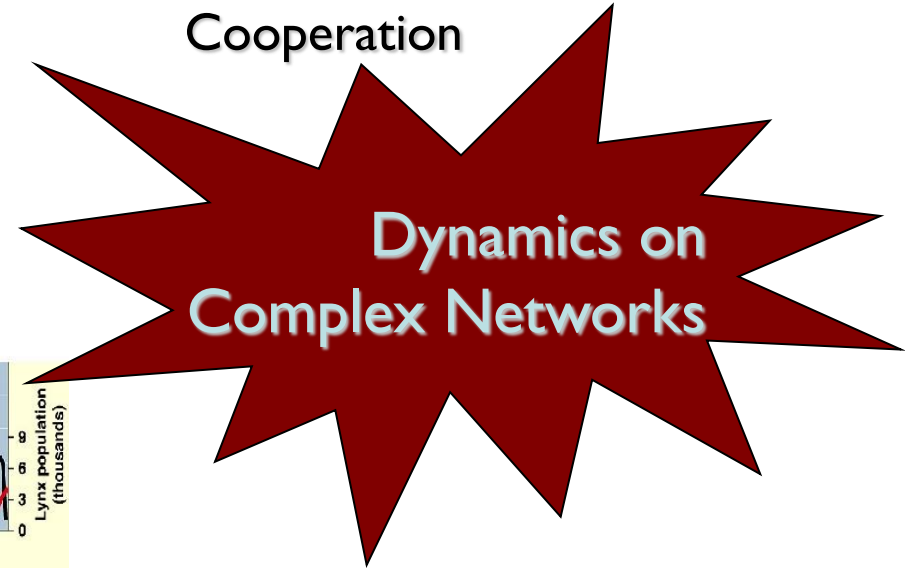
Ecologies

Complexity of dynamics

Nonlinear dynamics,
Game-theory strategies



Synchronization,
Consensus,
Cooperation



Ordering,
Diffusion,
Contagion Spreading

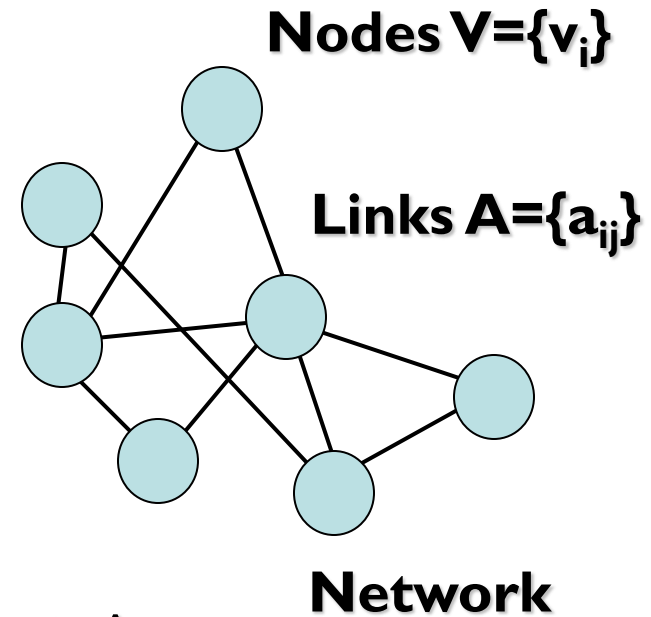
Complexity of interconnection

What is a network ?

Components = Nodes or Vertices

Interactions = Links or Edges

System = Network or Graph



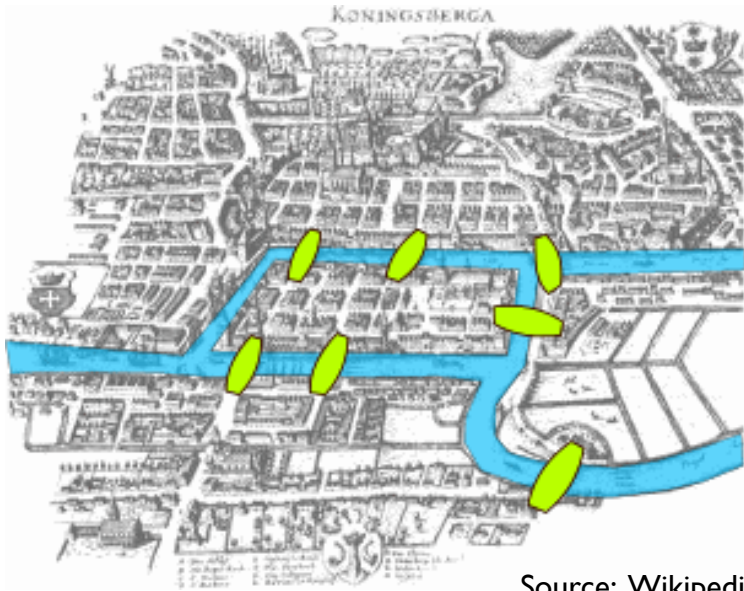
Network structure is defined by *adjacency matrix* A

$$A_{ij} = 1, \text{ if a link exists between } i \text{ and } j \text{ (} \neq i \text{)}$$
$$= 0, \text{ otherwise}$$

What is a *complex network* ?

