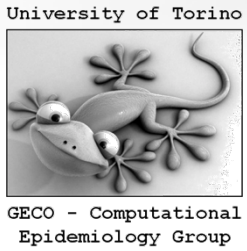
Course of Bioinformatics
Master in Cellular and Molecular Biology
University of Torino

Networks and models of gene regulation



Mario Giacobini

GECO - Computational Epidemiology Group,
Department of Veterinary Sciences

CBU - Computational Biology Unit, Molecular Biotechnology Center

ARCS - Applied Research on Computational Complex Systems Group,
Department of Computer Science

University of Torino, Italy

Artificial life

Artificial life ***complements the traditional analytic approach*** *of traditional biology* ***with a synthetic approach*** *in which, rather than studying biological phenomena by taking apart living organisms to see how they work, one attempts to put together systems that behave like living organisms.*

Alife [...] will result in not only ***better theoretical understanding*** *of the phenomena under study, but also in* ***practical applications of biological principles*** *in many technological domains.*

Christopher G. Langton

Discrete dynamical networks

when using networks to model technological, social, and biological phenomena, we can let **each node assume a discrete value** determining its state

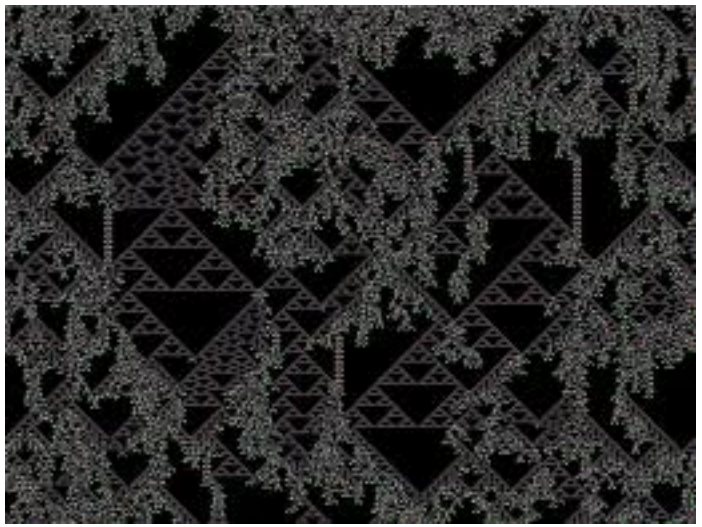
the ensemble of the states of all nodes of the graph, usually called **configuration of the network**, can be interpreted as the state of the whole system under analysis

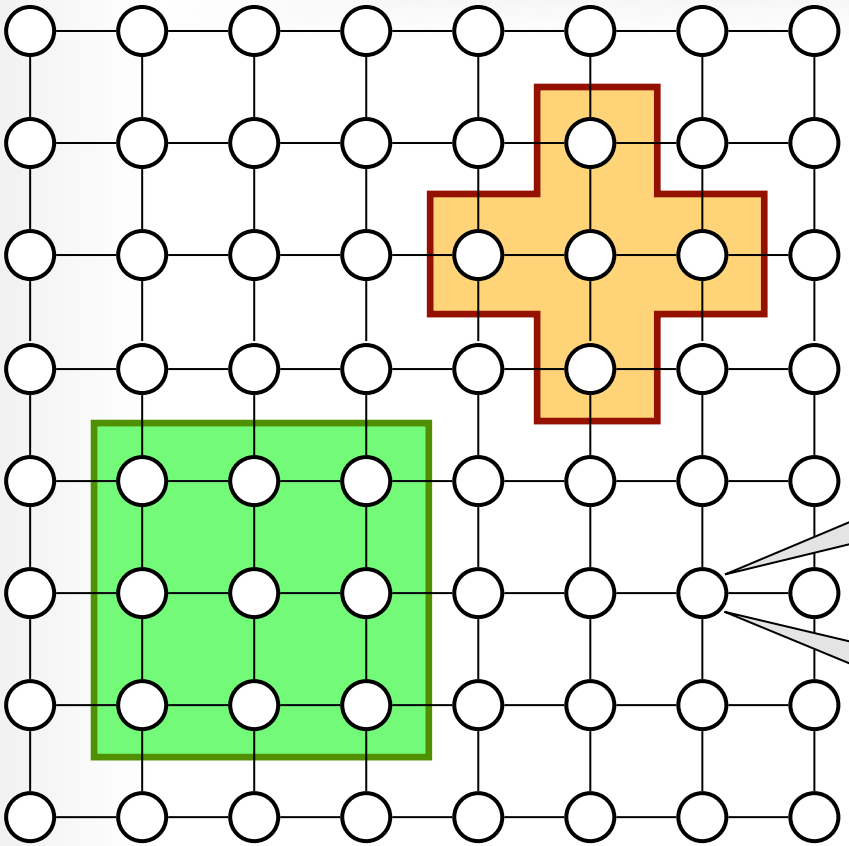
the system can thus **synchronously or asynchronously evolve in discrete time** by letting each node change its state according to a local function (i.e. a function those input values are the state of the neighbors of the considered nodes)

