Discrete and Indiscrete Models of Biological Networks

Winfried Just
Ohio University

November 17, 2010

Who are we?
What are we doing here?

Who are we?
What are we doing here?

- A population of interacting organisms.

Who are we?
What are we doing here?

- A population of interacting organisms.
- Behavior is controlled by an organ, the brain.

Who are we?
What are we doing here?

- A population of interacting organisms.
- Behavior is controlled by an organ, the brain.
- Brain output is based on the firing patterns of interconnected and interacting cells, the neurons.

Who are we?
What are we doing here?

- A population of interacting organisms.
- Behavior is controlled by an organ, the brain.
- Brain output is based on the firing patterns of interconnected and interacting cells, the neurons.
    - We would like to decipher the "neural code."

Who are we?
What are we doing here?

- A population of interacting organisms.
- Behavior is controlled by an organ, the brain.
- Brain output is based on the firing patterns of interconnected and interacting cells, the neurons.
    - We would like to decipher the "neural code."
      This is one of the big open problems in biology.

Who are we?
What are we doing here?

- A population of interacting organisms.
- Behavior is controlled by an organ, the brain.
- Brain output is based on the firing patterns of interconnected and interacting cells, the neurons.
  - We would like to decipher the "neural code."
    This is one of the big open problems in biology.
- The firing of an individual neuron is generated by changes in concentration of certain molecules and ions. Such concentrations change as the result of chemical reactions.

Who are we?
What are we doing here?

- A population of interacting organisms.
- Behavior is controlled by an organ, the brain.
- Brain output is based on the firing patterns of interconnected and interacting cells, the neurons.
    - We would like to decipher the "neural code."
      This is one of the big open problems in biology.
- The firing of an individual neuron is generated by changes in concentration of certain molecules and ions. Such concentrations change as the result of chemical reactions.
    - I am ignoring ion transport here.

Mathematical modeling is a process of selective ignorance. (Louis Gross)

Mathematical modeling is a process of selective ignorance. (Louis Gross)

Judiciously choosing what to stay ignorant about can help us in seeing the forest behind the trees.

Mathematical modeling is a process of selective ignorance. (Louis Gross)

Judiciously choosing what to stay ignorant about can help us in seeing the forest behind the trees.

With how much ignorance can we get away with and still discover something true and biologically relevant?

Who are we?
What are we doing here?

- A population of interacting organisms.
- Behavior is controlled by an organ, the brain.
- Brain output is based on the firing patterns of interconnected and interacting cells, the neurons.
- The firing of an individual neuron is generated by changes in concentration of certain molecules and ions. Such concentrations change as the result of chemical reactions.
- What is going on in this room is simply the dynamics of gigantic networks of chemical reactions. The rest is what biologists call emergent properties.

Networks in biology

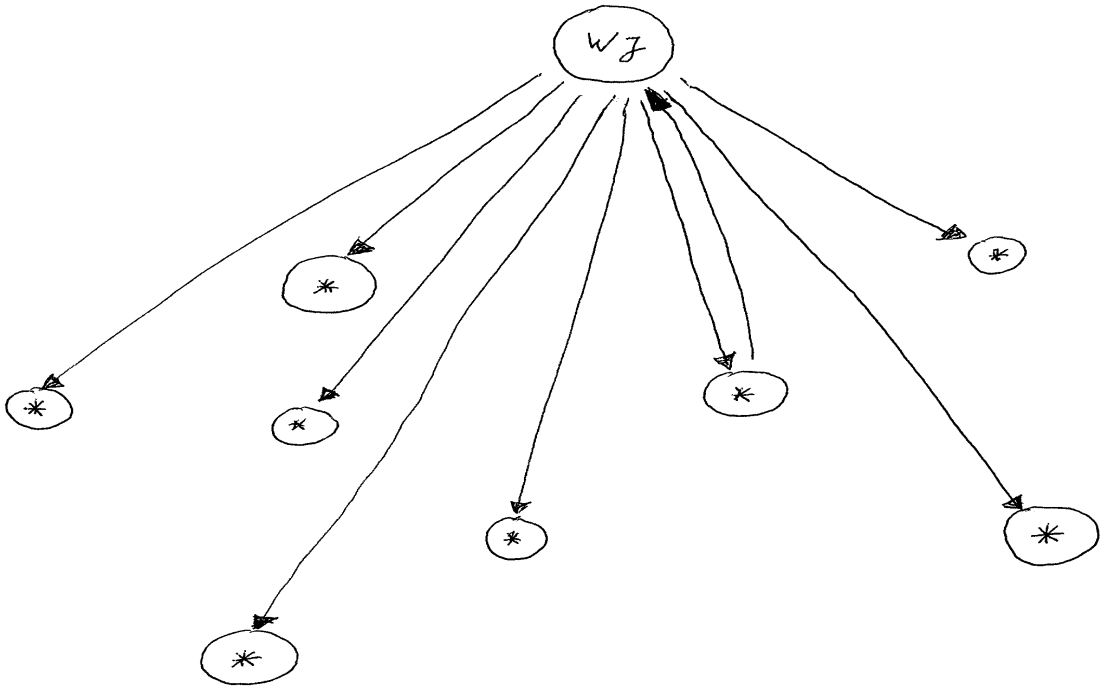
Networks occur at all levels of biological organization.

Many biological questions can be framed in terms of network dynamics.

What is a network anyway?

- The *connectivity* of a network is given by a digraph D.
- The nodes of D may represent organisms, neurons, chemical species or other types of *agents.*
- The arcs of D represent interactions.

The seminar network



What is a network anyway?

- The *connectivity* of a network is given by a digraph D.
- The nodes of D may represent organisms, neurons, chemical species or other types of *agents.*
- The arcs of D represent interactions.
- Each node has a *state* at any given time, the state of the whole network is the vector of all these individual states.
- The network dynamics is the change of the state vector over time.

Discrete and continuous networks

We will consider here only two types of networks:

- *Discrete networks* for which time takes on integer values and the state space is finite.
- *Continuous flows* for which time is modeled by reals and the state space is some n-dimensional manifold.

What is a network anyway?

- The *connectivity* of a network is given by a digraph D.
- The nodes of D may represent organisms, neurons, chemical species or other types of *agents.*
- The arcs of D represent interactions.
- Each node has a *state* at any given time, the state of the whole network is the vector of all these individual states.
- The network dynamics is the change of the state vector over time.
- The dynamics determined by any *initial state* is called the trajectory of this state.
- A steady state is a state whose trajectory remains constant.
- We will only consider networks here for which each trajectory approaches an attractor.