Other than regular networks (d-dimensional lattice) or homogeneous random networks

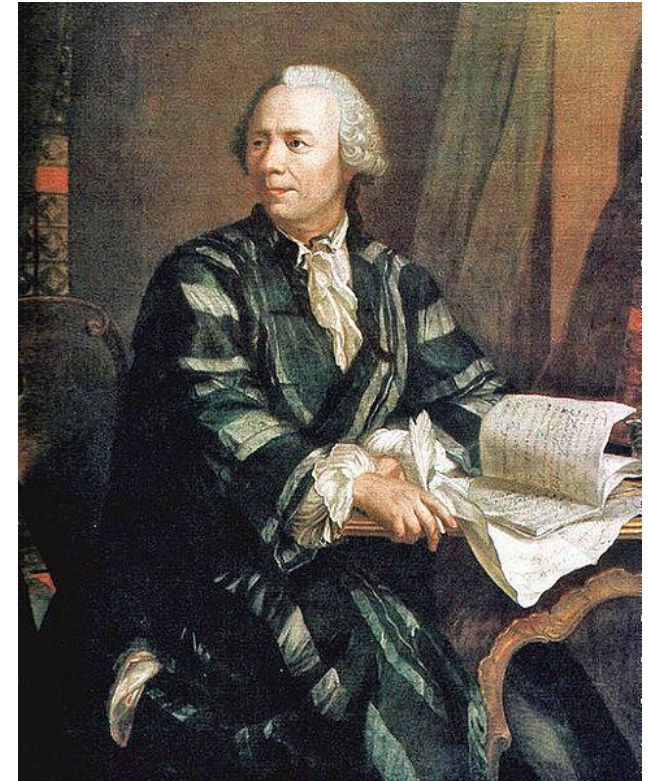
Using the concept of networks: From the 7 bridges of Konigsberg...



Source: Wikipedia

Problem: to find a walk through the city that would cross each bridge once and only once.

Euler: the problem has **no** solution.



Source: Wikipedia

Leonard Euler (1707-1783)

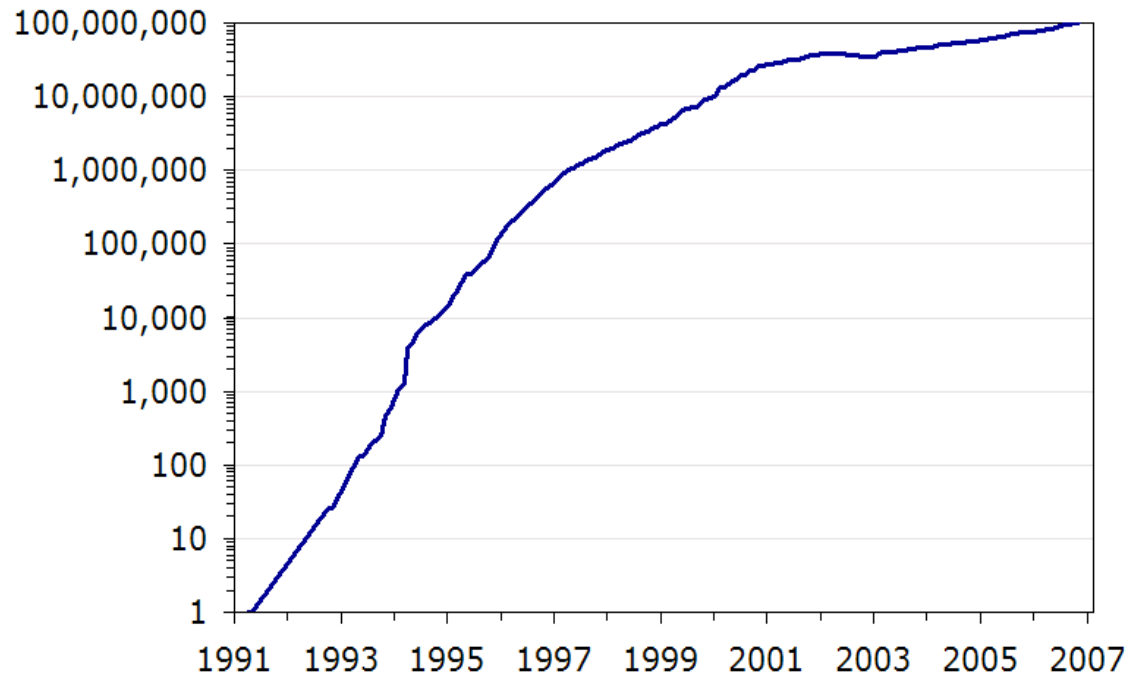
Led to foundation of **Graph theory**: the study of mathematical structures called graphs that model pairwise relations between objects

Using the concept of networks: ...to searching the World Wide Web

When the World Wide Web (WWW) was first proposed, it was widely assumed that its utility was limited because it would be impossible to efficiently access the vast quantity of information distributed throughout the entire network

Like searching for a needle in a haystack most of the time what you will find will be completely irrelevant to what you want !