Cellular automata

proposed by Ulam and von Neumann in the 40s, cellaular automata (CAs) are dynamical systems where both space and time are discrete:

- ✓ a set of cells disposed on a regular grid of dimension *d* (usually *d = 1, 2, 3*)
- ✓ at each time, each cell can have a state in a finite set of possible state
- ✓ each cell has a neighborhood
- ✓ the dynamics of the system is defined by a local transition rule

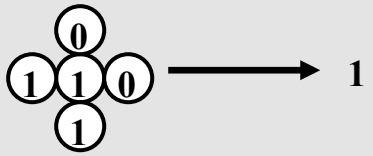




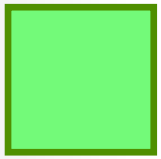
a bidimensional binary CA

each cell may have a state
in the set $S = \{0,1\}$

the state of a cell at the next
time step is determined
by a local transition rule



von Neumann



Moore

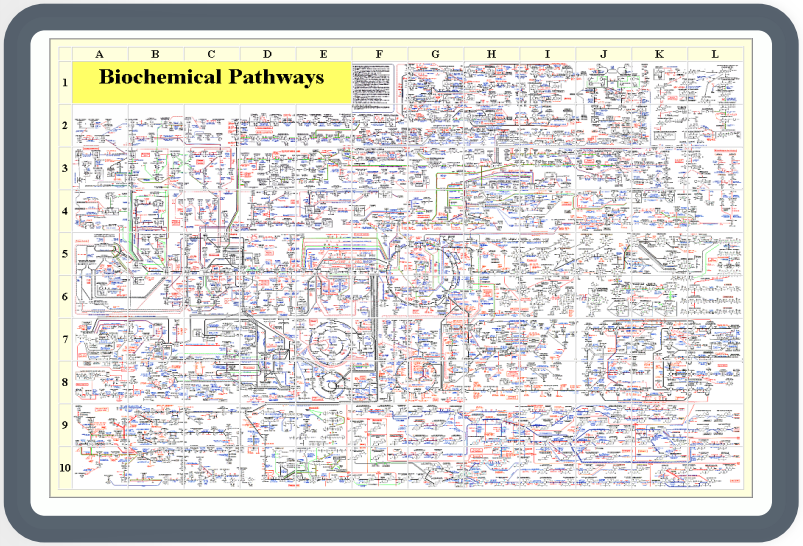
The game of life

in a bidimensional grid world with Moore's neighborhood, each cell may contain an organism (state 1) or not (state 0)

the evolution of such world is performed by using the following local transition rule:

- ✓ **survival:** each organism in a cell with two or three organisms in its neighborhood stays alive in the next time step
- ✓ **death:** each organism in a cell with four or more organisms in its neighborhood dies (for sparsity of resources), as each organism in a cell with one or no organism (for isolation)
- ✓ **birth:** each empty cell with three organisms in its neighborhood will contain an organism in the next time step

From gene regulation to models

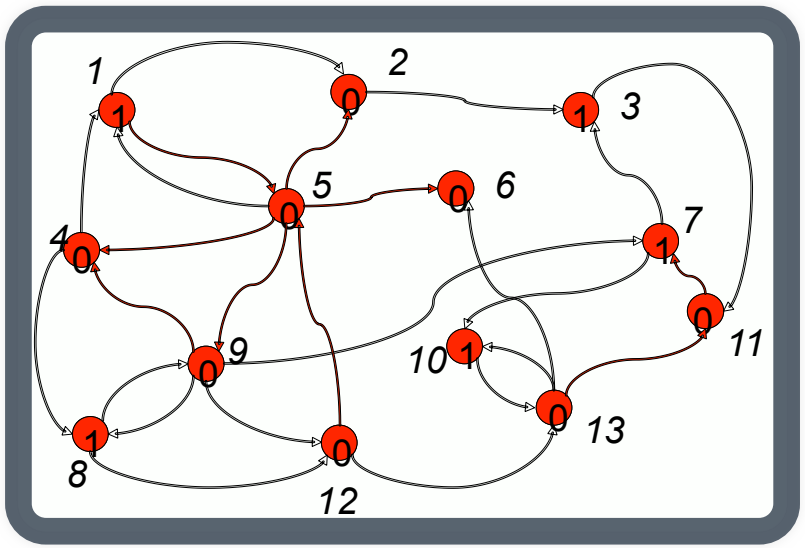


molecular biology

gene expression is determined by a combination of regulatory proteins

simplified model

the activation of a gene, that results in a protein, influences the activation of other genes



Random Boolean Networks

- random boolean networks (RBNs) were proposed by Stuart Kauffman in 1969 as an extremely simplified model to study the dynamics of networks of gene regulation

network:

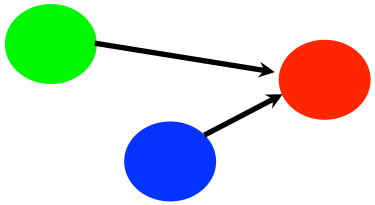
- ✓ a gene regulation network is seen as a set of N nodes, each representing a gene
- ✓ in the original model, each gene has fixed out-degree (K)
- ✓ the $K \times N$ edges represent the influence in regulation dynamics of a gene on another

boolean:

