Introduction to Random Boolean Networks

Carlos Gershenson
Centrum Leo Apostel, Vrije Universiteit Brussel.
Krijgskundestraat 33 B-1160 Brussel, Belgium
cgershen@vub.ac.be
http://homepages.vub.ac.be/~cgershen/rbn/tut
http://rbn.sourceforge.net

Topics (1)

- Introduction
- Classical Model (Kauffman)
  - Order, Chaos, and the Edge
    - Phase transitions in RBNs
  - Explorations
    - Attractor lengths
    - Convergence
  - Multi-valued Networks
  - Topologies
  - RBN Control

Topics (2)

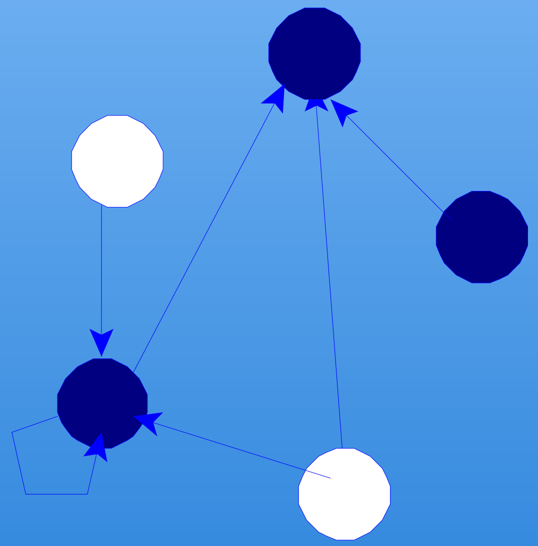
- Different Updating Schemes
  - Asynchronous RBNs
    - Rhythmic ARBNs
  - Deterministic Asynchronous RBNs
    - Thomas' ARBNs
  - Mixed-context RBNs
  - Classification of RBNs
- Applications
- Tools
- Future Lines of Research
- Conclusions

Introduction

- RBNs originally models of genetic regulatory networks (Kauffman, 1969; Kauffman, 1993)
- Random connectivity and functionality
  - Useful with very complex systems
- Possibility to understand holistically living processes (e.g. for disease treatment)
- Possibility to explore *possibilities* of living and computational systems.

Classical Model

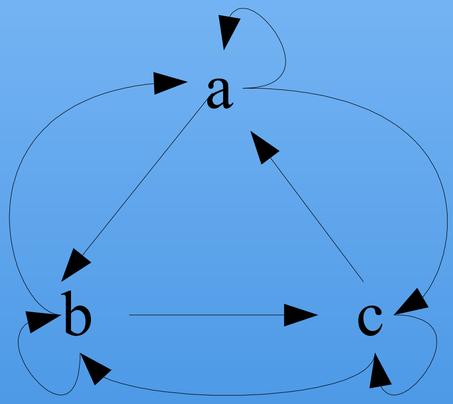
- Generalization of boolean CA
- *N* boolean nodes, *K* connections per node
- Connectivity and logical functions generated randomly
- Synchronous updating

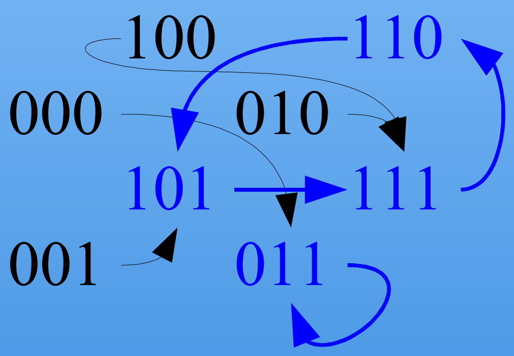


Example

N=3, K=3

abc(t)	abc(t+1)
000	011
001	101
010	111
011	011
100	111
101	111
110	101
111	110





- Finite states (2^N) => **attractors** (dissipative system)
  - Point and cycle attractors

Computational "Difficulties"

- Practically infinite possible networks
  - 2^2^K possible logic functions per node
  - N!/(N −K)! possible ordered combinations per node

$$possible\ nets = \left(\frac{2^{2^K} N!}{(N-K)!}\right)^N$$

  - though many equivalent (Harvey and Bossomaier, 1997)
- Can extract general properties with statistical samples
  - All possible initial states but small networks OR
  - Large networks but few initial states

Order, Chaos, and the Edge (1)

- Neighbouring nodes in lattice
  - If changing, green; if static, red
  - Order: few green “islands”, surrounded by a red “frozen sea”
  - Chaos: green sea of changes, typically with red stable islands
  - Edge: green sea breaks into green islands, and the red islands join and percolate through the lattice (Kauffman, 2000, pp. 166-167)

Order, Chaos, and the Edge (2)

- Network stability
  - "sensitivity to initial conditions"
  - "damage spreading"
  - "robustness to perturbations"
- Make a change in a state, connection, or rule, and see how changes affect the "normal" behaviour
  - Order: "Perturbed" network goes to the same dynamical path as "normal" net. Changes stay in green islands, damage does not spread

Order, Chaos, and the Edge (2.5)

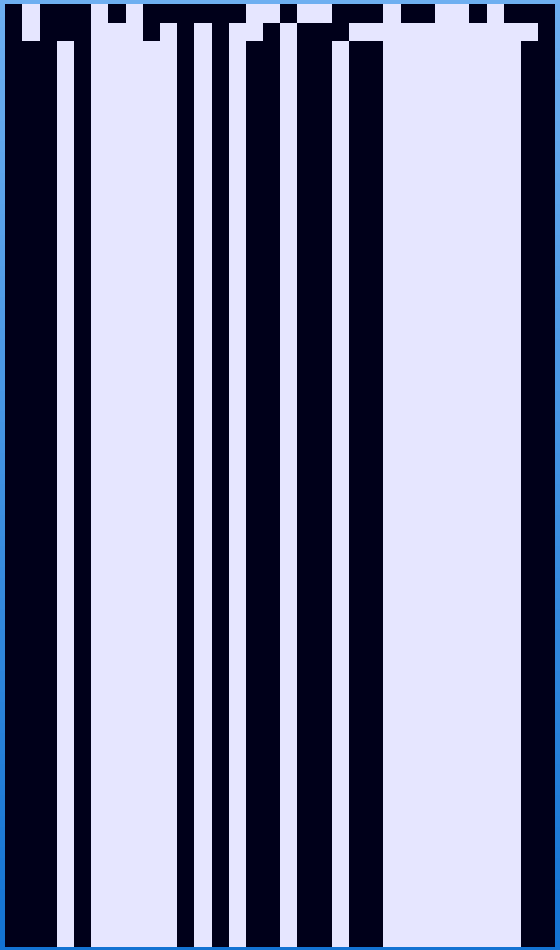
- Chaos: changes propagate, high sensitivity. Damage percolates through green sea
- Edge: changes can propagate, but not necessarily through all the network (Kauffman, 2000, pp. 168-170)

Order, Chaos, and the Edge (3)

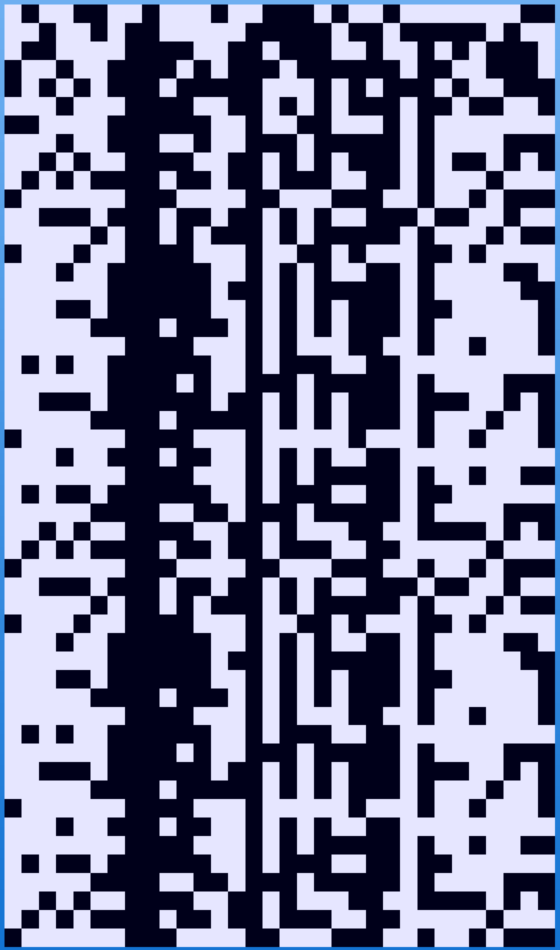
- Convergence vs. Divergence of Trajectories
  - Order: Similar similar states tend to converge to the same state
  - Chaos: similar states tend to diverge
  - Edge: nearby states tend to lie on trajectories that neither converge nor diverge in state space (Kauffman, 2000, p. 171)