Growth of number of web pages between 1991-2007



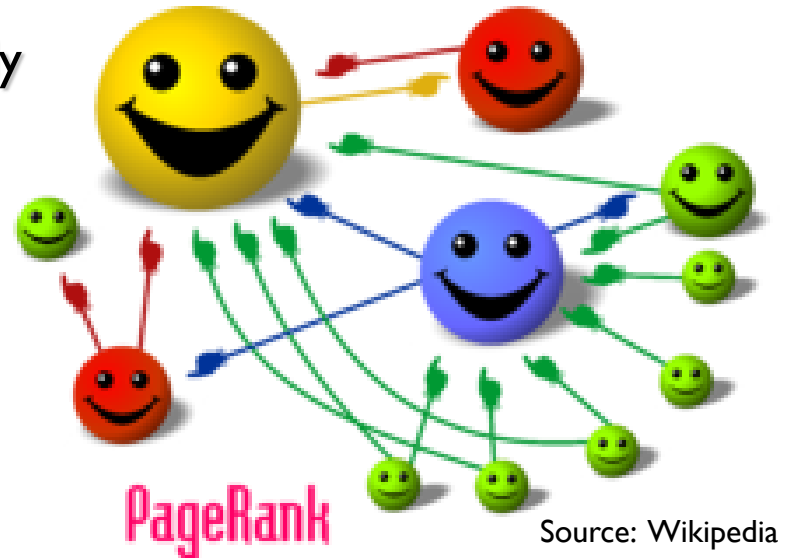
Source: www.useit.com/alertbox/web-growth.html

Using the theory of networks

Google's answer to data deluge

Manually indexed search engines were clearly incapable of handling the exponentially increasing amount of information in the web...

until Google came up with an efficient automated search algorithm that **weights the importance of each page according to how many other pages are pointing to it**



- The page rank algorithm is directly inspired by the **graph theoretic** concept of **node centrality**
- importance of a node measured in terms of how extensively it is connected to other important (i.e., having high centrality) nodes.

The basic idea behind PageRank is that the importance of a page is determined by how many other important pages link to it

If that seems a circular definition of importance, it turns out that one can define it rigorously using linear algebra by using the concept of

Eigenvector centrality

Let us denote the centrality (“importance”) of each node i by x_i – for which we initially just make a guess (let’s say $x_i = 1$ for all i)

We can then improve our guess by noting that $x_i = \sum A_{ij} x_j$ (as the centrality of each node is determined by the centrality of its neighbors, the information about neighborhood being provided by the adjacency matrix \mathbf{A})

Thus our improved guess for the x_i s will be $\mathbf{x}' = \mathbf{A} \mathbf{x}$ (where \mathbf{x} and \mathbf{x}' are vectors)

Repeating the iterative procedure for n times we get $\mathbf{x}(n) = \mathbf{A}^n \mathbf{x}(0)$

We can always express $\mathbf{x}(0)$ as a linear combination of the eigenvectors \mathbf{v}_i of \mathbf{A}
 $\mathbf{x}(0) = \sum c_i \mathbf{v}_i$ (for some appropriate choice of c_i s)

Thus, $\mathbf{x}(n) = \mathbf{A}^n \sum c_i \mathbf{v}_i = \sum c_i \lambda_i^n \mathbf{v}_i = \lambda_{\max}^n \sum c_i [\lambda_i/\lambda_{\max}]^n \mathbf{v}_i$ where λ_i are the eigenvalues of \mathbf{A} with λ_{\max} being the largest eigenvalue.

As, by definition $\lambda_i/\lambda_{\max} < 1$ for all i , all terms in the sum other than the first will become negligible for large n

Thus in the limit of infinite iterations, the centrality is given by $\mathbf{A} \mathbf{x} = \lambda_{\max} \mathbf{x}$

Thinking about biological problems in terms of networks

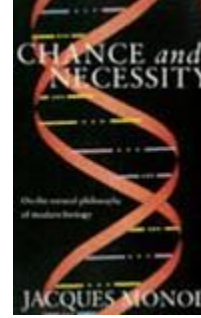
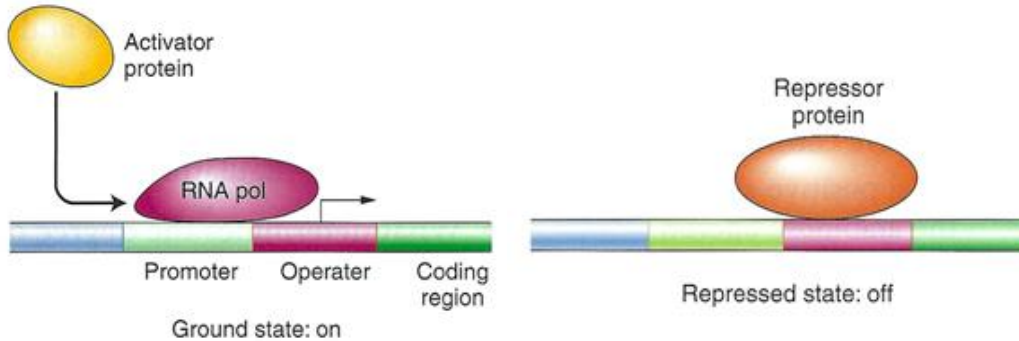
Example:

The *system* of genes in a cell that switch each other ON/OFF resulting in a cellular phenotype

Network of genes controlling each other

Jacob & Monod: Explained the genetic regulation of enzyme synthesis with their work on *lac* operon in *E coli*

Overview of prokaryotic transcription regulation



Jacques Monod
(1910 – 1976)