- ✓ each gene is modeled as a unit that can be active/activated (expressed gene) or not/repressed (unexpressed gene)
- ✓ each node can therefore assume two states (binary node)

aleatoria:

- ✓ for each gene, the K outgoing edges are directed to other K randomly chosen nodes: the networks therefore follows the Erdös-Rényi model with fixed out-degree
- ✓ the local transition function assigned to each node is randomly generated as a function of its in-degree and of a probability p of a node to be expressed



input	stato
off / off	on
off / on	off
on / off	off
on / on	off

Dynamics of random boolean networks

- at each discrete time step, each node changes its state according to its local transition rule
- in the original model, an RBNs determine the state of all its nodes for the following discrete time step according to a synchronous scheme
- the state or configuration $S(t)$ of an RBN at time t is the set of the states $s_i(t)$ of its singles nodes at time t:

$$S(t) = \{s_1(t), s_2(t),..., s_N(t)\}$$

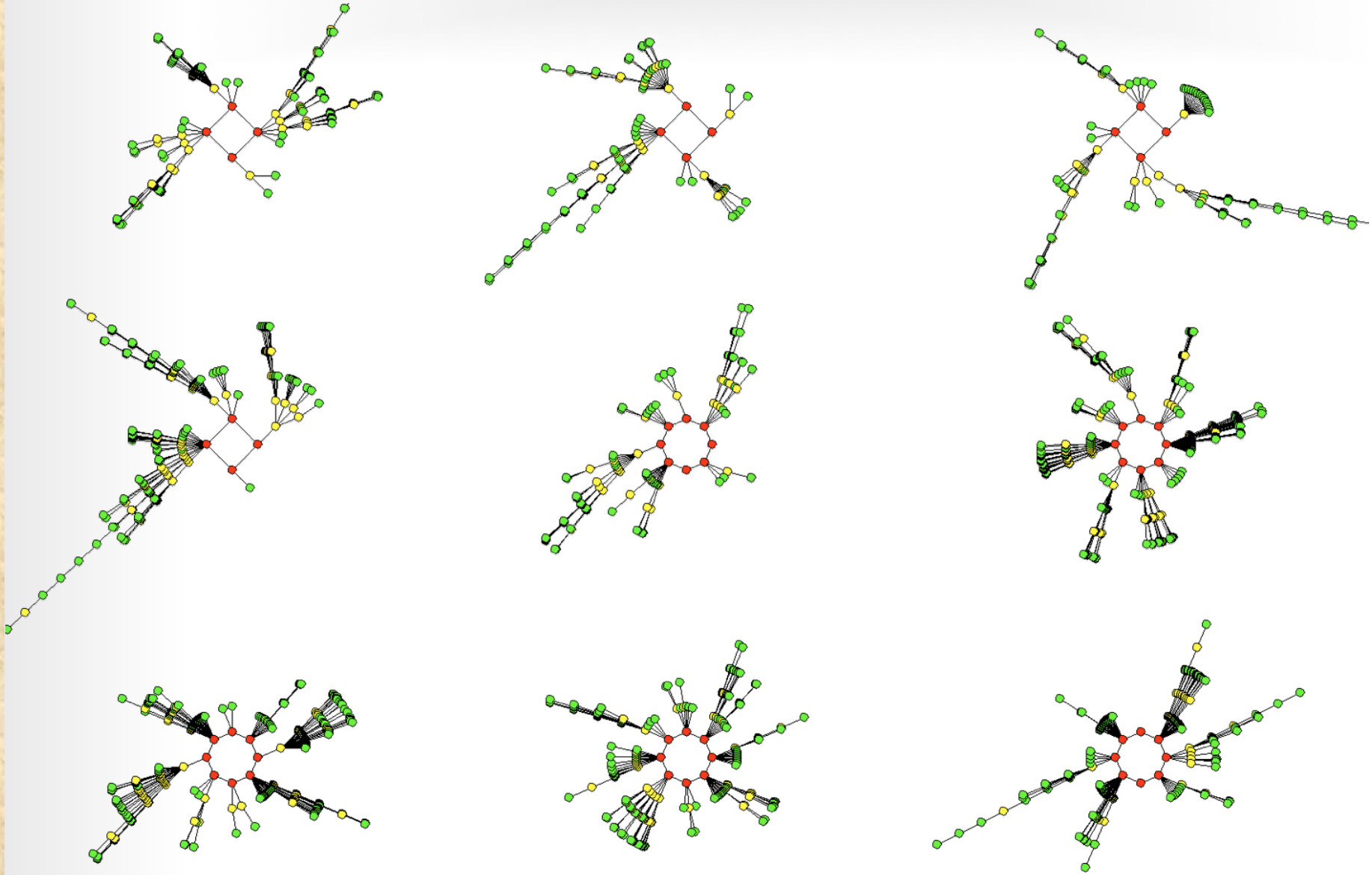
- each configuration $S(t)$ of an RBN has a unique state $S(t+1)$ to which it can evolve
- an RBN follows therefore a deterministic dynamics