Life and Computation at the Edge of Chaos (Langton, 1990; Kauffman, 1993; 2000)

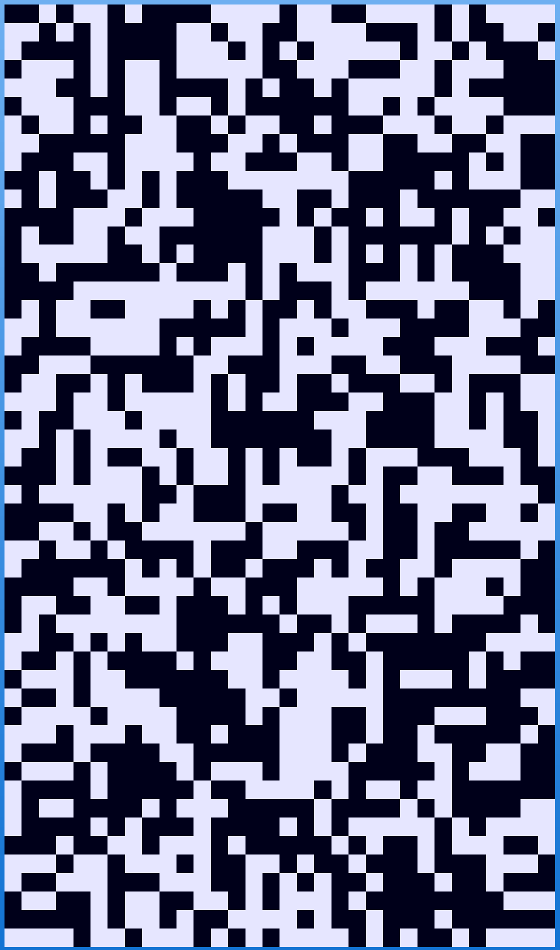
Phase Transitions in RBNs



Ordered



Edge



Chaos

Derrida's Annealed Approximation (1)

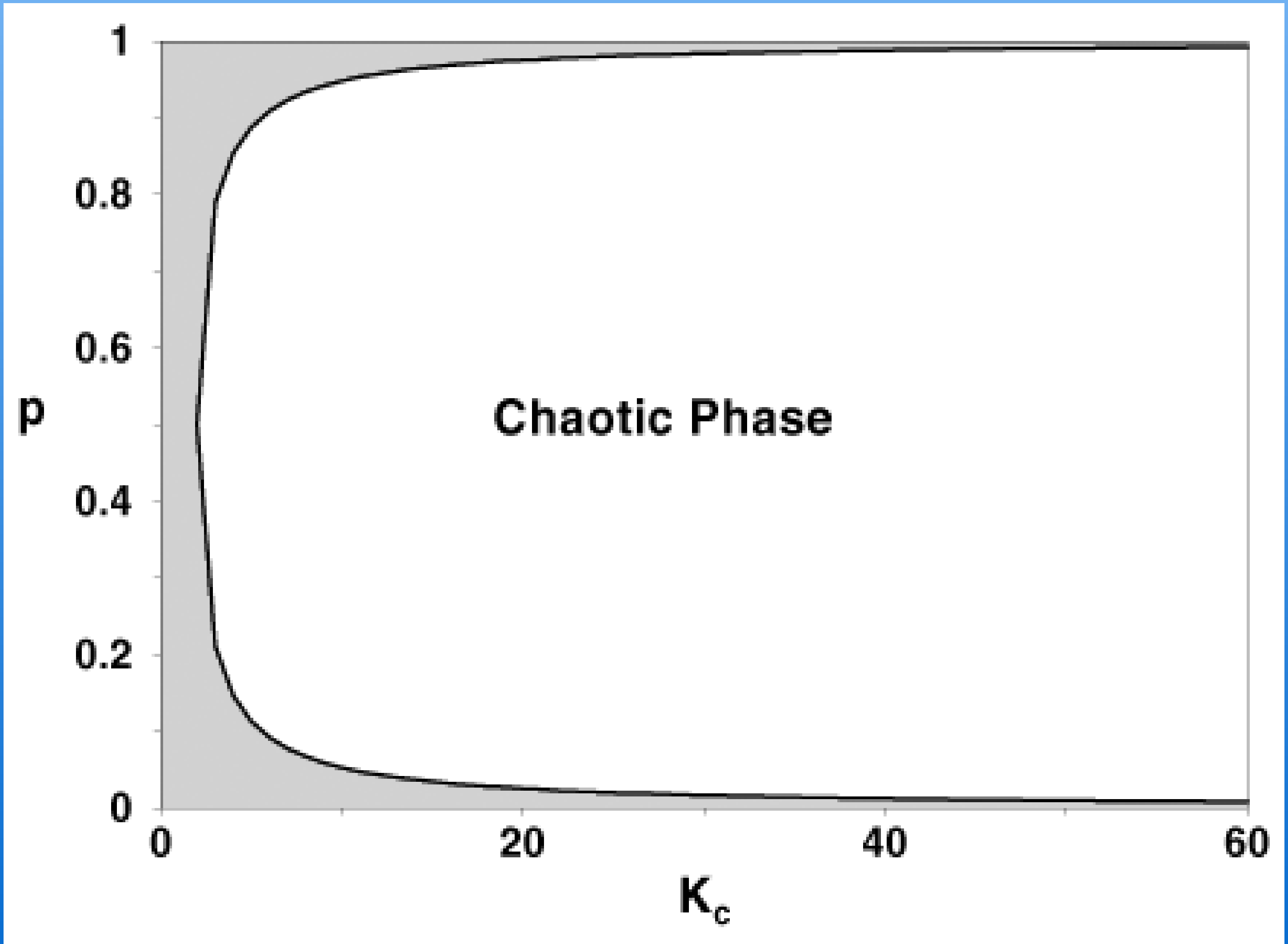
- Phase transition when *K*=2 (Derrida and Pomeau, 1986)
- Also generalized for mean *K* and probability *p*
- Measure overlap of state at t with state at t+1 using normalized Hamming distance:

$$H(A,B)=\frac{1}{n}\sum_{i}^{n}|a_i-b_i|$$

- Then choose new rules and connections
- One dimensional map
  - At *p*=0.5, converges to 0 when K<2 (ordered)
  - when *K*>2, divergence (chaos), critical phase *K*=2

Derrida's Annealed Approximation (2)