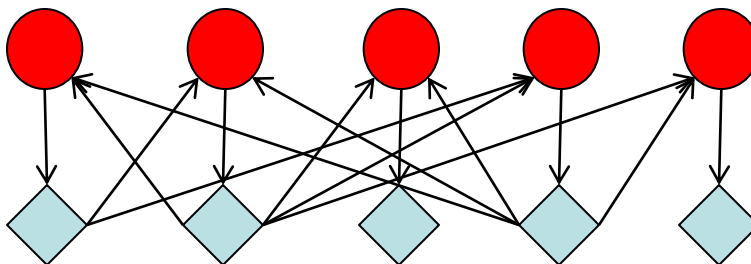


Francois Jacob
(1920 – 2013)

Led to the idea of a **bipartite** network of genes and transcription factors that regulate the expression (switching ON/OFF) of genes: collective activity of the network – expression of different proteins – in a cell differs depending on the cell type (e.g., neuron vs muscle cell)

Genes

TFs



Random Boolean Network

Model for explaining how the collective switching of genes can lead to different cellular phenotypes

S A Kauffman (1969) Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of Theoretical Biology*, 22:437-467.

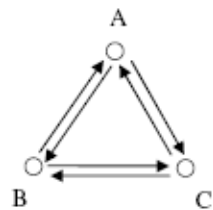
N Boolean variables each of whose states are determined by a boolean function taking inputs from a subset of size K variables



www.anisn.it

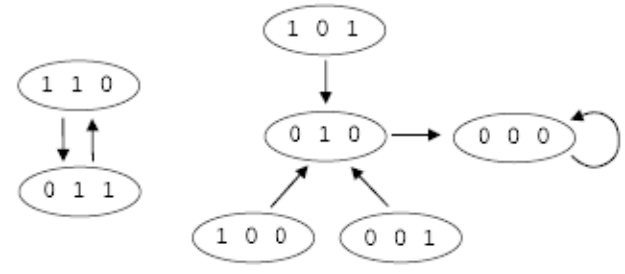
Stuart Kauffman (1939 -)

NK Network (completely connected) with $N=3$ and $K=2$

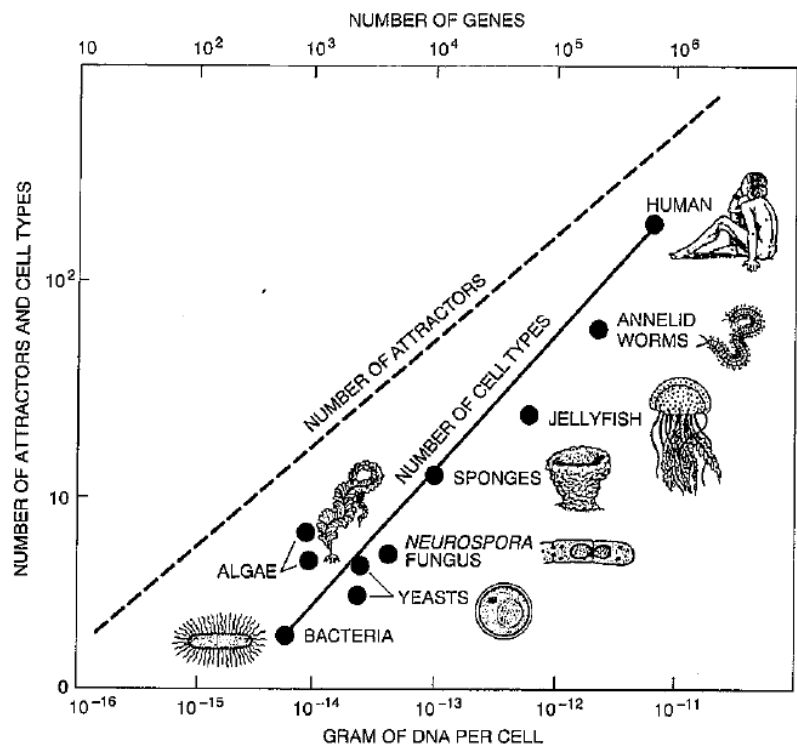


A	B	C	→	A	B	C
0	0	0		0	0	0
0	0	1		0	1	0
0	1	0		0	0	0
0	1	1		1	1	0
1	0	0		0	1	0
1	0	1		0	1	0
1	1	0		0	1	1
1	1	1		1	1	1

Transition Rules
 $A = \text{AND}(B \ \& \ C)$
 $B = \text{OR}(A \ | \ C)$
 $C = \text{AND}(A \ \& \ B)$



wiki.cas-group.net

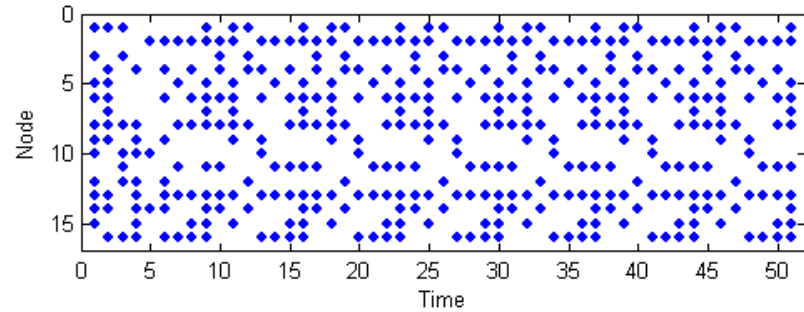


Kauffman, Scientific American (1991)