- the set of all possible configurations of an RBN coupled with all the transitions among them is called state space

- the state space of an RBN with N nodes is composed by 2^N points (nodes), linked by directed edges representing the transitions among states

- for an RBN, if the number of nodes N is sufficiently small, it is feasible to study in an exhaustive way the topology of its state space

- there exist 3 kind of states in a state space:
  - ✓ ***gardens of Eden***: states that can not be reached from other states
  - ✓ ***transients***: intermediate states
  - ✓ ***attractors***: single states (point attractors) or sets of states (cyclic attractors) absorbing the dynamics of an RBN



the transients and gardens of Eden of an attractor are called **basin of attraction**

Dynamical regimes

ordered regime

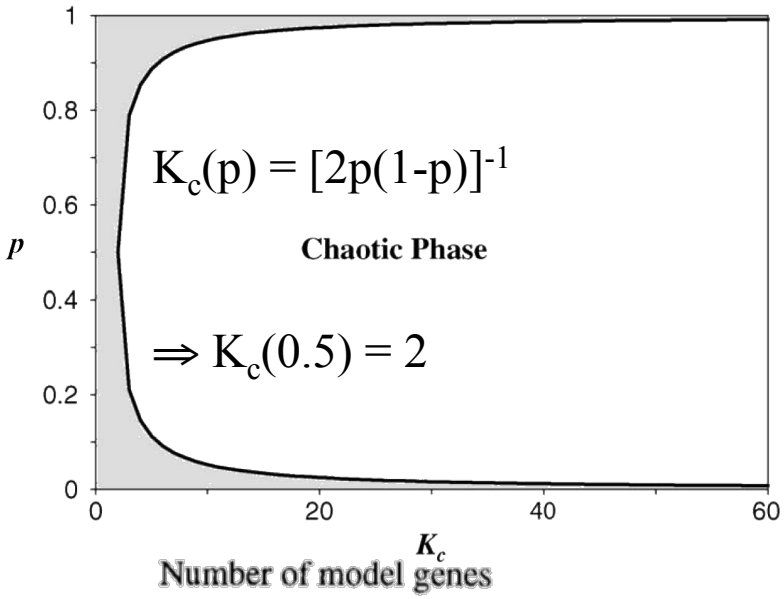
- attractors are stable to small perturbations of nodes' states
- attractor period lengths scale as a small power of the number of nodes N of the RBN

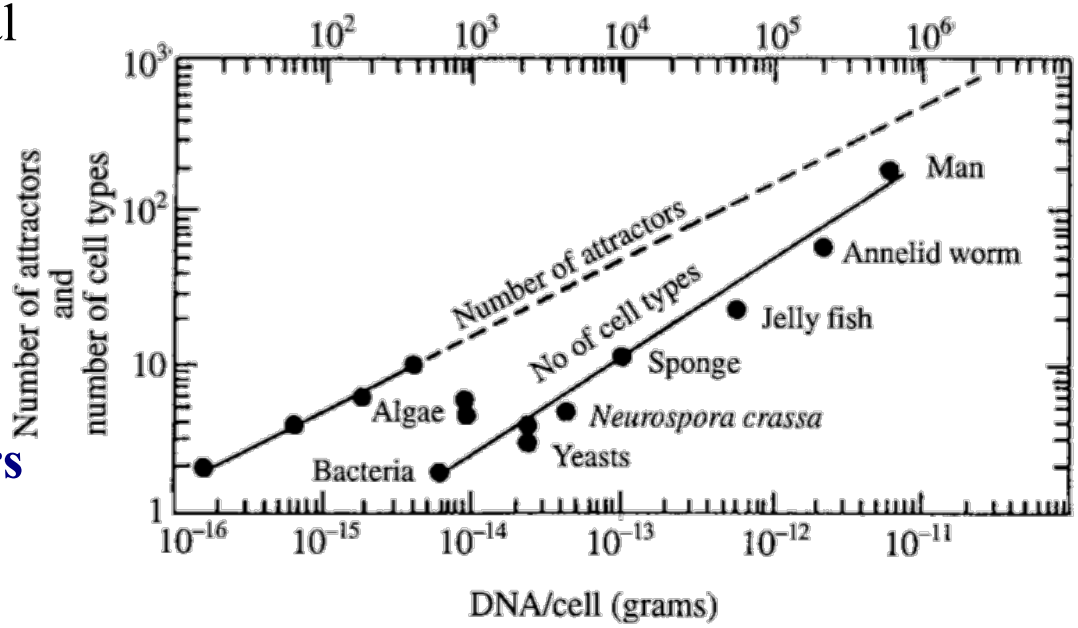
chaotic regime

- attractors are sensible to initial conditions of the dynamics
- attractor period lengths grow exponentially with the number of nodes N of the RBN

in the ordered regime, at the edge of chaos, lies the **critical regime**

- ✓ systems in the critical regime show a good compromise between stability to small perturbations and the possibility to evolve (***evolvability***)
- ✓ according to Kauffman hypothesis **living organisms** lie in critical regimes
- ✓ different **point attractors** in the state space of an RBN are interpreted as cell types
- ✓ **cyclic (or periodic) attractors** correspond to cellular cycles