Critical *K*



$$\langle K \rangle = \frac{1}{2p(1-p)}$$

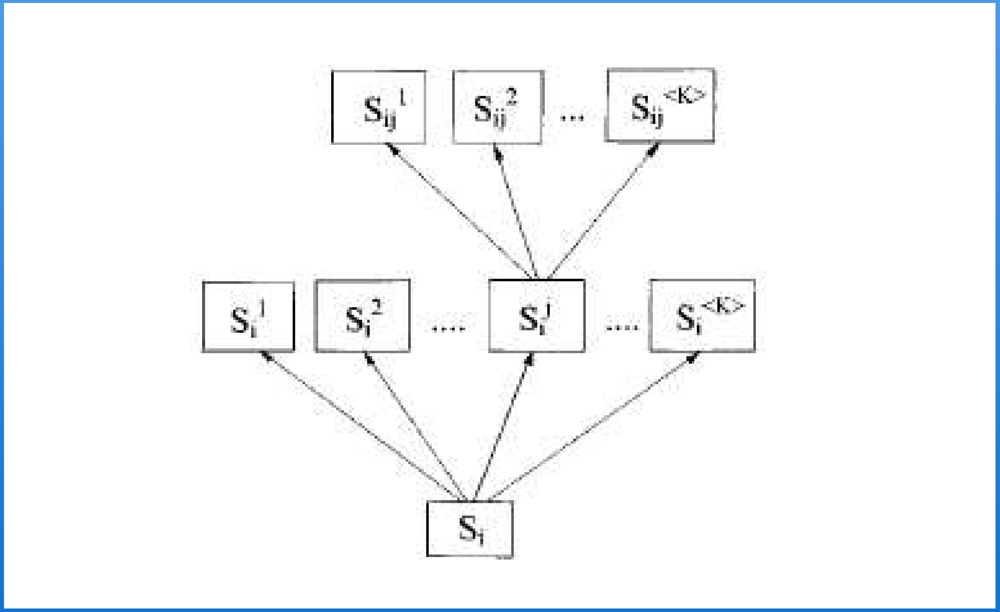
(Aldana, 2003)

A Simpler Analytical Determination (1)

- Damage spreading when single nodes are perturbed (Luque and Solé, 1997b)
- Consider trees of nodes that can affect the state of other nodes in time
- As a node has more connections, there will be an increase in the probability that a damage in a single node ($0 \rightarrow 1$ or $1 \rightarrow 0$) will percolate through the network.

A Simpler Analytical Determination (2)

- Let us focus only in one node *i* at time *t*, and a node *j* of the several *i* can affect at time $t+1$. There is a probability *p* that *j* will be one, and a damage in *i* will modify *j* towards one with probability $1-p$. The complementary case is the same. Now, for *K* nodes, we could expect that at least one change will occur if $\langle K \rangle 2p(1-p) \geq 1$, which leads to Derrida's result



- This method can be also used for other types of networks

Lyapunov exponents in RBNs

- Using the concept of boolean derivative (Luque and Solé, 2000)

$$\lambda = \log\left(2p(1-p)K\right)$$

- where $\lambda < 0$ represents the ordered phase, $\lambda > 0$ the chaotic phase, and $\lambda = 0$ the critical phase.

- **Beware**: Very high standard deviations!
- Theory can differ from practice...

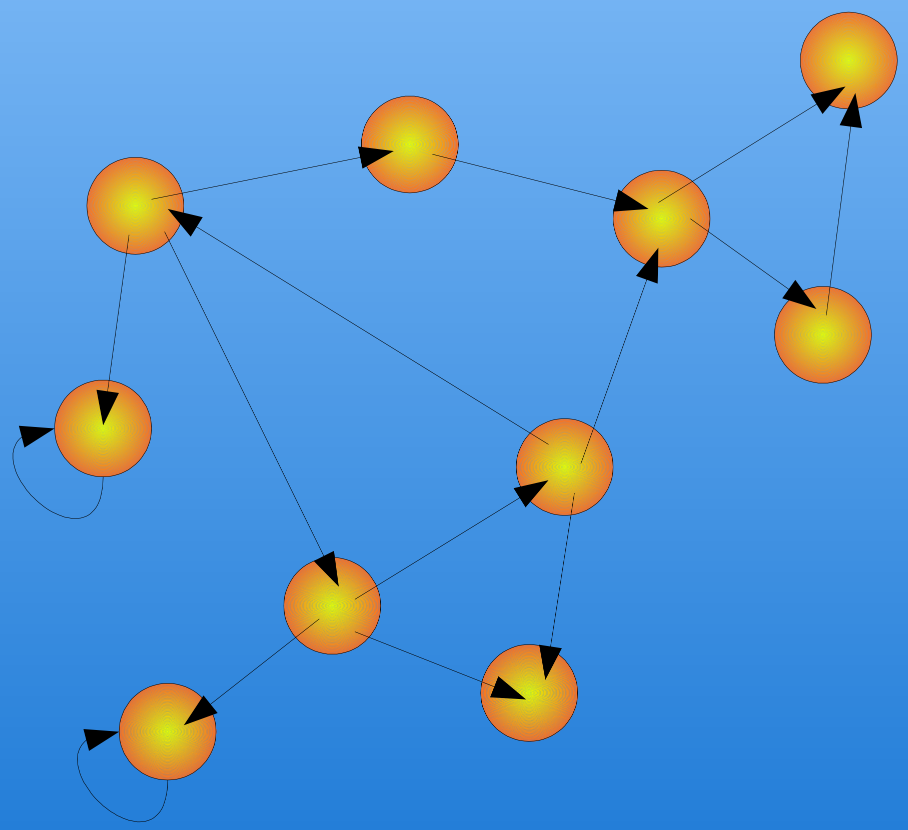
Explorations of the Classical Model

- E.g. number and length of attractors, sizes and distributions of their basins, and their dependence on different parameters (N, K, p, or topology) (Wuensche, 1997; Aldana et al., 2003)
- Analytical or statistical?

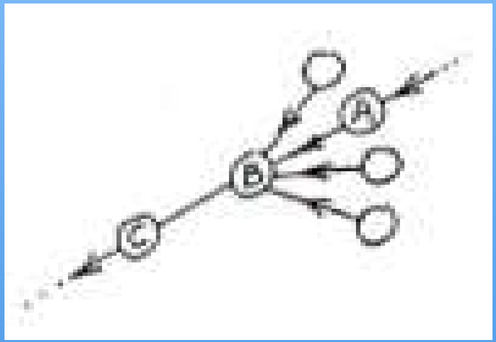
Node Structure

- Descendants
- Ancestors
- Linkage loops
- Linkage trees

- More connections, more loops, less stability



State Space Structure



(Wuensche, 1998)

- A *predecessor* of B
- C *successor* of B
  - Only one successor => CRBNs dissipative systems
- *In-degree*: number of predecessors
- *Garden-of-Eden* states: in-degree=0
- *Transient*: trajectory towards attractor

Attractor Lengths (1)

- Analytic solutions of RBNs for K = 1 (Flyvbjerg and Kjaer, 1988), and for K = N (Derrida and Flyvbjerg, 1987), but not for general case
- Statistical studies (p=0.5) (Kauffman, 1969; 1993; Bastolla and Parisi, 1998; Aldana et al., 2003; ...)
  - ***K*=1** probability of having long attractors decreases exponentially with N. Avg. number of cycles seems to be independent of N. The median lengths of state cycles are of order $\sqrt{(N/2)}$.

Attractor Lengths (2)

- ***K* ≥ *N***, average length of attractors and the transient times required to reach them grow exponentially (numerical investigations only of small networks). Typical cycle length grows proportional to $2^{N/2}$.
- ***K* = 2**, (critical phase), both typical attractor lengths and average number of attractors grow algebraically with *N*.
  - $\sqrt{N}$? - undersampling (Kauffman, 1969; Kauffman, 1993; Bastolla and Parisi, 1998)
  - *N* ? - small networks (Bilke and Sjunnesson, 2002; Gershenson, 2002)
  - Needs more research

Convergence (1)

- Measured with *G*-density, in-degree frequency distribution (histogram), etc. (Wuensche, 1998).
- **ordered** phase, very high *G*-density, high in-degree frequency => high convergence, very short transient times. The basins of attraction are very compact, with many states flowing into few states.
- **critical** phase, in-degree distribution approximates a power-law. There is medium convergence.