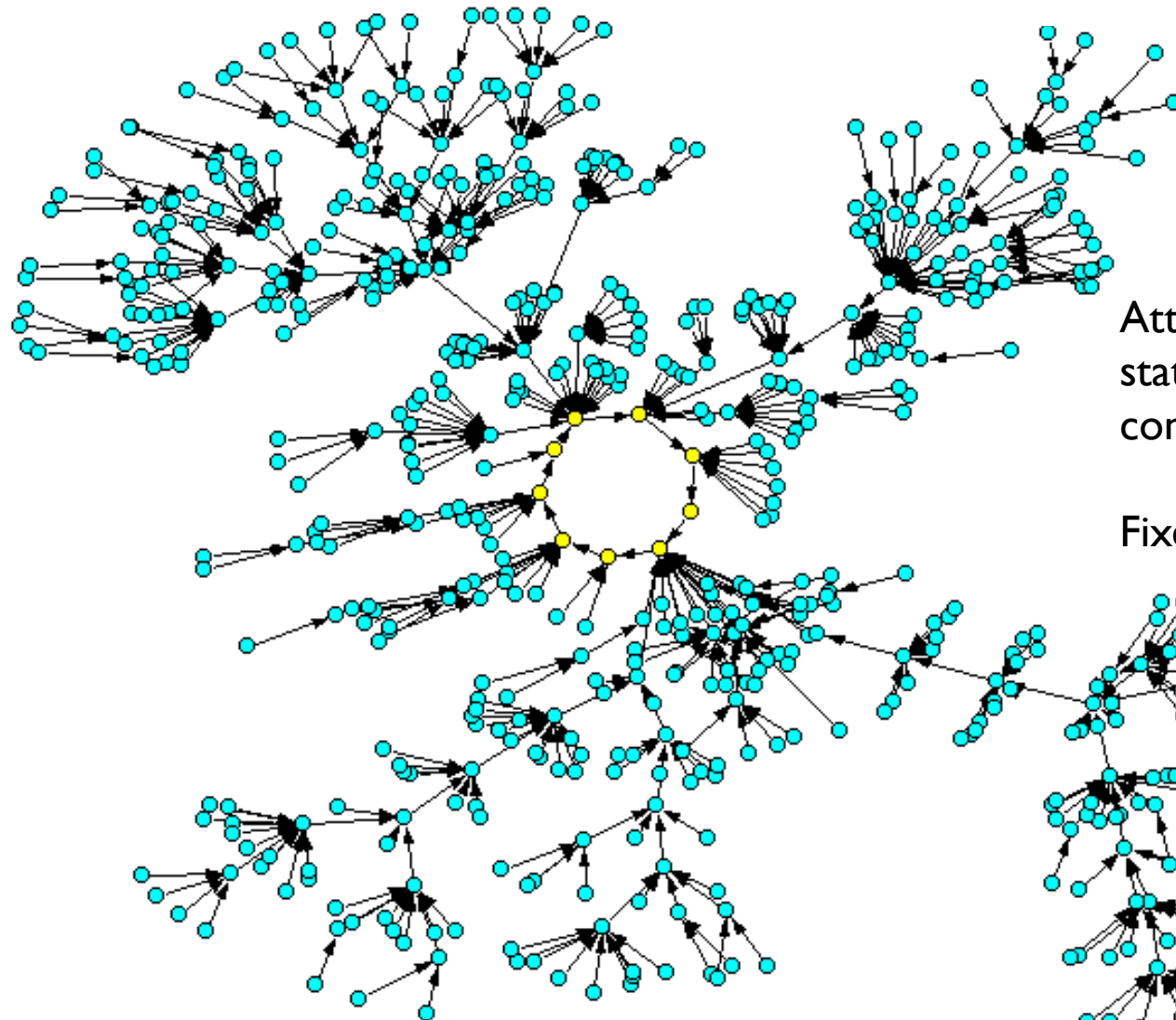
The set of functions defines a directed network on N nodes

2^N possible network states for binary variables

Attractors of the Boolean network dynamics can be periodic

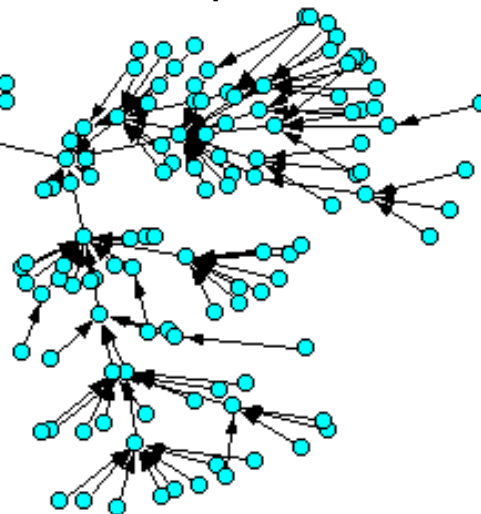


<http://fias.uni-frankfurt.de/~willadsen/RBN/>



Attractor is a set of dynamical states to which the system converges asymptotically