Kauffman's main assumptions

1. the nodes contain a Boolean value and not a graded function

   this is an acceptable approximation for *threshold phenomena*

2. the connection topology is a random graph

   this does not agree with nowadays knowledge on biological networks

3. the dynamics of the system is synchronous

   this does not agree with experimental data on gene expression in gene regulatory networks

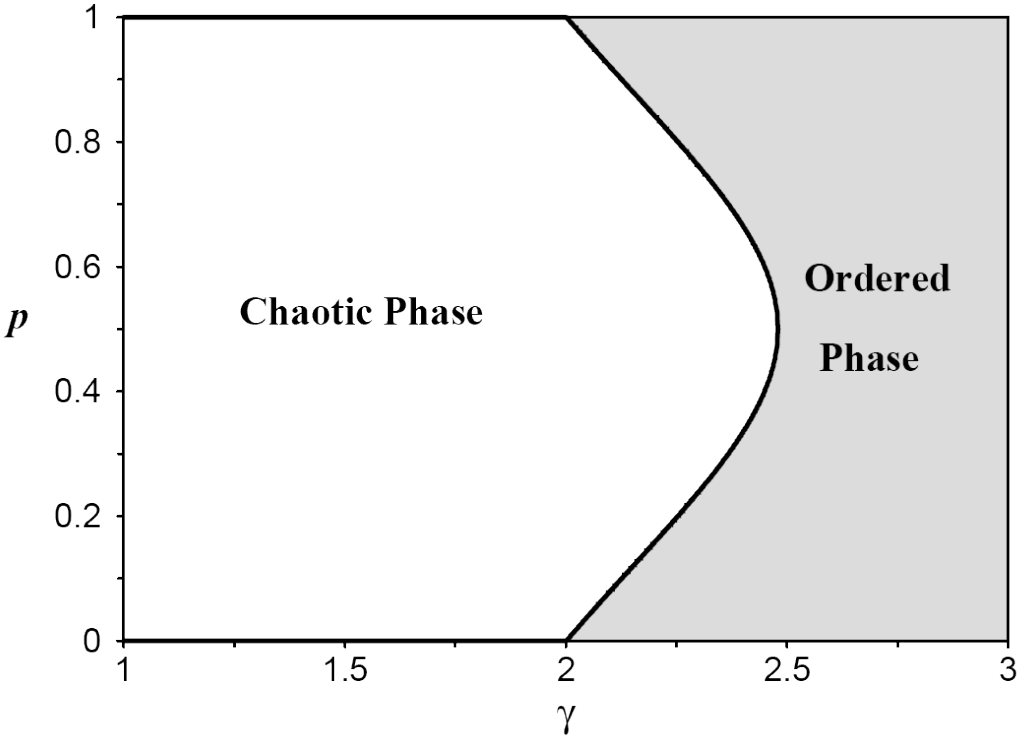
4. the local transition function is random

   this is not acceptable given recent discoveries on the activation functions of genes

RBNs with scale-free topologies

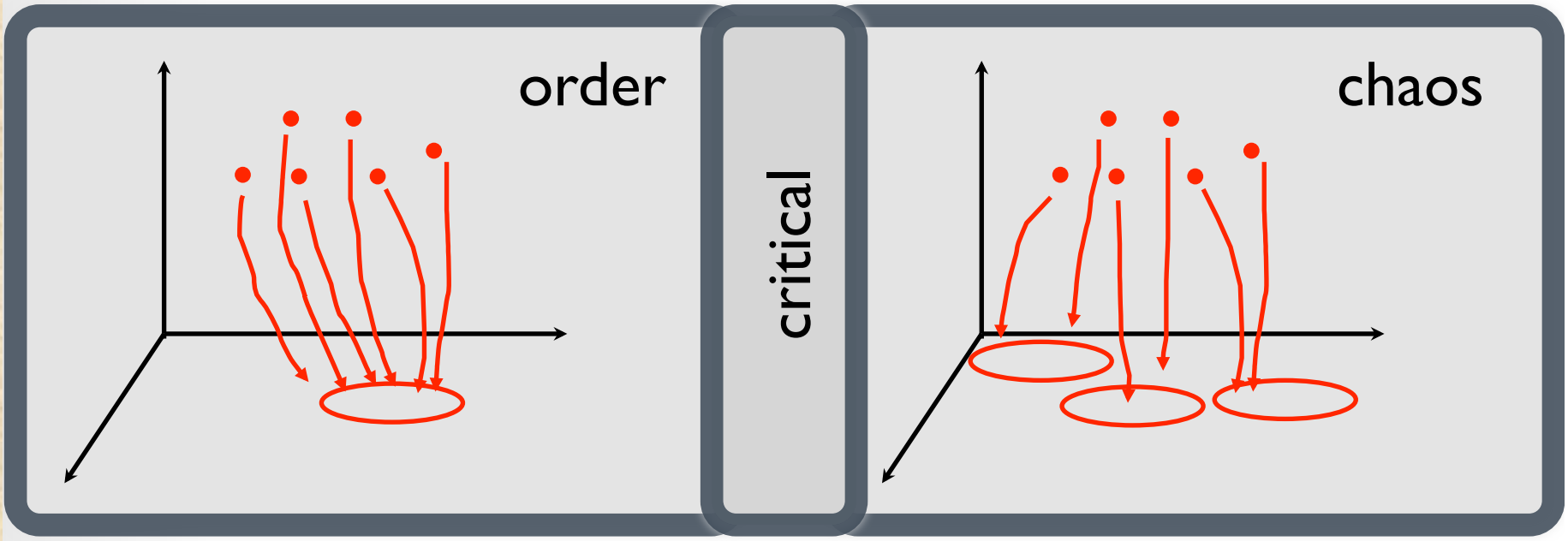
according to recent studies, several genetic regulatory networks show a Poissonian distribution of the in-degrees, and a scale-free distribution of the out-degrees