Convergence (2)

- **chaotic** phase, relatively lower *G*-density, and high frequency of low in-degrees. Basins of attraction are very elongated => very long average transient times. Low convergence.
- Other measures of convergence:
  - Walker's "internal homogeneity" (Walker and Ashby, 1966)
  - Langton's λ parameter (Langton, 1990)
  - Wuensche's *Z* parameter (Wuensche, 1999).
    - Automatic classification of rule-space

Multi-Valued Networks

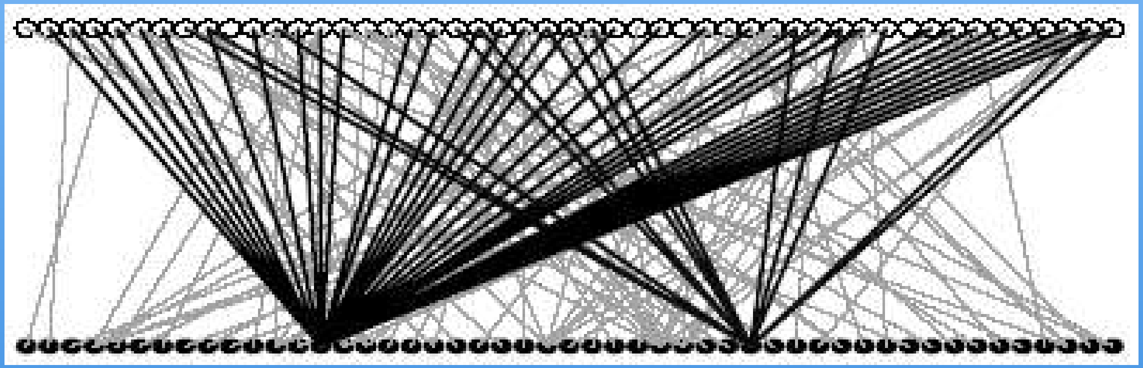
- More than 2 values per node (Solé et al., 2000; Luque and Ballesteros, 2004)
- results of Derrida are recovered for 2 values
- In nature, certain systems better described with more than two states. Particular models should go beyond the boolean idealization.
- However, for theoretical purposes, we could combine several boolean nodes to act as a multi-valued one
  - e.g. codifying in base two its state

Topologies

- Topology can change drastically properties of RBNs (Oosawa and Savageau, 2002):
- more uniform rank distributions exhibit more and longer attractors and less entropy and mutual information (less correlation in their expression patterns)
- more skewed topologies exhibit less and shorter attractors and more entropy and mutual information
- A topology based on *E. coli* (scale-free), balances the parameters to avoid the disadvantages of the extreme topologies
- Most RBN studies use uniform rank distributions

RBNs with scale-free topology (1)

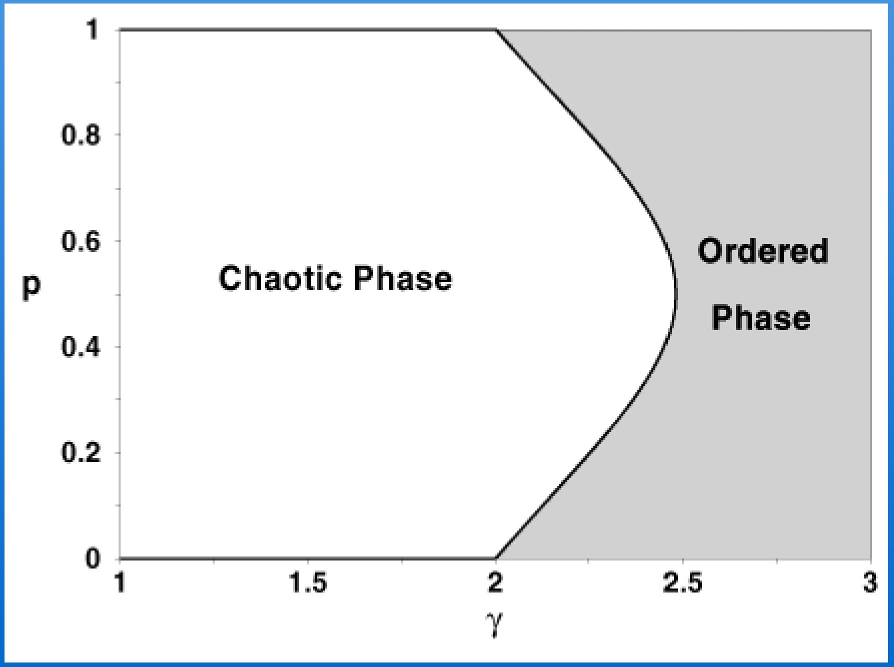
- $P(k) = [\zeta(\gamma)k^{\gamma}]^{-1}, \gamma>1$ (Aldana, 2003)



(Aldana, 2003)

- Using Derrida's method:

$$2p(1-p)\frac{\zeta(\gamma_c-1)}{\zeta(\gamma_c)}=1$$



(Aldana, 2003)

RBNs with scale-free topology (2)

- The network properties at each phase (e.g. number and length of attractors, transient times) are analogous to homogeneous RBNs.
- **Evolvability** has more space in scale-free networks, since these can adapt even in the ordered regime, where changes in well-connected elements do propagate through the network.
- However, experimental evidence shows that most biological networks are scale free with exponent $2 < \gamma < 2.5$, i.e. near edge of chaos

RBN Control (1)

- External inputs? (e.g. molecular clocks related to sunlight)
- Methods of chaos control have been successfully applied to chaotic RBNs (Luque and Solé, 1997a; 1998; Ballesteros and Luque, 2002)
- Use a periodic function to drive a very chaotic network into a stable pattern. If a periodic function determines the states of some nodes at some time, these will have a regularity that can spread through the rest of the network, developing into a global periodic pattern

RBN Control (2)