Moving down to base level

- ecosystems
- populations
- organisms
- organs
- tissues
- cells
- organelles
- molecules

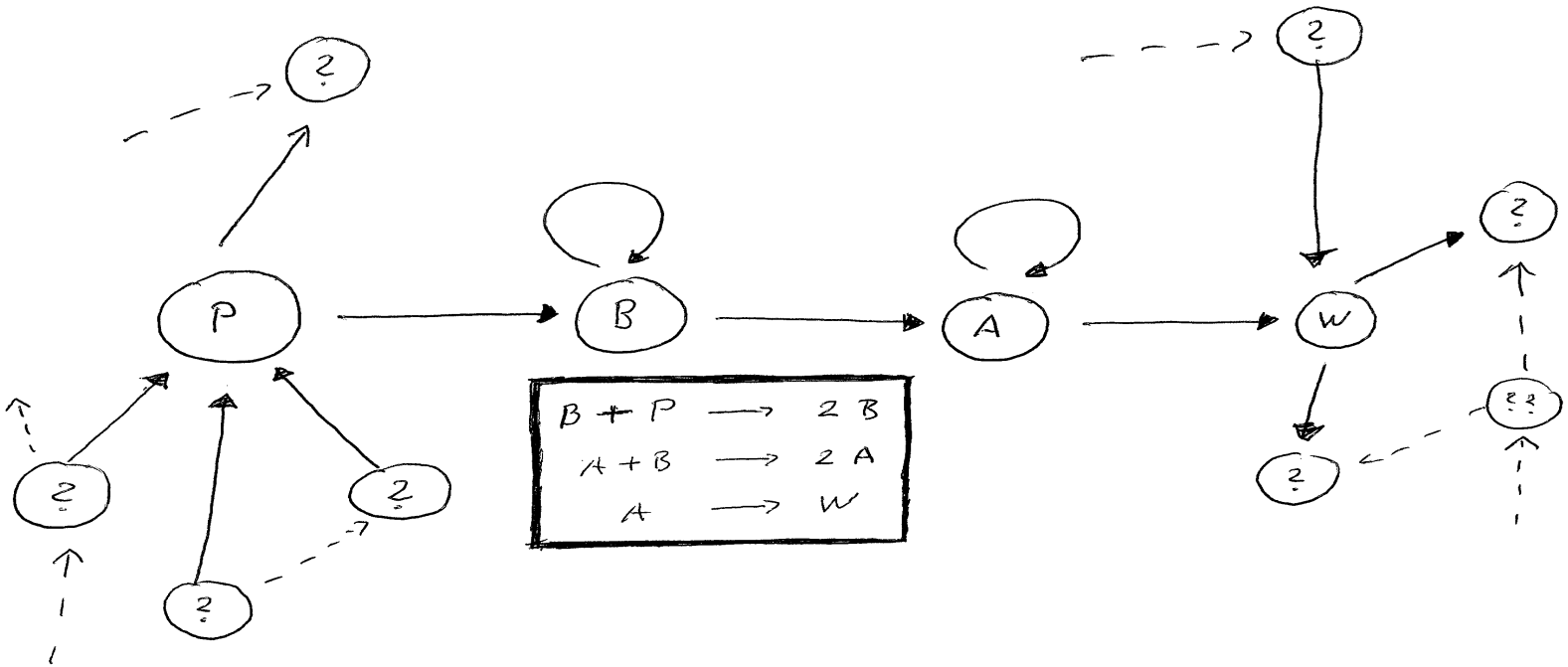
A chemical reaction network

Suppose A, B represent chemical species of interest who participate only in the following reactions:

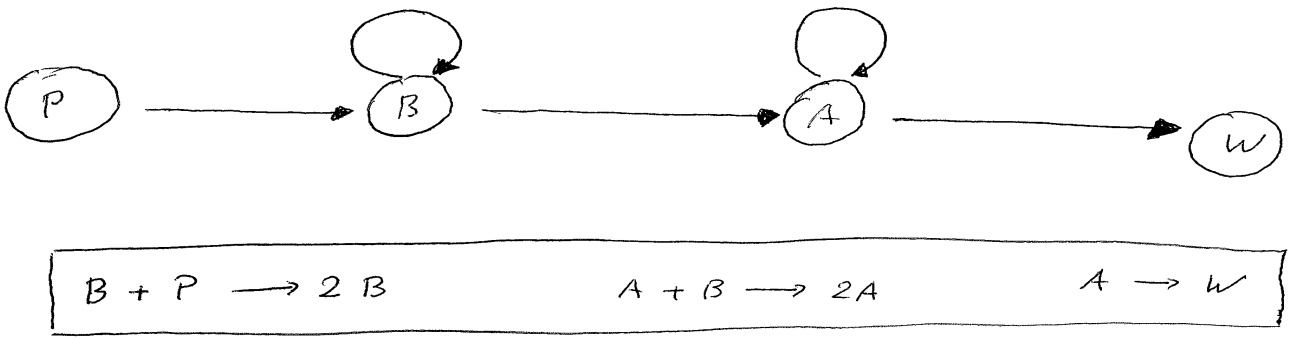
$$A + B \to 2A, \quad B + P \to 2B, \quad A \to W.$$

Can we isolate these reactions from the larger network? How?

The larger network



First try: Ignore everything else



First try: Ignore everything else

Let the state of a chemical species X be represented by its concentration $[X]$; assume *mass action kinetics.* Then

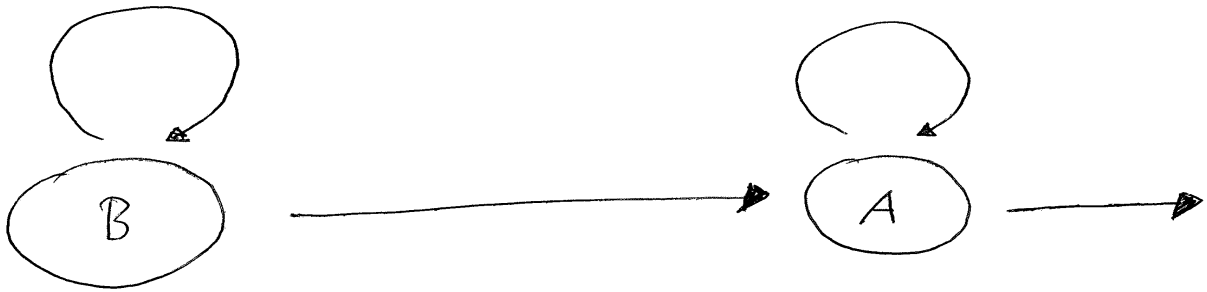
$$A + B \to 2A, \qquad B + P \to 2B, \qquad A \to W$$

define a continuous flow

$$\begin{aligned}
\frac{d[A]}{dt} &= k_1[A][B] - k_2[A] \\
\frac{d[B]}{dt} &= k_3[B][P] - k_1[A][B] \\
\frac{d[P]}{dt} &= -k_3[B][P] \\
\frac{d[W]}{dt} &= k_2[A].
\end{aligned}$$

The only attractor is the steady state with $[A] = [B] = [P] = 0$.

Second try: Ignore P and W as well



Second try: Ignore P and W as well

Suppose P is always plentiful and present in practically the same concentration. Then our system becomes

$$A + B \to 2A, \qquad B \to 2B, \qquad A \to \emptyset,$$

which defines a continuous flow

$$\frac{d[A]}{dt} = k_1[A][B] - k_2[A]$$
$$\frac{d[B]}{dt} = k_3[B] - k_1[A][B].$$

For suitable choices of k_1, k_2, k_3 the system has a cyclic attractor.

A general question

Network modeling in biology usually focuses on a small subnetwork of interest and ignores the embedding of this subnetwork in much larger ones. This approach seems to work quite well much of the time.

A general question

Network modeling in biology usually focuses on a small subnetwork of interest and ignores the embedding of this subnetwork in much larger ones. This approach seems to work quite well much of the time.

Why?

A general question

Network modeling in biology usually focuses on a small subnetwork of interest and ignores the embedding of this subnetwork in much larger ones. This approach seems to work quite well much of the time.

Why?

Under what conditions should we be able to isolate the small subnetwork of interest?

A general question

Network modeling in biology usually focuses on a small subnetwork of interest and ignores the embedding of this subnetwork in much larger ones. This approach seems to work quite well much of the time.

Why?

Under what conditions should we be able to isolate the small subnetwork of interest?
Why should there even exist a small subnetwork of interest?