Aldana [2003] analysed the phase transition of synchronously updated RBNs with *scale-free* topology as a function of the outdegree power-law exponent γ

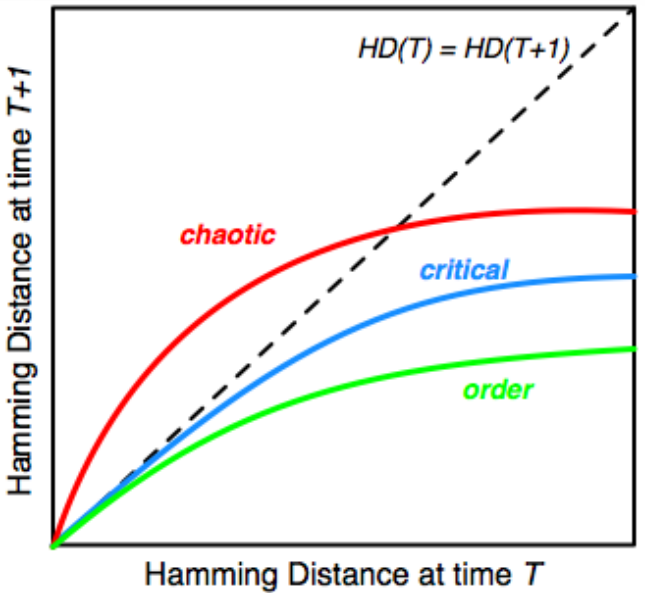


Tolerance to perturbations

- they model transcription errors of gene *knock outs*
- they consist in errors in the output values of nodes' transition functions
- their impact on the dynamics of the system can be measured
  - ✓ does the system converge to different attractors?



Derrida plot



- it is possible to study the dynamical regime of an RBN using the Derrida plots
- a Derrida plot is a two-dimesional graphic where points have:
  - ✓ as x-coordinate the Hamming distance (HD) between two configurations at time *t*