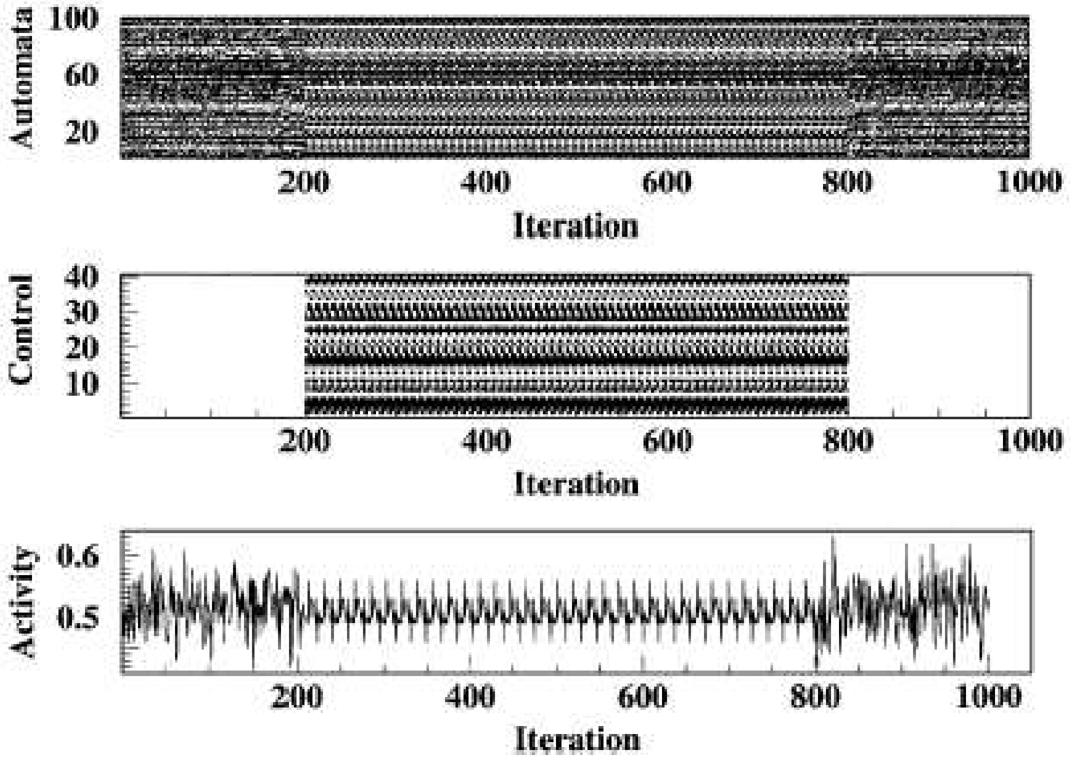
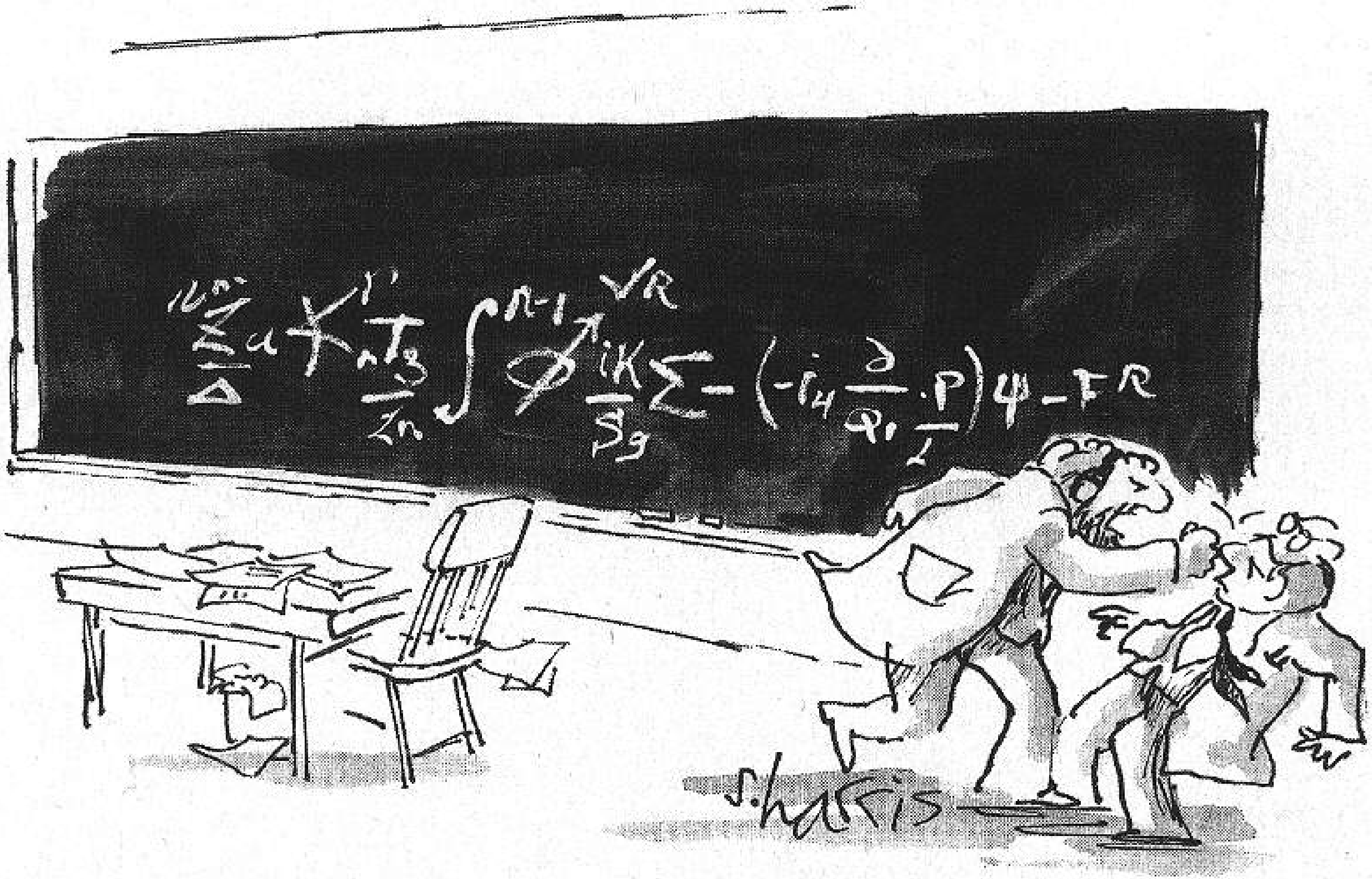


Fig. 6. RBN of size $n = 40$ in ordered state ($K = 2$, $p = 0.5$) and period $\tau = 9$ controlling the chaotic RBN described in Fig. 2. It can be seen how the first one induces an ordered behaviour of period 18 in the chaotic RBN.

RBN Control (3)

- A high percentage of nodes should be controlled to achieve periodic behaviour. However, once we control a small chaotic network, we can use this to control a larger chaotic network, and this one to control an even larger one, and so on
- This shows that it is possible to *design* chaotic networks controlled by few external signals to force them into regular behaviour
- And scale-free chaotic RBNs? Could control only high-ranking nodes?

Intermission...



"YOU WANT PROOF? I'LL GIVE YOU PROOF!"

RBN attractors as cell types, lengths as replication time?

- “order for free” (Kauffman, 1969; Kauffman, 1993)
- Drawbacks:
  - Precise number of genes, junk DNA, redundancy
  - Attractor number linear or sqrt dependence?
  - Scale-free topology
  - Biased functions
  - **Genes do not march in step!**

Updating Schemes

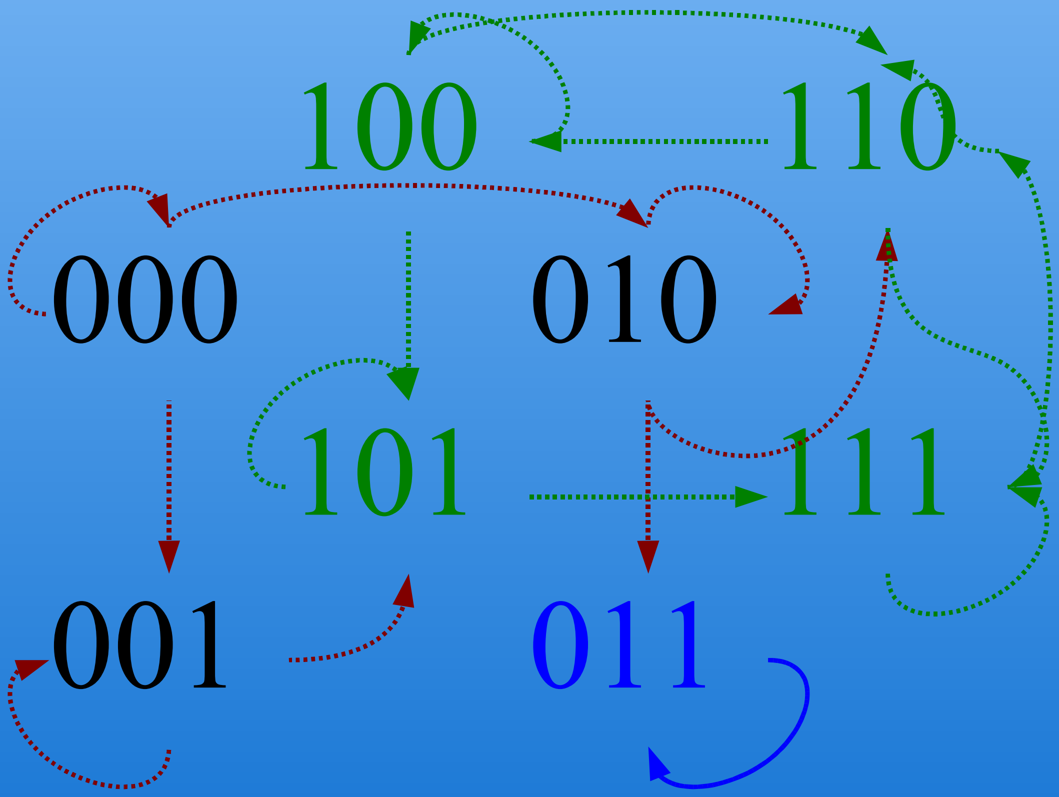
- Change of updating scheme can change drastically behaviour of a system
  - prisoner’s dilemma (Huberman and Glance, 1993)
  - Conway’s game of life (Bersini and Detours, 1994)

Asynchronous RBNs (1)

- **ARBNs**: Pick up a node randomly, update network (Harvey and Bossomaier, 1997)
  - Asynchronous AND non-deterministic
- No cycle attractors
  - Point attractors (the same than CRBNs)
    - In theory, on average 1 per net. In practice, less.
  - “loose” attractors (K>1)
- Different from **CRBN** behaviour
  - RBN useful genetic model???