Moving to the top level

- ecosystems
- populations
- organisms
- organs
- tissues
- cells
- organelles
- molecules

Food webs

- The nodes represent biological species.
- The arcs represent predation (who eats whom).
- The state of a node usually represents the population size.

Food webs

- The nodes represent biological species.
- The arcs represent predation (who eats whom).
- The state of a node usually represents the population size.

For example, species A might be amoeba who feed on bacteria B. The state variable a represents the number of amoeba, the state variable b the number of bacteria.

Food webs

- The nodes represent biological species.
- The arcs represent predation (who eats whom).
- The state of a node usually represents the population size.

For example, species A might be amoeba who feed on bacteria B. The state variable a represents the number of amoeba, the state variable b the number of bacteria.
Let us assume bacteria have plenty of food and never die except when eaten by amoeba. They multiply by cell division. Amoeba may die at random times and need to eat a certain number of bacteria to accumulate the energy for cell division.

First try: a discrete model

Clearly, a and b are integers. Cell division and predation are discrete events. So let us try to build a discrete individual-based model where time moves from one cell division or predation event to the next.

First try: a discrete model

Clearly, a and b are integers. Cell division and predation are discrete events. So let us try to build a discrete individual-based model where time moves from one cell division or predation event to the next. There are two problems:

- We need to somehow record the fraction of energy that an individual amoeba has already accumulated towards cell division.

First try: a discrete model

Clearly, a and b are integers. Cell division and predation are discrete events. So let us try to build a discrete individual-based model where time moves from one cell division or predation event to the next. There are two problems:

- We need to somehow record the fraction of energy that an individual amoeba has already accumulated towards cell division.
- The type of the *next* event that will happen (a cell division, death of an amoeba, or predation event) is not completely determined by the current state.

First try: a discrete model

Clearly, a and b are integers. Cell division and predation are discrete events. So let us try to build a discrete individual-based model where time moves from one cell division or predation event to the next. There are two problems:

- We need to somehow record the fraction of energy that an individual amoeba has already accumulated towards cell division.
- The type of the *next* event that will happen (a cell division, death of an amoeba, or predation event) is not completely determined by the current state.

This kind of modeling gives us a Markov Chain with a very large state space and the model may be difficult to simulate and analyze.

Second try: a continuous model

Clearly, a and b are large integers. Let us treat them as fractions of, say, one million and pretend the state space is continuous.

Second try: a continuous model

Clearly, a and b are large integers. Let us treat them as fractions of, say, one million and pretend the state space is continuous. The increase in a due to cell division as a result of successful predation will be proportional to ab, the decrease in a due to random death will be proportional to a.

Second try: a continuous model

Clearly, a and b are large integers. Let us treat them as fractions of, say, one million and pretend the state space is continuous. The increase in a due to cell division as a result of successful predation will be proportional to ab, the decrease in a due to random death will be proportional to a.

$$\frac{da}{dt} = k_1 ab - k_2 a.$$

The increase in b due to cell division will be proportional to b, the decrease in b due to predation will be proportional to ab.

Second try: a continuous model

Clearly, a and b are large integers. Let us treat them as fractions of, say, one million and pretend the state space is continuous. The increase in a due to cell division as a result of successful predation will be proportional to ab, the decrease in a due to random death will be proportional to a.

$$\frac{da}{dt} = k_1 ab - k_2 a.$$

The increase in b due to cell division will be proportional to b, the decrease in b due to predation will be proportional to ab.

$$\frac{db}{dt} = k_3 b - k_1 ab.$$

Second try: a continuous model

Clearly, a and b are large integers. Let us treat them as fractions of, say, one million and pretend the state space is continuous. The increase in a due to cell division as a result of successful predation will be proportional to ab, the decrease in a due to random death will be proportional to a.

$$\frac{da}{dt} = k_1 ab - k_2 a.$$

The increase in b due to cell division will be proportional to b, the decrease in b due to predation will be proportional to ab.

$$\frac{db}{dt} = k_3 b - k_1 ab.$$

Looks familiar?

Moving down a few levels

- ecosystems
- populations
- organisms
- organs
- brain tissues
- cells
- organelles
- molecules

An ODE Model of Neuronal Networks

by Terman D, Ahn S, Wang X, Just W, Physica D. 2008

Each excitatory (E-) cell satisfies

$$\begin{aligned}
\frac{dv_i}{dt} &= f(v_i, w_i) - g_{EI} \sum s_j^I (v_i - v_{syn}^I) \\
\frac{dw_i}{dt} &= \epsilon g(v_i, w_i) \\
\frac{ds_i}{dt} &= \alpha(1 - s_i) H(v_i - \theta_E) - \beta s_i.
\end{aligned}$$

Each inhibitory (I-) cell satisfies

$$\begin{aligned}
\frac{dv_i^I}{dt} &= f(v_i^I, w_i^I) - g_{IE} \sum s_j (v_i^I - v_{syn}^E) - g_{II} \sum s_j^I (v_i^I - v_{syn}^I) \\
\frac{dw_i^I}{dt} &= \epsilon g(v_i^I, w_i^I) \\
\frac{dx_i^I}{dt} &= \epsilon \alpha_x (1 - x_i^I) H(v_i^I - \theta_I) - \epsilon \beta_x x_i^I \\
\frac{ds_i^I}{dt} &= \alpha_I (1 - s_i^I) H(x_i^I - \theta_x) - \beta_I s_i^I.
\end{aligned}$$

Mathematical Neuroscience is Difficult!

- Individual neurons are usually modeled by the the Hodgkin-Huxley Equations.
  - Nonlinear ODEs involving multiple time scales.
  - Hard to analyze both mathematically and computationally.
- Neuronal networks involve a large number of individual neurons.
  - Details of the connectivity not usually known.
  - Hard to analyze how connectivity influences ODE dynamics.

Mathematical Neuroscience is Difficult!

- Individual neurons are usually modeled by the the Hodgkin-Huxley Equations.
  - Nonlinear ODEs involving multiple time scales.
  - Hard to analyze both mathematically and computationally.
- Neuronal networks involve a large number of individual neurons.
  - Details of the connectivity not usually known.
  - Hard to analyze how connectivity influences ODE dynamics.

Fortune cookie: Doing the impossible is kind of fun.

A manageable problem?

Recordings from certain neuronal tissues reveal the following pattern: Time seems to be partitioned into episodes with surprisingly sharp boundaries. During one episode, a group of neurons fires, while other neurons are at rest. In the next episode, a different group of neurons fires. Group membership may vary from episode to episode, a phenomenon called "dynamic clustering."

Why?

A manageable problem?

Recordings from certain neuronal tissues reveal the following pattern: Time seems to be partitioned into episodes with surprisingly sharp boundaries. During one episode, a group of neurons fires, while other neurons are at rest. In the next episode, a different group of neurons fires. Group membership may vary from episode to episode, a phenomenon called "dynamic clustering."

Why?

Can we mathematically explain this phenomenon?

Some Simple Facts

The following is true in at least some neuronal networks.

- Neurons fire or are at rest.
- After a neuron has fired, it has to go through a certain *refractory period* when it cannot fire.
- A neuron will fire when it has reached the end of its refractory period and when it receives firing input from a specified minimal number of other neurons.

Let us build a simple model of neuronal networks based on these facts.

A Discrete Dynamical System Model

A directed graph $D = [V_D, A_D]$ and integers n (size of the network), p_i (refractory period), th_i (firing threshold).