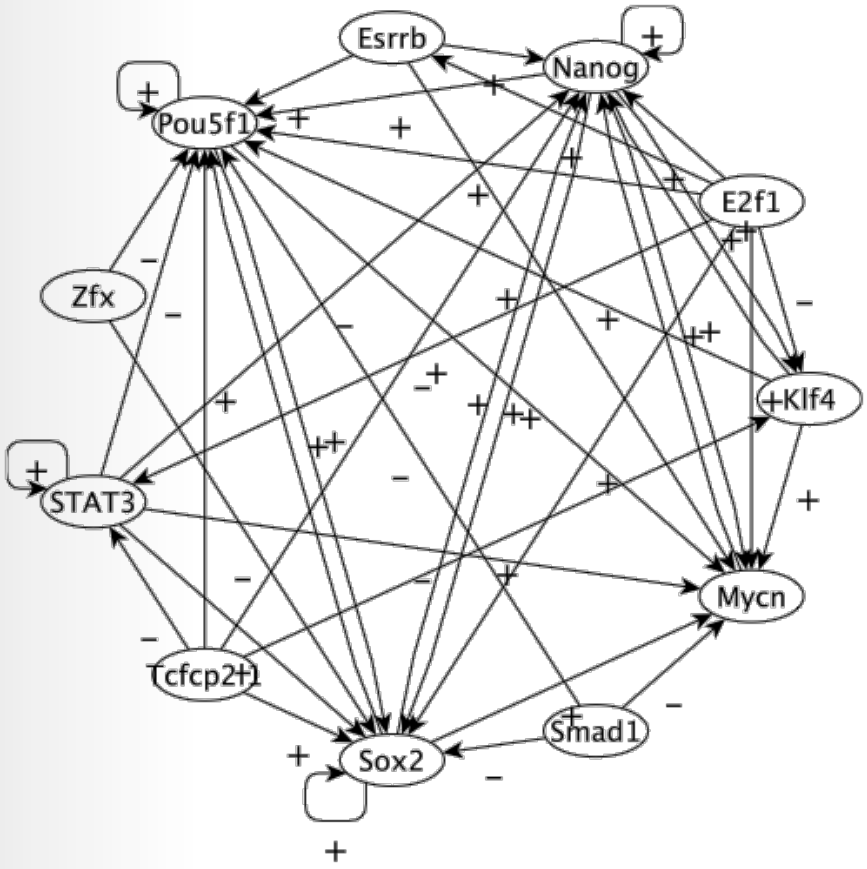
$$HD(S_a(t),S_b(t))$$

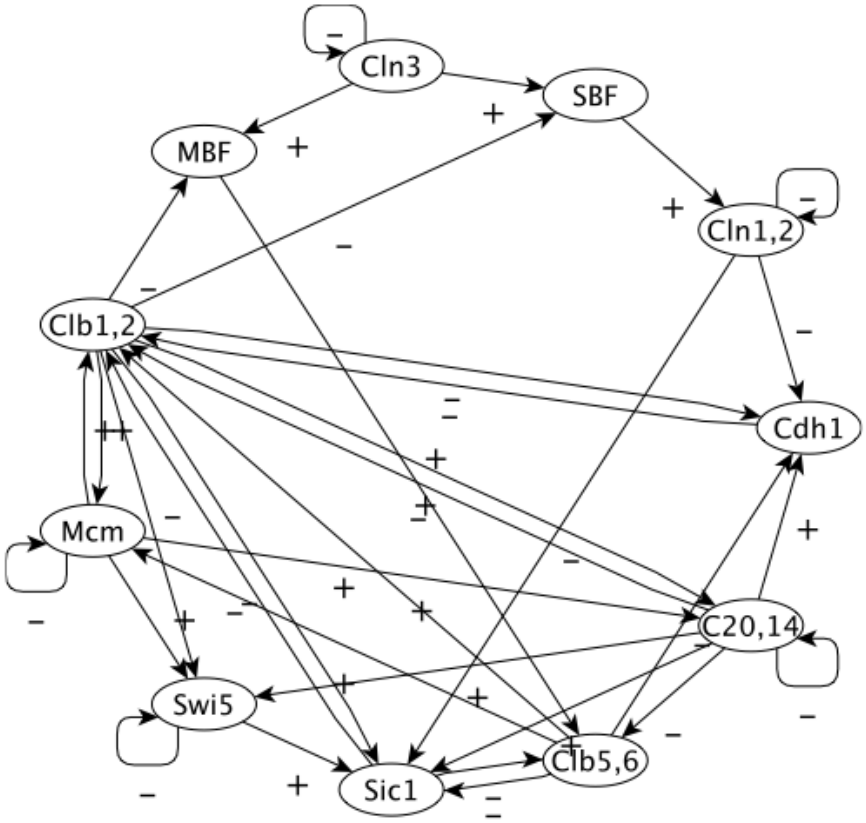
  - ✓ as y-coordinate the HD between the two configurations at time t+1 evolved from those considered for the x-coordinate

$$HD(S_a(t+1),S_b(t+1))$$

The two case studies

mouse embryonic stemm cells

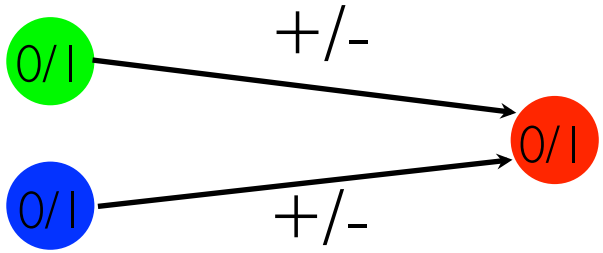




yeast cell cycle

Update functions

- real-life update function often unknown, thus inspired by Stoll *et al.*, we propose a new one that :
  - ✓ uses the extra information and is closer to biological reality
  - ✓ applies a threshold to regulate the gene expression value
  - ✓ takes into account the promoting and repressing factors
  - ✓ reduces the number of possible functions to one per node, making models state space exhaustively portrayable