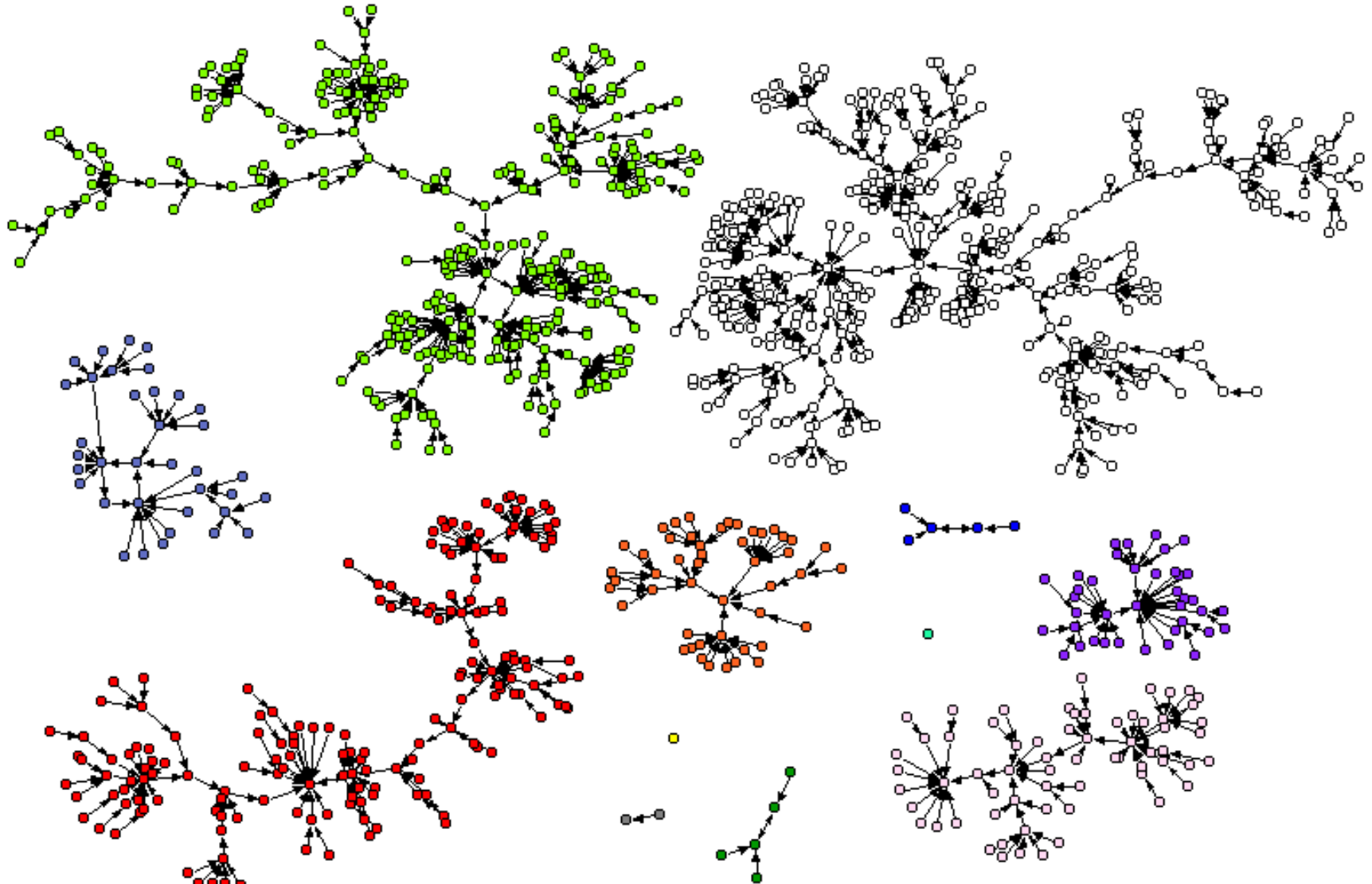
Fixed point is Period-1 attractor

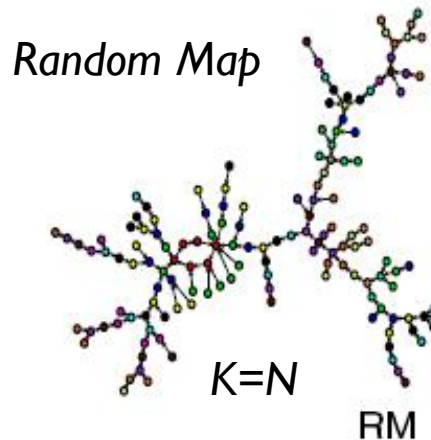
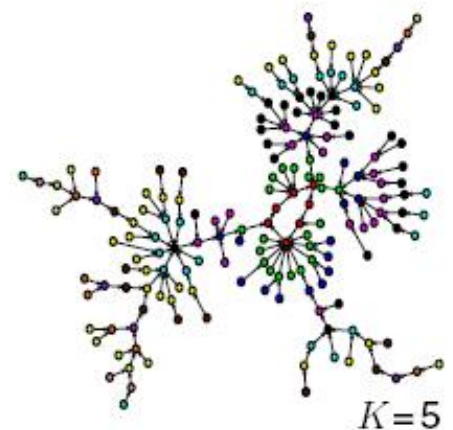
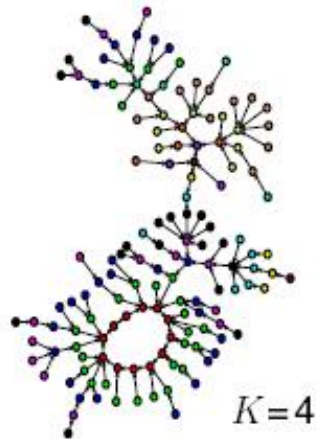
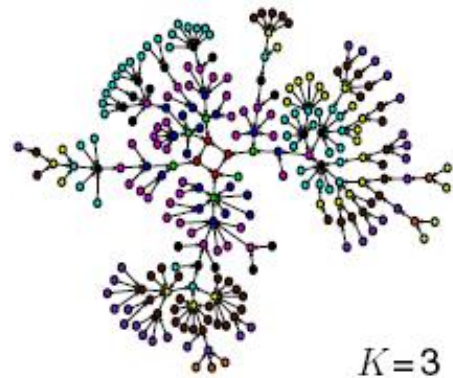
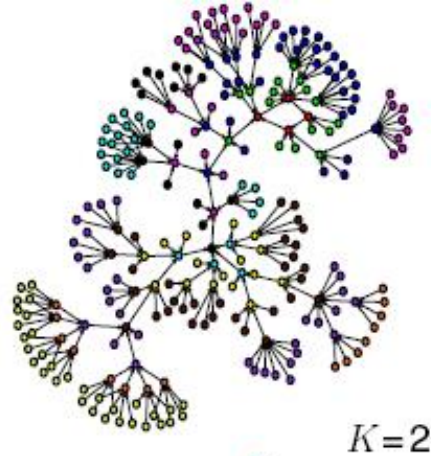
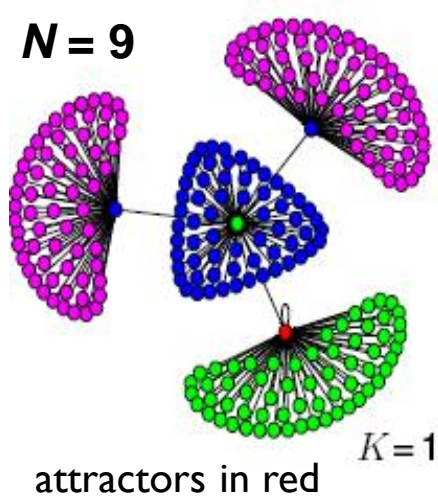


<http://fias.uni-frankfurt.de/~willadsen/RBN/>

A Boolean Network can have multiple attractors

Each attractor state is defined by a basin of attraction corresponding to states which flow to the attractor as a result of time-evolution of the Boolean network dynamics





Complex Network Analysis of Random Boolean Networks

A directed state space network (SSN) of a Boolean network of N elements can be constructed by considering each of the 2^N dynamical states as a node, and links as connections to the next state occurring in the time-evolution of the system