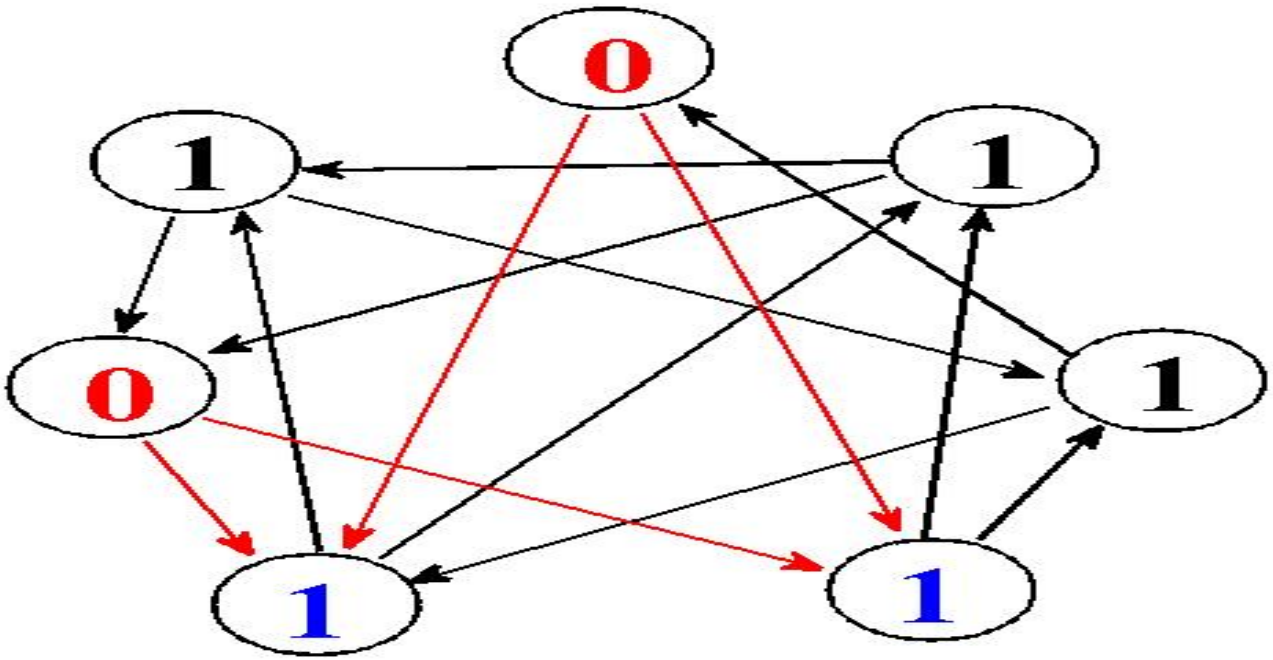
A state $\vec{s}(t)$ at the discrete time t is a vector:
$\vec{s}(t) = [s_1(t), \ldots, s_n(t)]$ where $s_i(t) \in \{0, 1, \ldots, p_i\}$ for each i.
The state $s_i(t) = 0$ means neuron i fires at time t.

Dynamics on the discrete network:

- If $s_i(t) < p_i$, then $s_i(t+1) = s_i(t) + 1$.
- If $s_i(t) = p_i$, and there exists at least th_i neurons j with $s_j(k) = 0$ and $< j, i > \in A_D$, then $s_i(t+1) = 0$.
- If $s_i(t) = p_i$ and there do not exist th_i neurons j with $s_j(t) = 0$ and $< j, i > \in A_D$, then $s_i(t+1) = p_i$.

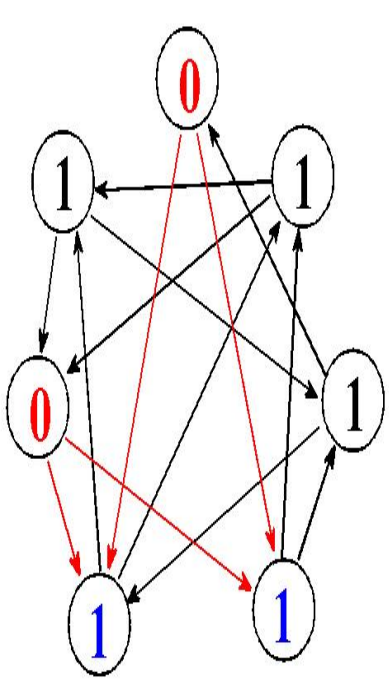
An Example

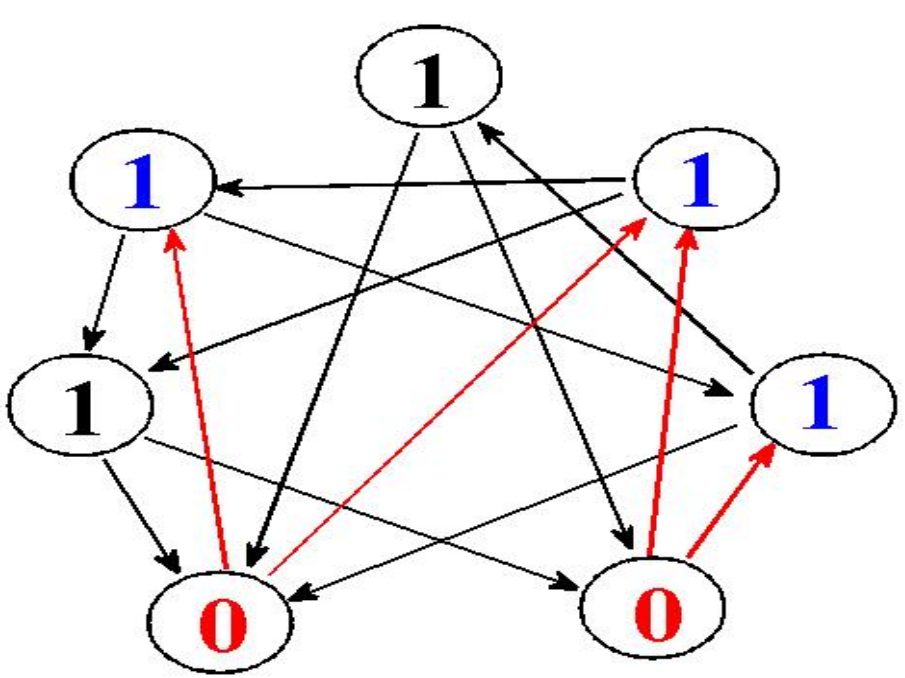
Assume that refractory period= 1 and threshold= 1.