Example

abc(t)	abc(t+1)
000	011
001	101
010	111
011	011
100	111
101	111
110	101
111	110



Rhythmic Asynchronous RBNs (1)

- If cells asynchronous, how could they achieve rhythm?
- Evolve RBNs and see... (Di Paolo, 2001)
- "Ring" topology (Rholfshagen and Di Paolo, 2004)
  - Only one linkage loop in pruned net
  - "Medusa" topologies found in yeast (Lee et al., 2002)
- What about more than one rhythmic attractor?

Rhythmic Asynchronous RBNs (2)

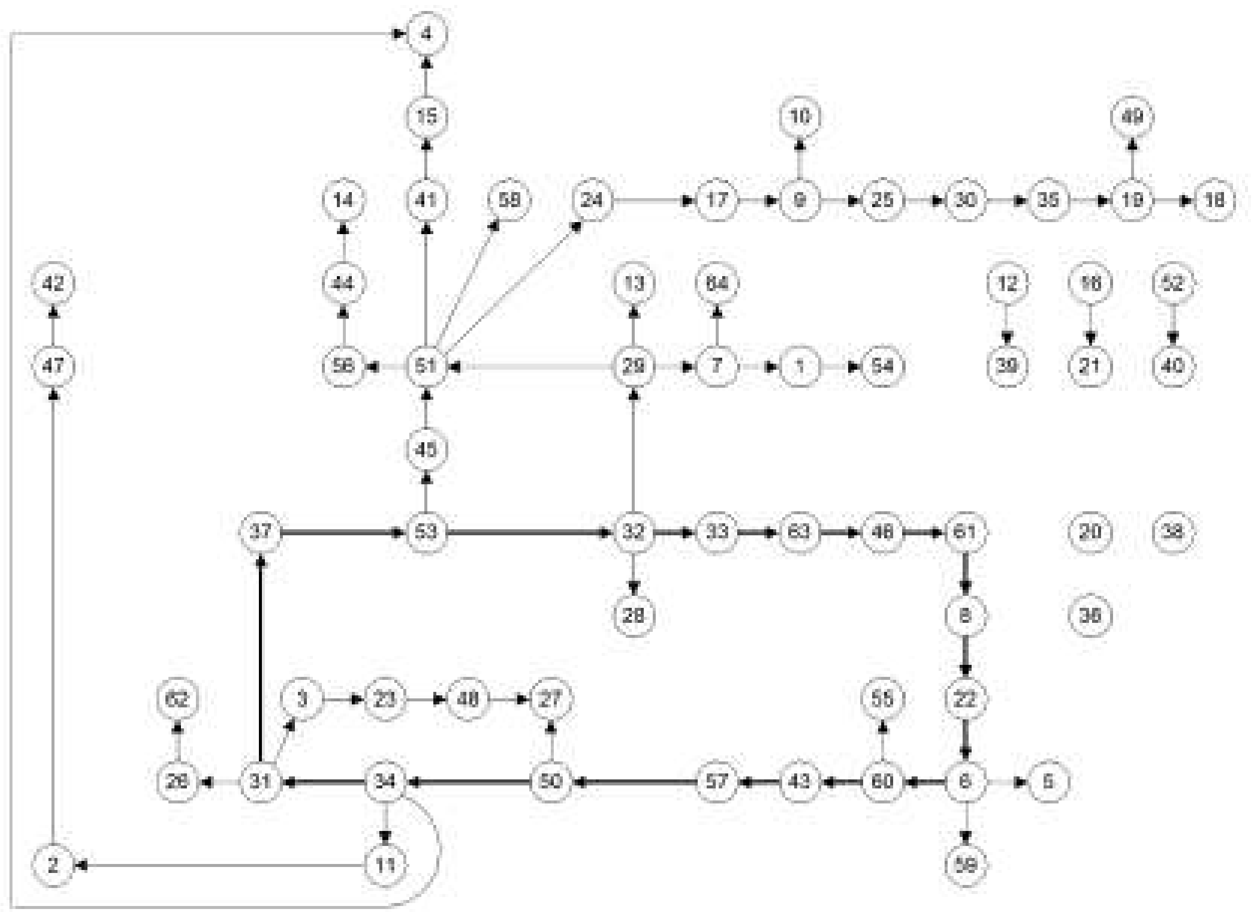


Fig. 6. A network with $N = 64$, $K = 2$ and $P = 32$. A central ring of 16 nodes underlies the rhythm produced by this network (ring: [37, 53, 32, 33, 63, 46, 61, 8, 22, 6, 60, 43, 57, 50, 34, 31]).

Deterministic Asynchronous RBNs

- Cells not synchronous, but not purely stochastic
- **DARBNs**: Introduce parameters P_i and Q_i per node

$$P_i, Q_i \in \mathbb{N}, P_i > Q_i, P_{max} \geq P_i, Q_{max} \geq Q_i$$

- Update a node when mod of time over $P_i == Q_i$
  - P_i - period
  - Q_i – translation
- If more than one node should be updated at a time step, do this in a sequential order
- Asynchronous, deterministic (Gershenson, 2002)

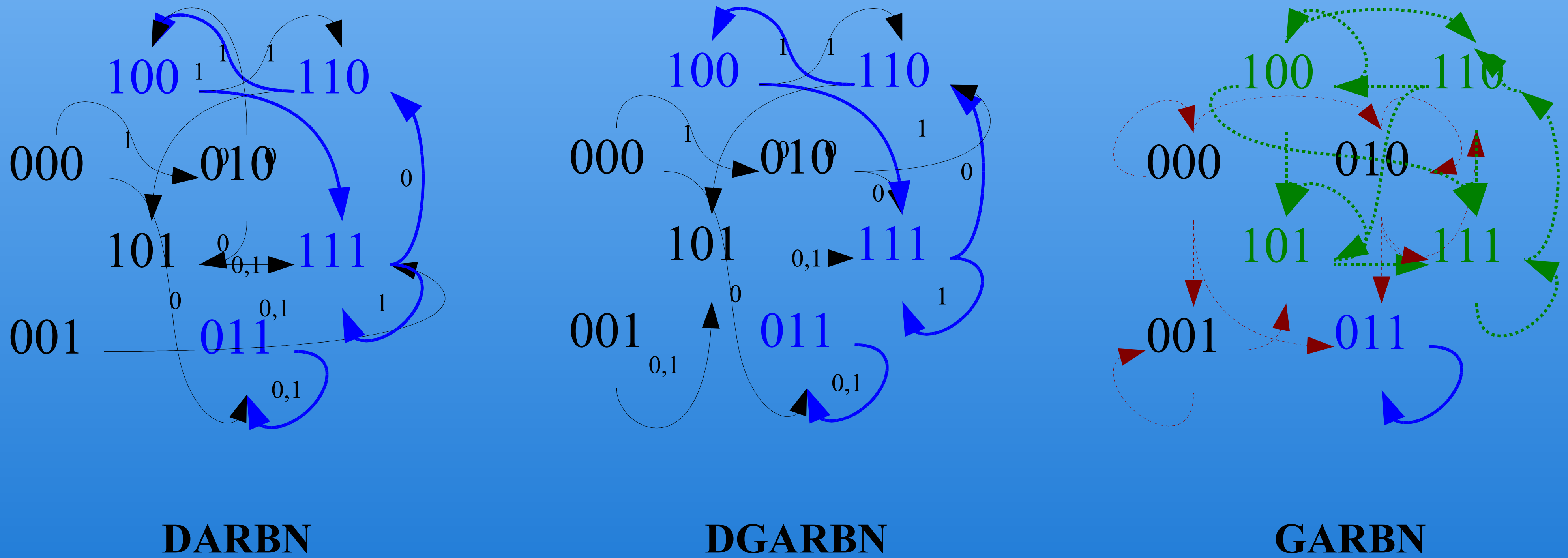
Deterministic Generalized Asynchronous RBNs

- **DGARBNs**: Like DARBNs, but if more than one node should be updated, do this synchronously
- Semi-synchronous, deterministic (Gershenson, 2002)

Generalized Asynchronous RBNs

- **GARBNs**: Like ARBNs, but select randomly nodes to update synchronously
- Semi-synchronous, non-deterministic (Gershenson, 2002)