“Garden of Eden” (GoE) configurations which can only occur as initial states and cannot be obtained as a result of time evolution starting from another state \equiv SSN nodes with in-degree zero

- ❑ For $K = 1$ SSN all nodes are either GoE states or hubs, all hubs having same in-degree \Rightarrow “modular star networks”
- ❑ The SSN of $K=N$ Boolean Network is a ER random graph with $\langle k \rangle = 1$

As K increases, SSN transits from connected stars to branched structures

Connecting to Biology: the Yeast Regulatory Network

S. cerevisiae

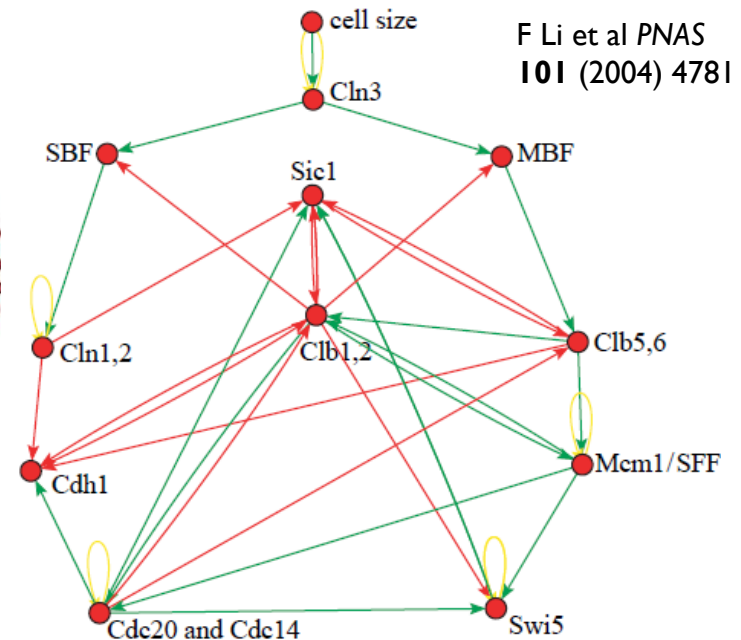
Interactions are classified into:

- activation (green)
- repression (red)

The gene states can be considered as binary:

- Expressing or ON (1)
- Not expressing or OFF (0).