An Example

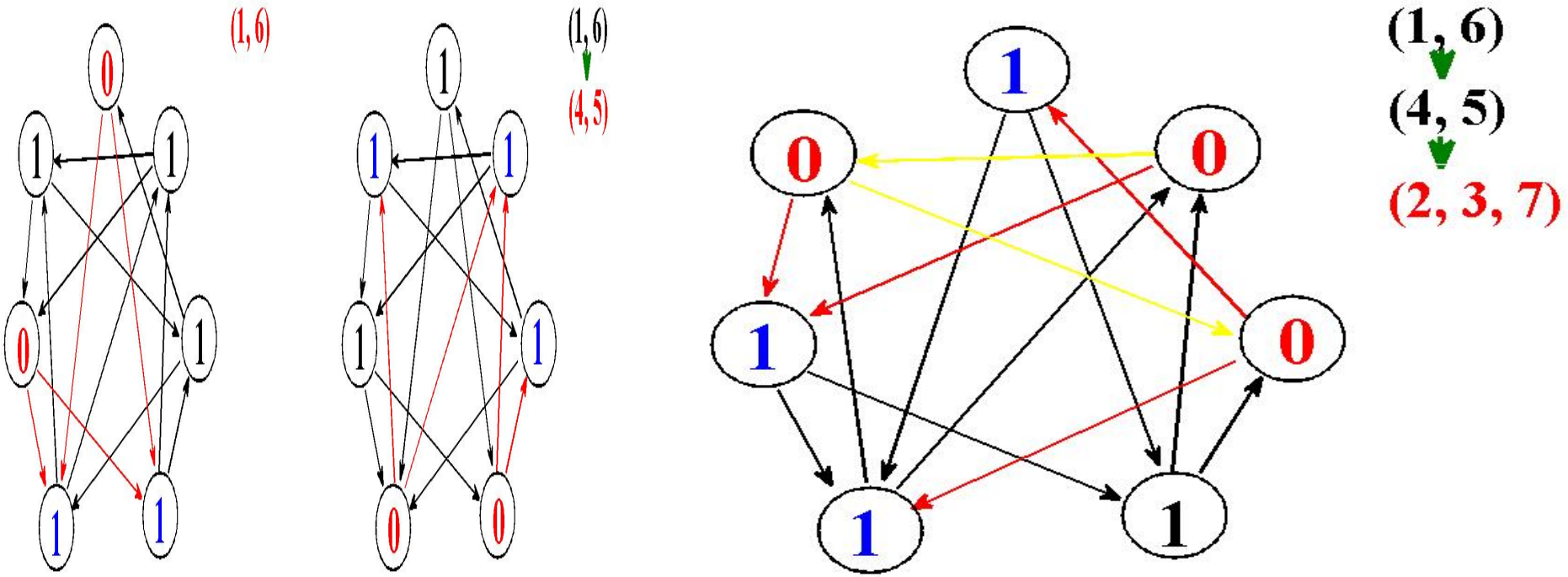
Assume that refractory period= 1 and threshold= 1.





An Example

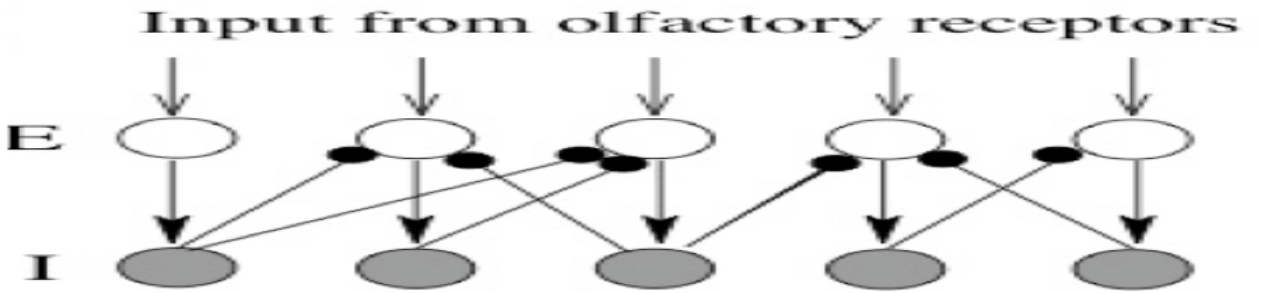
Assume that refractory period= 1 and threshold= 1.

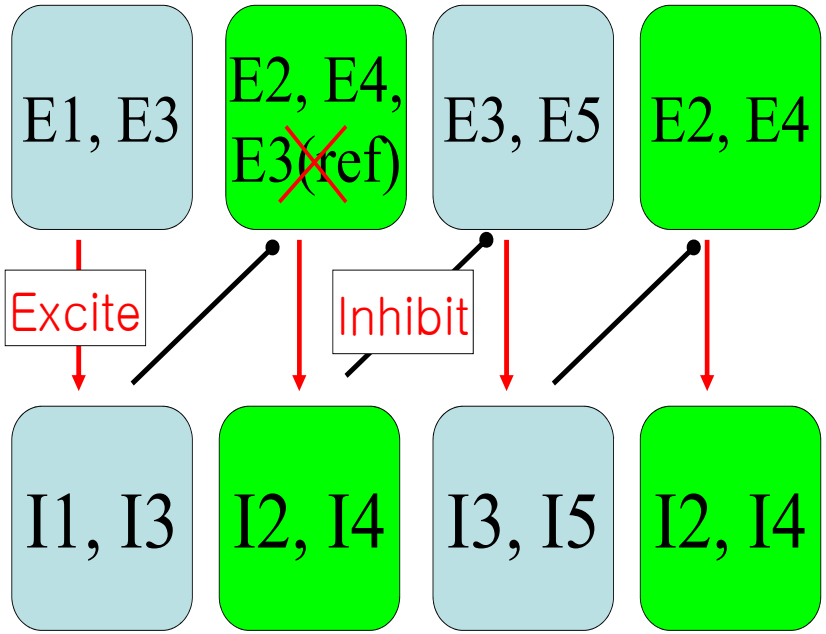


Is this model realistic?

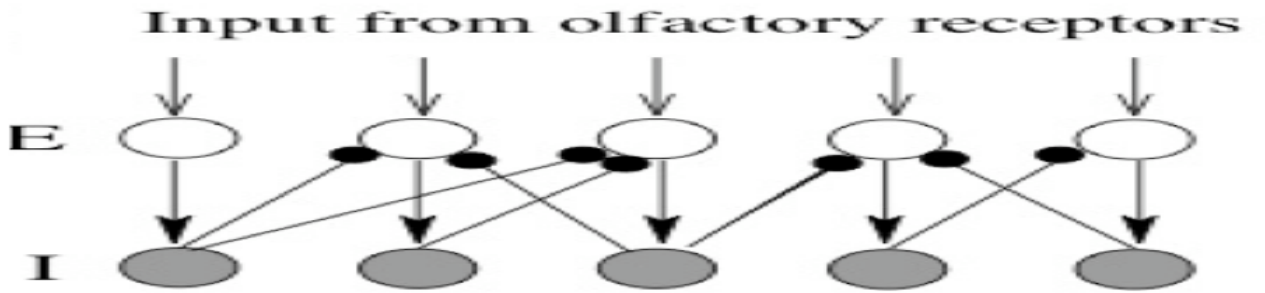
The model we just described does not explain dynamic clustering, since we built this phenomenon into the model right from the outset. But we would have a plausible explanation for the phenomenon if we could show, at least for some types of neuronal networks, that there is an exact correspondence between the ODE dynamics and the dynamics predicted by our discrete model.