Examples



P's={1,1,2}, Q's={0,0,0}

RBNs and Updating Schemes

- Many properties change drastically (Gershenson, 2002)
- All RBNs share point attractors, but basins can change
- More difference in attractor length due to determinism than synchronicity
- All have similar “edge of chaos” (Gershenson, 2004a,b)
- All perform *complexity reduction* (Gershenson, 2004b)
  - Including loose attractors
- Can map any deterministic RBN into a CRBN (Gershenson, 2002)
  - Encode in base two the periods, add new nodes

Thomas' ARBNs

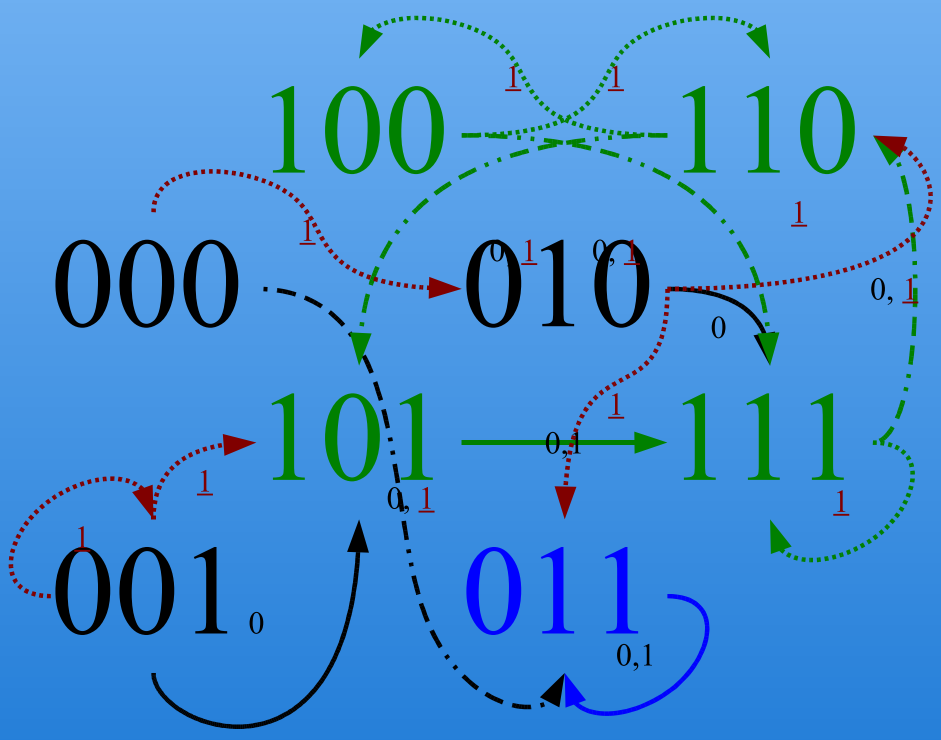
- ARBNs using delays (deterministic or stochastic) (Thomas, 1973; Thomas, 1978; Thomas, 1991)
- Used for analysis of precise networks, their circuits, and feedback loops. For ensembles???
- A positive feedback loop (direct or indirect autocatalysis) implies two point attractors
  - *Multistationarity*
- A negative feedback loop implies periodic behaviour, i.e. point or cycle attractors
  - *Homeostasis*

Mixed-context RBNs (1)

- Sets *P* and *Q* (P*i*'s and Q*i*'s) as *context* of a network (Gershenson, Broekaert, and Aerts, 2003)
  - External factors can change precise updating periods
- Same DGARBN can have different behaviours with different contexts
- **MxRBNs**: *M* "pure" contexts, one chosen randomly at each *R* time steps
- Semi-synchronous, "quantum-like"
  - Non-determinism introduced in a very particular and controlled fashion