Boolean network model for Yeast cell cycle control



Robustness of the Yeast cell-cycle as a network property

Cell-cycle: process by which one cell grows and divides into two daughter cells

Consists of 4 phases: G_1 (cell grows and, under appropriate conditions, commits to division), S (DNA is synthesized and chromosomes replicated), G_2 (a “gap” between S and M), and M (chromosomes are separated and the cell is divided into two). After the M phase, the cell enters the G_1 phase, hence completing a “cycle.”

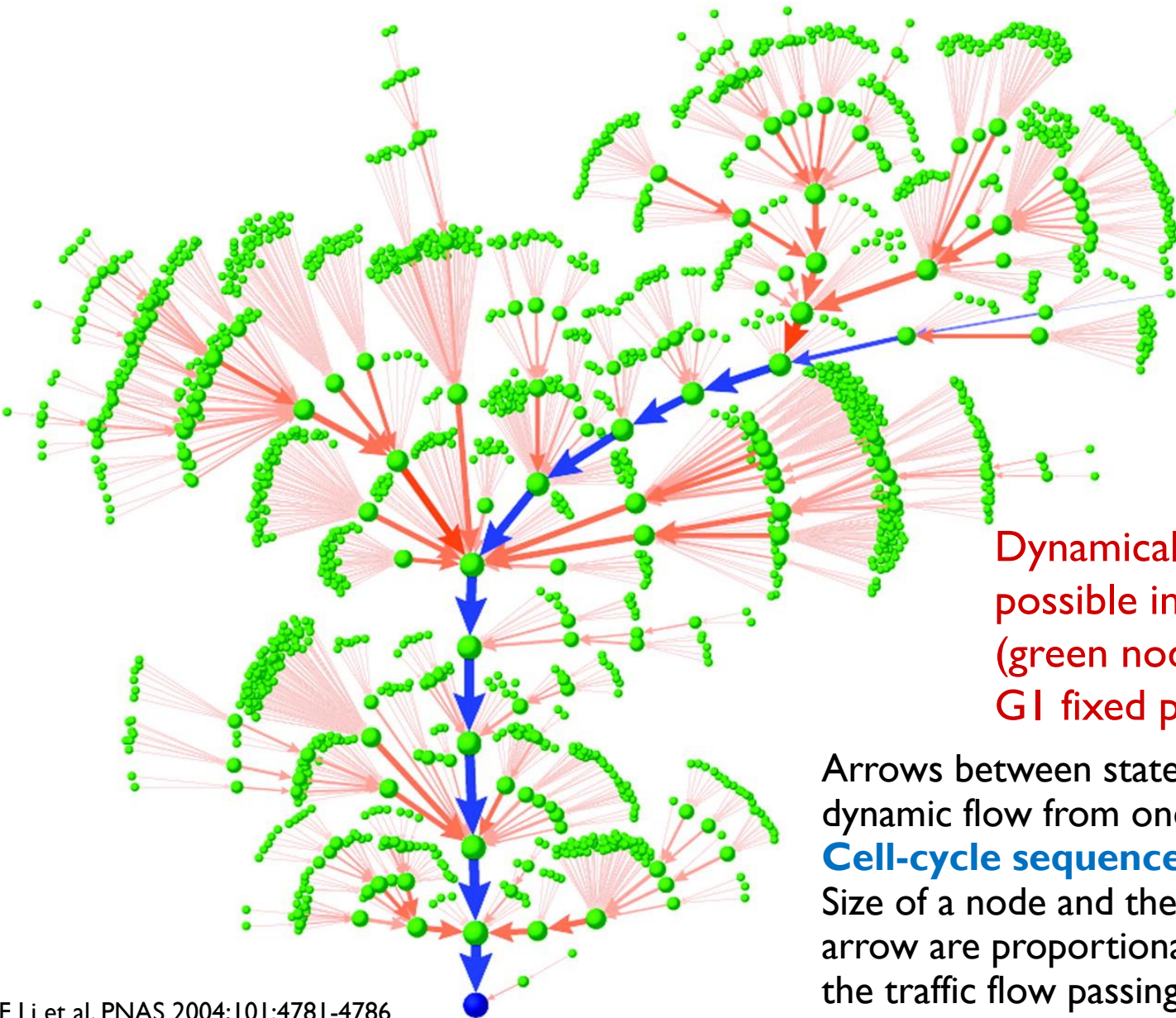
Fixed points of the Yeast cell-cycle network dynamics

The protein states of the fixed point with largest basin correspond to G_1 state.

Basin size	Cln3	MBF	SBF	Cln1,2	Cdh1	Swi5	Cdc20	Clb5,6	Sic1	Clb1,2	Mcm1
1764	0	0	0	0	1	0	0	0	1	0	0
151	0	0	1	1	0	0	0	0	0	0	0
109	0	1	0	0	1	0	0	0	1	0	0
9	0	0	0	0	0	0	0	0	1	0	0
7	0	1	0	0	0	0	0	0	1	0	0
7	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	1	0	0	0	0	0	0

2048 total possible states

Robustness of the Yeast cell-cycle as a network property



Dynamical trajectories of 1764 possible initial protein states (green nodes) flowing to the G1 fixed point (blue node).

Arrows between states indicate direction of dynamic flow from one state to another. **Cell-cycle sequence is colored blue.** Size of a node and the thickness of an arrow are proportional to the logarithm of the traffic flow passing through them.