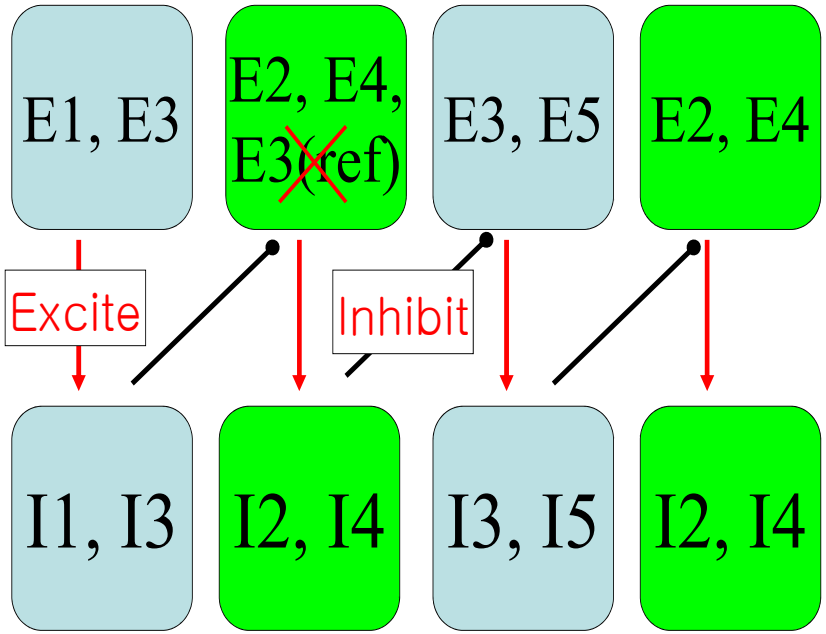
An Architecture





An Architecture



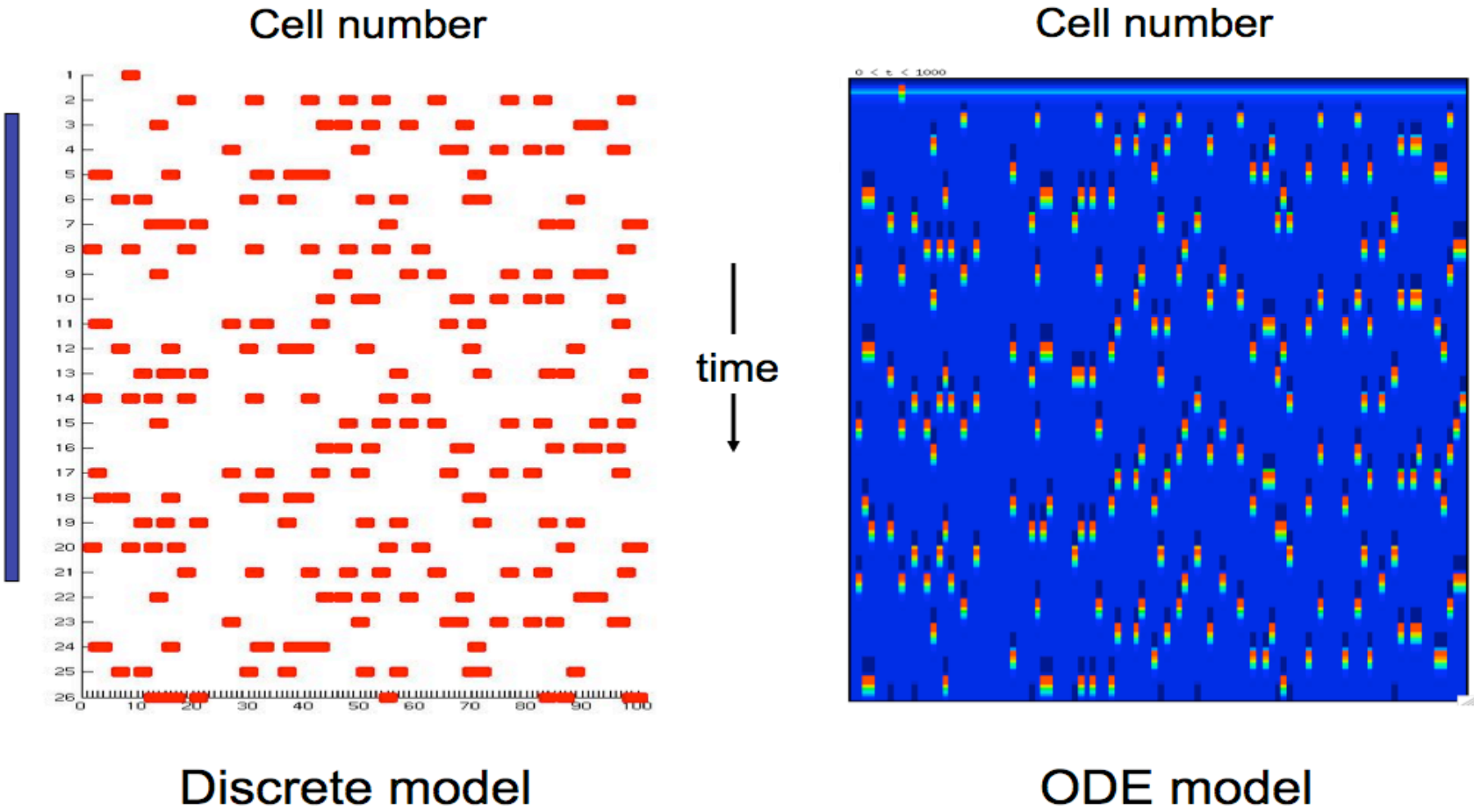


$$
\begin{aligned}
E1 &\rightarrow E2, E3 \\
E2 &\rightarrow E3 \\
E3 &\rightarrow E2, E4 \\
E4 &\rightarrow E5 \\
E5 &\rightarrow E4
\end{aligned}
$$

Assume: E-cells can excite one another v
interneurons.

The ODE Model Predicts Discrete Episodes

Consider 100 E-cells and 100 I-cells. Each E-cell excites one I-cell and each I-cell inhibits nine E-cells.



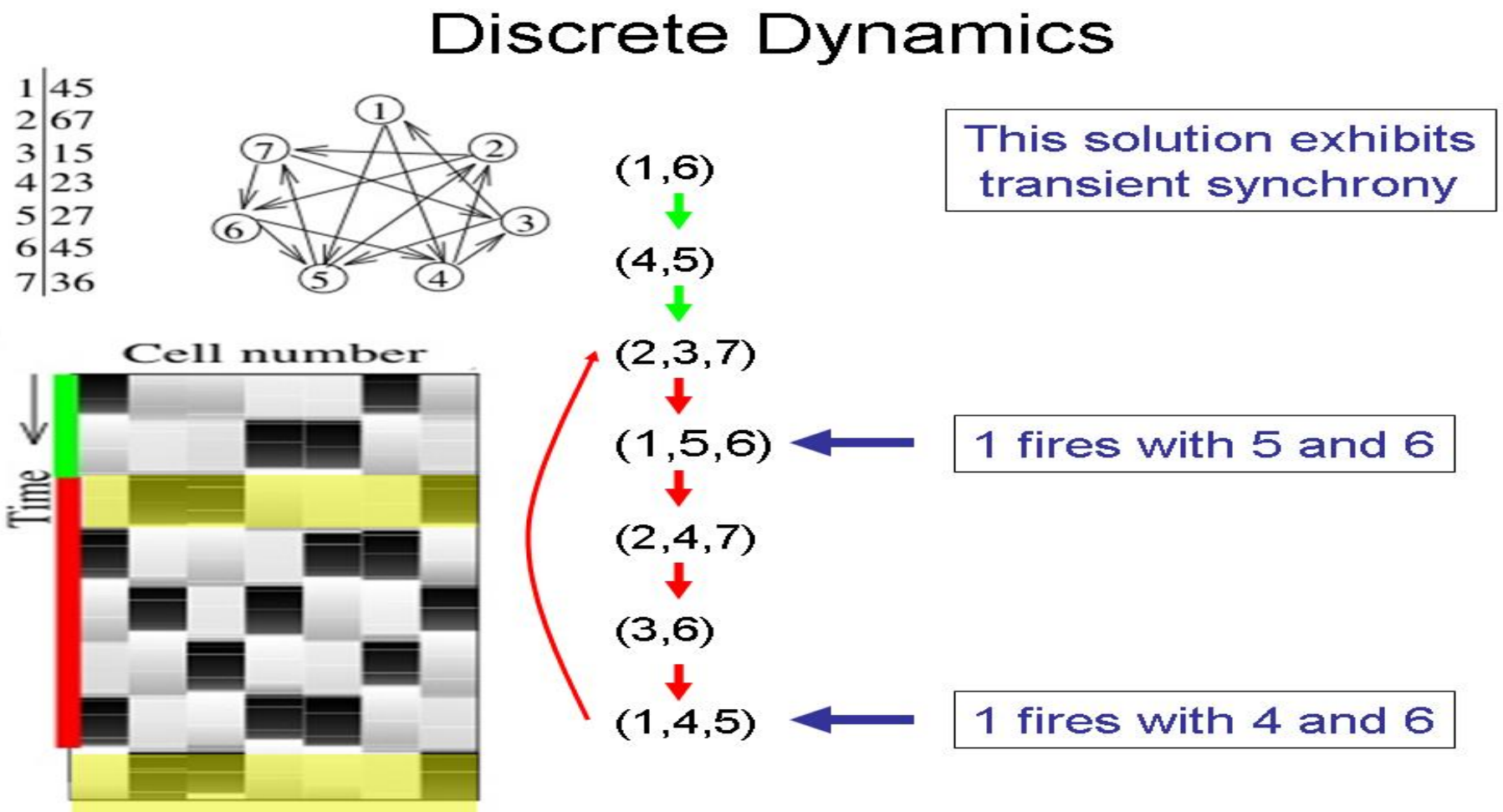
Reducing Neuronal Networks to Discrete Dynamics,