Example

abc(t)	abc(t+1)
000	011
001	101
010	111
011	011
100	111
101	111
110	101
111	110

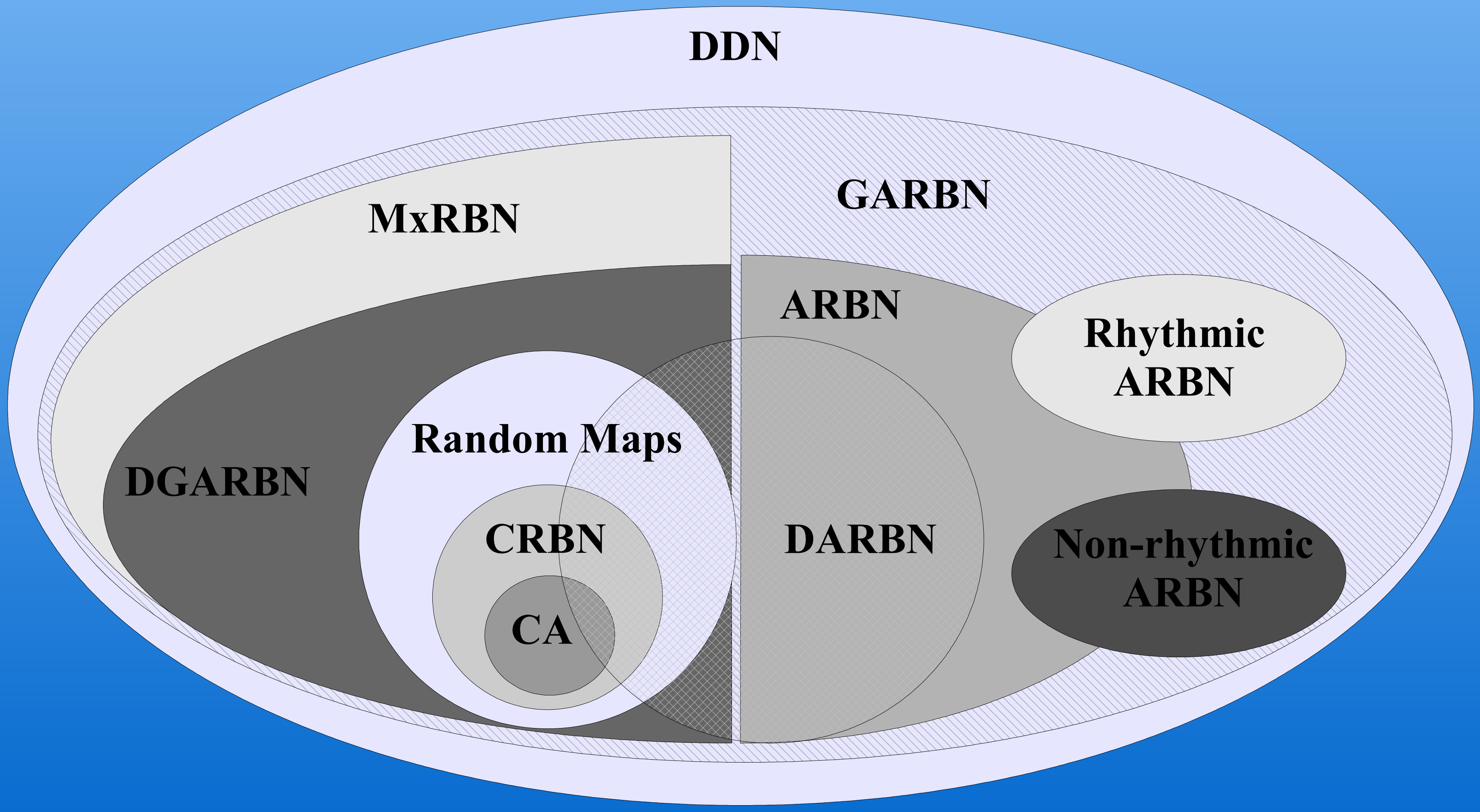


P1's={1,1,2}, Q1's={0,0,0}
P2's={2,1,1}, Q2's={0,0,0}

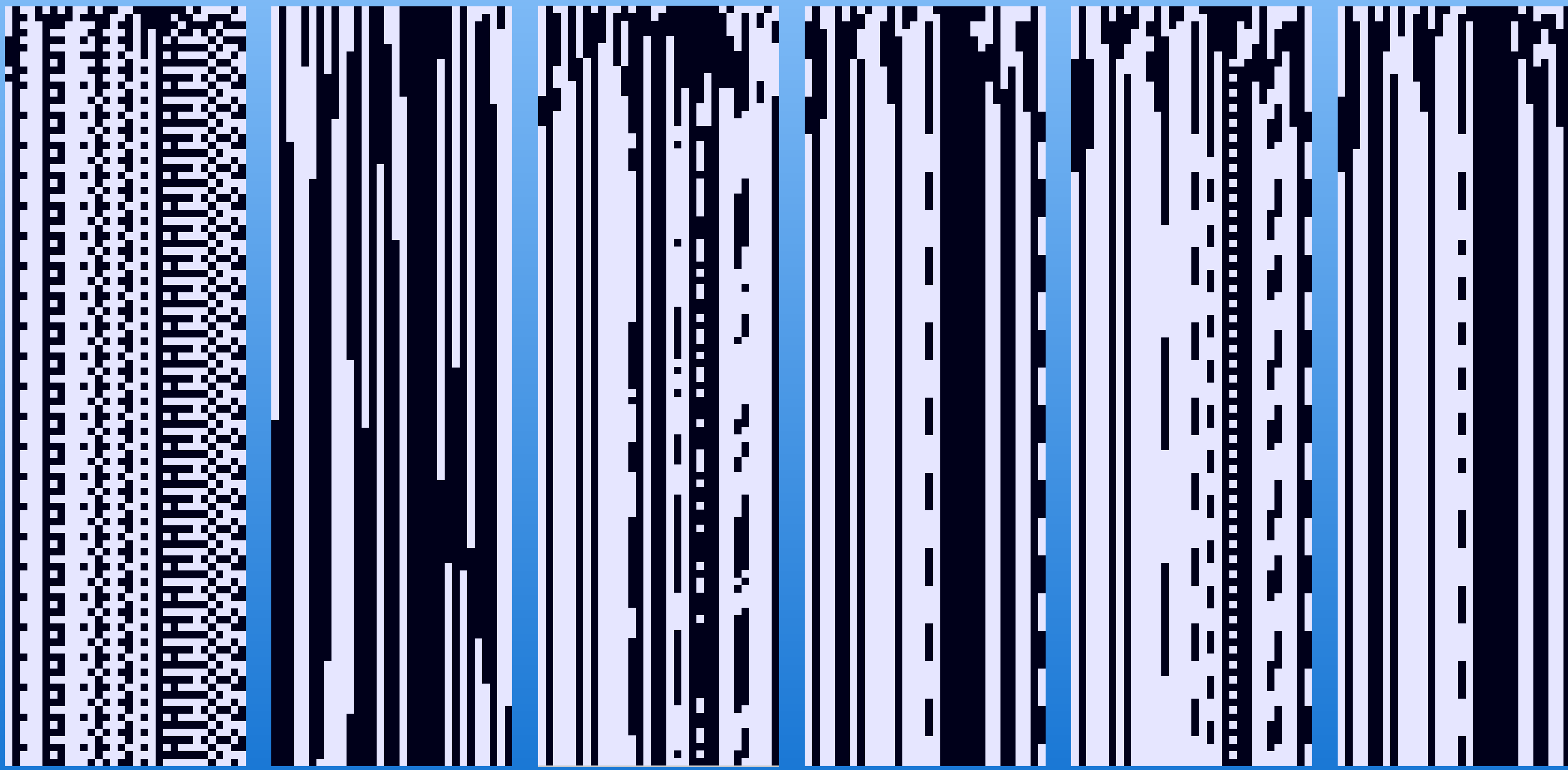
How much non-determinism?

- GARBNs: *N* “coin flips” per time step
- ARBNs: one coin flip per time
- MxRBNs: one coin flip per *R* time steps
  - The higher the value of *R* and the lower number of *M* contexts, the less stochasticity there will be
- MxRBNs similar number of attractors than ARBNs and GARBNs, but much more complexity reduction (even more than CRBNs)
  - Information can be “thrown” into the context

Classification of RBNs



Dynamics Example



CRBN **ARBN** **GARBN** **DARBN** **DGARBN** **MxRBN**

Same net & initial state. $N = 32$, $K = 2$, $p = 0.5$, $P_{max} = 5$, $Q_{max} = 4$, $M = 2$, $R = 10$.

Applications

- Genetic regulatory networks
- Evolution and computation
- Neural networks (Huepe and Aldana, 2002)
- Social modelling (Shelling, 1971)
- Robotics (Quick et al., 2003)
- Music generation (Dorin, 2000).
- Mathematics
  - Cellular automata (von Neumann, 1966; Wolfram, 1986; Wuensche and Lesser, 1992)
  - Percolation theory (Stauffer, 1985)
- ...

Genetic Regulatory Networks (1)

- Nodes as genes: “on-off” (activation), interaction via proteins (Kauffman, 1969)
- Generic properties in ensemble studies (Kauffman, 2004)
- Analysis and prediction of genomic interaction, data mining (Somogyi and Sniegoski, 1996; Somogyi et al., 1997; D’haeseleer et al.,1998)
- probabilistic boolean networks (PBNs): infer possible gene functionality from incomplete data (Shmulevich et al., 2002)

Genetic Regulatory Networks (2)

- Experimental evidence of cell types as attractors of RBNs (Huang and Ingber, 2000)
  - Very strong correlation for some genes as a cell type is mechanically forced
  - Not all genes determine cell type (but e.g. metabolism)
- Continuos states GRN models (Glass and Kauffman, 1973; Kappler et al., 2002).
  - Use of differential equations in which gene interactions are incorporated as logical functions
  - no need for a clock to calculate the dynamics
  - Ensemble studies???

Evolution and Computation (1)

- Evolvability is expected at the edge of chaos
- Network evolvability properties:
  - robustness, redundancy, degeneracy, modularity (Fernández and Solé, 2004)
- Life performs computations (Hopfield, 1994)
  - Understanding computation networks helps us to understand life and its possibilities
  - "How can computational networks be evolved?" close to "How could life evolve?"

Evolution and Computation (2)

- Evolution of RBNs using genetic algorithms (Stern, 1999; Lemke et al., 2001)
- Evolvable hardware (Thompson, 1998)
  - Evolution of logical circuits in reconfigurable hardware
- Issues of evolvability also interesting for genetic algorithms, genetic programming, etc.
- ...

Tools (1)

- **DDLab** (Andy Wuensche)
  - synchronous RBNs and CA, multi-valued networks
  - Dynamics and basins of attraction visualization
  - It includes a wide variety of measures, data, analysis and statistics
  - Very well documented, runs on most platforms.
  - http://www.ddlab.com
- **RBN Toolbox for Matlab** (Christian Schwarzer and Christof Teuscher)
  - Simulation and visualization of RBNs.
  - Different updating schemes, statistical functions, etc.
  - http://www.teuscher.ch/rbntoolbox

Tools (2)

- **RBNLab** (Carlos Gershenson)
  - Simulation and visualization of RBNs with different updating schemes
  - Point, cycle, and loose attractors, other statistics...
  - Java, code and program at http://rbn.sourceforge.net
- BN/PBN Toolbox for Matlab (Harri Lähdesmäki and Ilya Shmulevich)
  - CRBNs and PBNs.
  - functions for simulating network dynamics, computing statistics (a lot), inferring networks from data, visualization and printing, intervention, membership testing of Boolean functions, etc.
  - http://www2.mdanderson.org/app/ilya/PBN/PBN.htm

Future Lines of Research

- Ensemble approach (Kauffman, 2004)
- RBNs for data mining and GRN analysis
- Evolvability and adaptability at an abstract level
- Generalizations, combinations, and refinements of the different types of RBNs
  - e.g. scale-free multi-valued DGARBNs, etc
- General analytical solution for CRBNs
- ...

Conclusions

- Tutorial only overview, but main topics covered
- RBNs interesting due to generality
  - Many conclusions with few assumptions
  - Illustrate order-complexity-chaos for non-physicists
- Which model is best? It depends...
- Theory vs. practice? Balance!
- An inviting research topic