Additive update functions

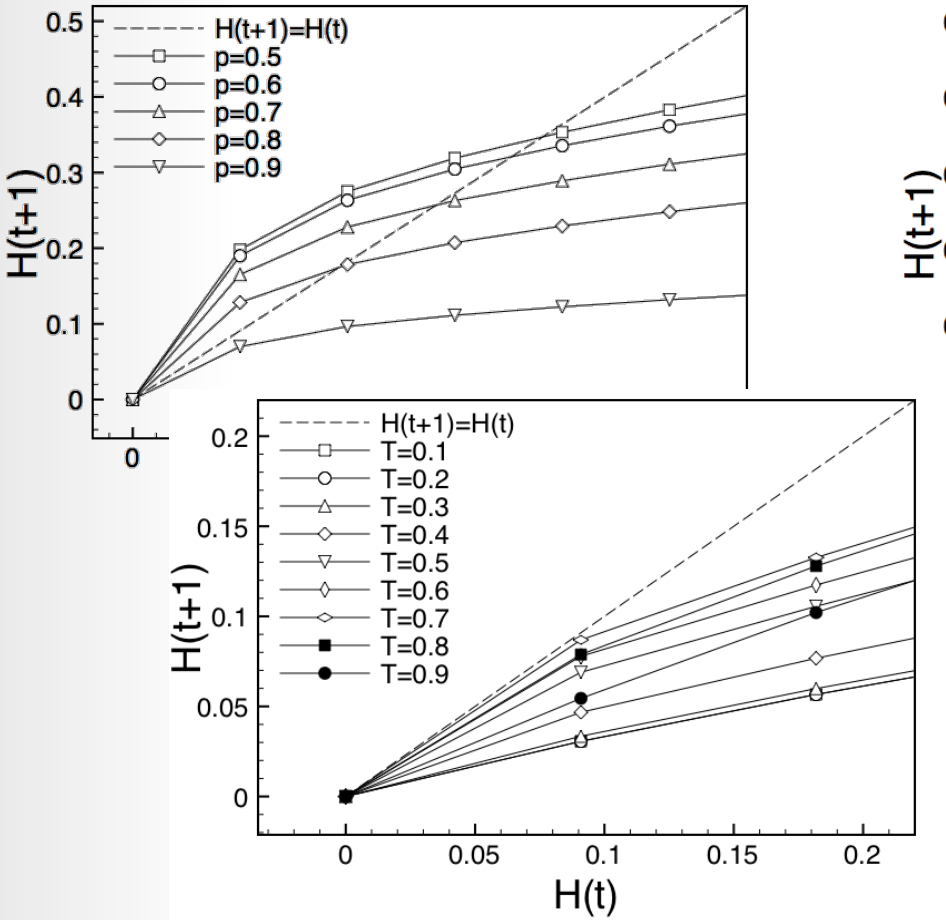
$$S_i(t+1) = \begin{cases} active\ (1) & \text{if } \sum_j S_j^+ > T \times (\sum_j S_j^+ + \sum_j S_j^-) \\ inactive\ (0) & \text{if } \sum_j S_j^+ < T \times (\sum_j S_j^+ + \sum_j S_j^-) \\ S(t) & \text{otherwise} \end{cases}$$

where:

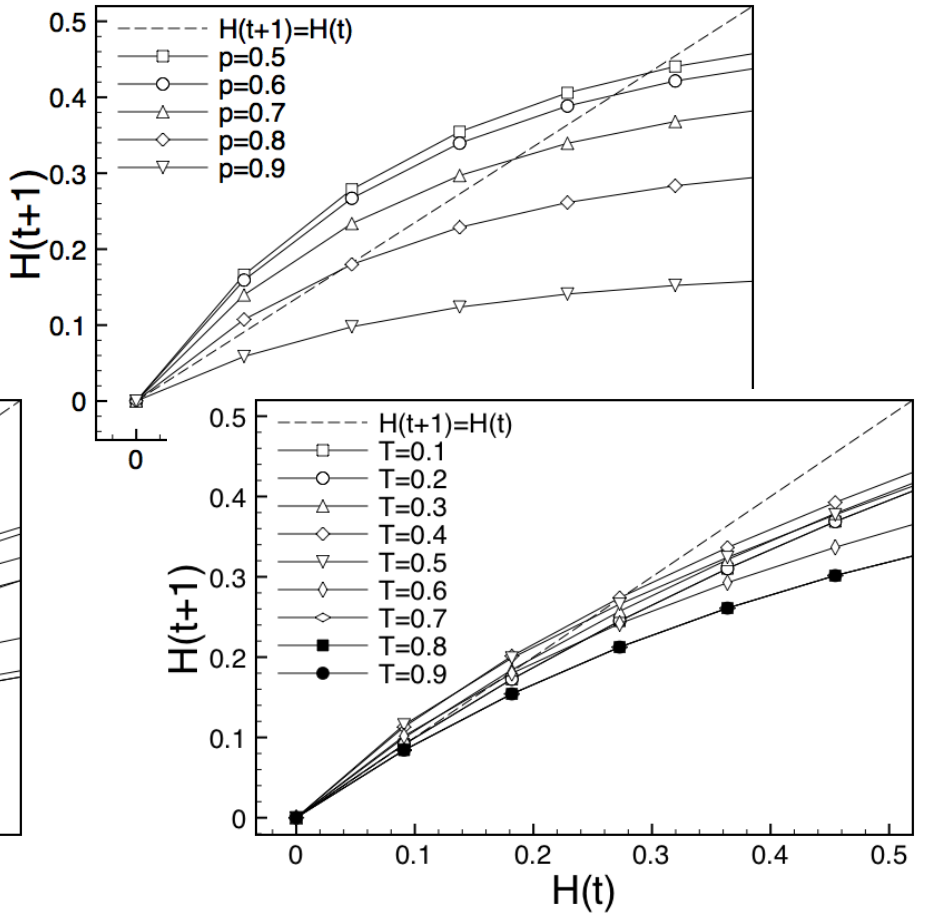
- S^+ represents the state of an activator gene
- S^- represents the state of an repressor gene
- T is a threshold value that gives different weights to + and -

Derrida plots of the two case studies

mouse embryonic stemm cells



yeast cell cycle



Essentially, all models are wrong, but some are useful.

George E.P. Box, 1987