by Terman D, Ahn S, Wang X, Just W, Physica D. 2008

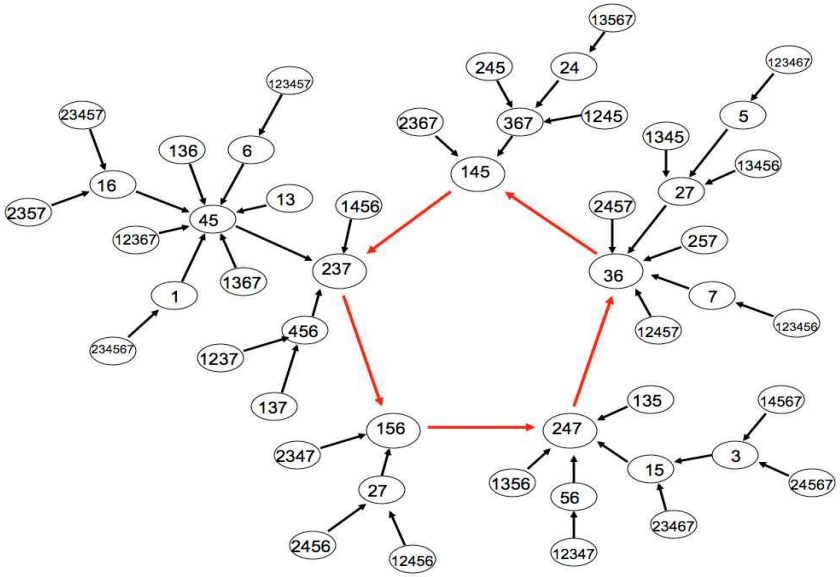
Theorem

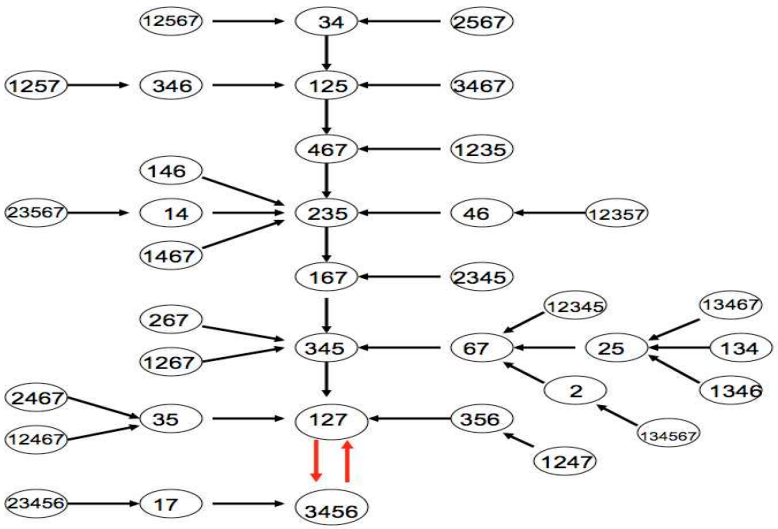
For the network architecture described above, if the intrinsic and synaptic properties of the cells are chosen appropriately, then there is an exact correspondence between solutions of the continuous and discrete systems for any connectivity between the excitatory and inhibitory cells.

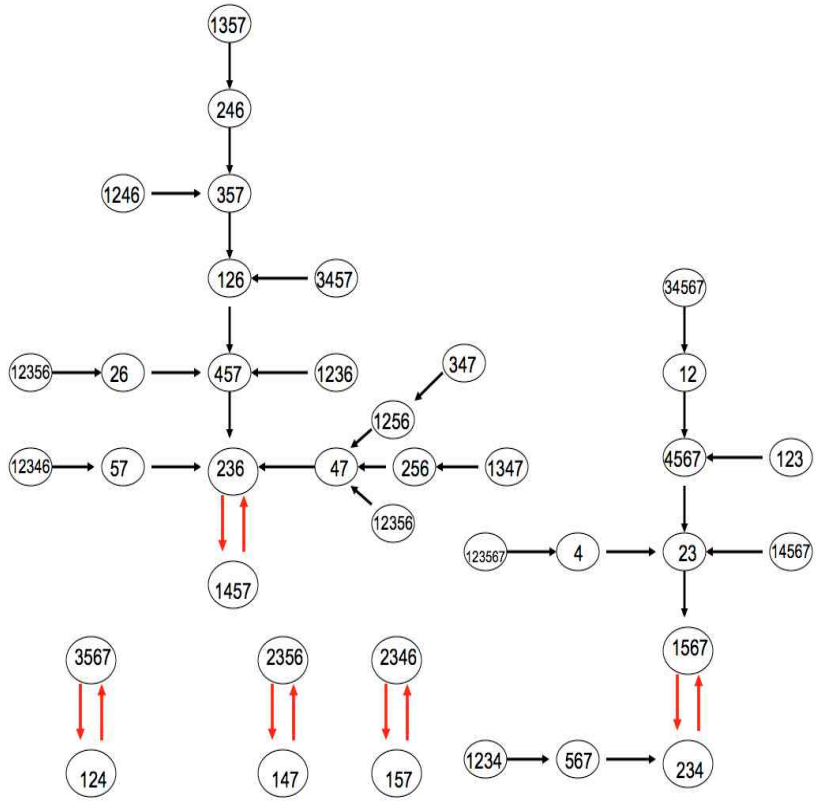
Assume that refractory period= 1 and threshold= 1.



Different Transients and Attractors







Studying the discrete model

For a given discrete model $N = < D, \vec{p}, \vec{th} >$ we may ask about the (possible, maximal, average)

- lengths of the attractors,
- number of different attractors,
- lengths of transients.

Some Special Digraphs

- Cyclic digraphs.
- Cyclic digraphs with one shortcut.
- Strongly connected digraphs: There is a directed path from every node to every other node.
- ....

What kind of dynamical properties are implied by these special connectivities?

Cyclic Digraph on n Nodes Sungwoo Ahn, Ph. D. Thesis 2010

Theorem

Let $\vec{p} = [p_1, \ldots, p_n]$, $\vec{th} = [1, \ldots, 1] >$, *and* $p^* = \max \vec{p}$. *Then*

- *If* $\vec{p} = [p, \ldots, p]$ *is constant, the length of any transient is at most* $2p - 1$.
- *If* $p^* < n$, *the length of any transient is at most* $n + p^* - 3$.
- *If* $p^* \geq n$, *the length of any transient is at most* $\max\{n + p^* - 1, 3n - 2\}$.

Theorem

Let $\vec{p} = [p_1, \ldots, p_n]$, $\vec{th} = [1, \ldots, 1] >$, *and* $p^* = \max \vec{p}$.

The number of different attractors is equal to the number of different necklaces consisting of n black or red beads where all the red beads occur in blocks of length that is a multiple of $p^* + 1$. *It is equal to*

$$\sum_{k=1}^{\lfloor \frac{n}{p^*+1} \rfloor} \left[\frac{1}{n - kp^*} \sum_{a \in \{\text{divisors of } \gcd(k, n-kp^*)\}} \phi(a) \binom{\frac{n-kp^*}{a}}{\frac{k}{a}} \right] + 1,$$

where ϕ *is Euler's phi function.*

Cyclic Digraph on n Nodes Sungwoo Ahn, Ph. D. Thesis 2010

Theorem

Let $\vec{p} = [p_1, \ldots, p_n]$, $\vec{th} = [1, \ldots, 1] >$, *and* $p^* = \max \vec{p}$.

The length of any attractor is a divisor of n.

Cyclic Digraph on n Nodes Sungwoo Ahn, Ph. D. Thesis 2010

Theorem

Let $\vec{p} = [p_1, \ldots, p_n]$, $\vec{th} = [1, \ldots, 1] >$, *and* $p^* = \max \vec{p}$.

The length of any attractor is a divisor of n.

Surprise: Numerical exploration suggest that the same is true for any strongly connected digraph.

Two Conjectures Sungwoo Ahn, Ph. D. Thesis 2010

Let n be the number of nodes; assume that all refractory periods and all firing thresholds are 1.

1. **Conjecture 1.** In strongly connected digraphs any attractor has length at most n.
2. **Conjecture 2.** In cyclic digraphs with one shortcut any attractor has length at most n.

Two Conjectures Sungwoo Ahn, Ph. D. Thesis 2010

Let n be the number of nodes; assume that all refractory periods and all firing thresholds are 1.

1. **Conjecture 1.** In strongly connected digraphs any attractor has length at most n.
2. **Conjecture 2.** In cyclic digraphs with one shortcut any attractor has length at most n.

Conjecture 2 was proved to be true in the thesis.

Two Conjectures Sungwoo Ahn, Ph. D. Thesis 2010

Let n be the number of nodes; assume that all refractory periods and all firing thresholds are 1.

1. **Conjecture 1.** In strongly connected digraphs any attractor has length at most n.
2. **Conjecture 2.** In cyclic digraphs with one shortcut any attractor has length at most n.

Conjecture 2 was proved to be true in the thesis.

Conjectures 1 is still open.

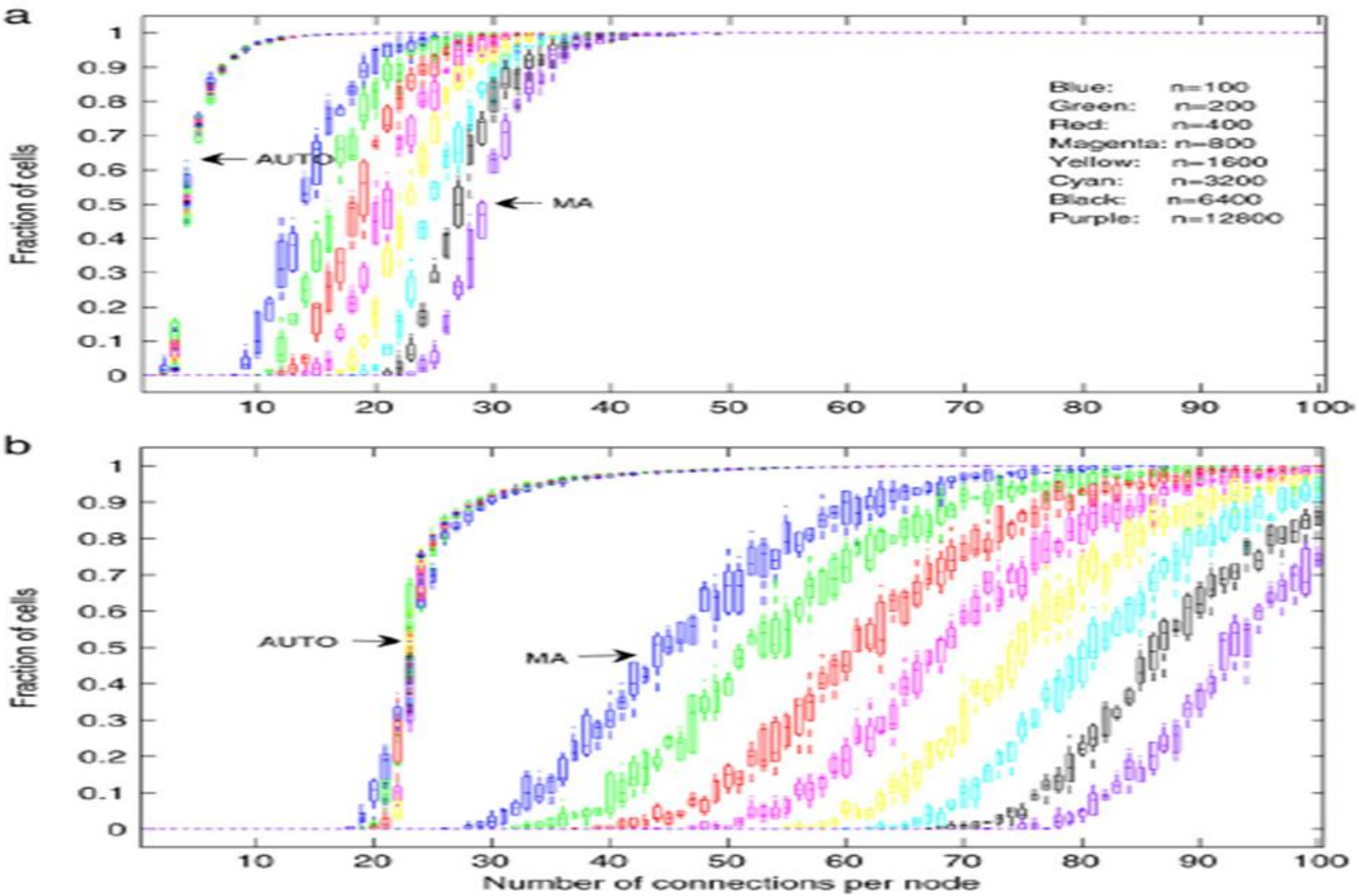
Some Special Objects

- $\vec{s}_{\vec{p}} = [p_1, \ldots, p_n]$ is the only steady state attractor.
- A *minimal attractor* is one in which each neuron either never fires or fires as soon as it reaches the end of its refractory period.
- An *autonomous set* consists of neurons that fire as soon as they reach the end of their refractory periods, regardless of the dynamics of neurons outside of this set.

Random Connectivities

- For given n, we randomly generate a digraph with n nodes by including each possible arc $< i, j >$ with probability $\rho(n)$; independently for all arcs (Erdős-Rényi random digraph).
- We randomly generate many initial conditions and collect statistics on minimal attractors and the size of the largest autonomous set.
- How do these properties depend on $\rho(n)$?

Results of the Simulations Just W, Ahn S, Terman D, Physica D. 2008



Minimal Attractors in Digraph System Models of Neuronal Networks, by Just W, Ahn S, Terman D, Physica D. 2008

Theorem

1. *The first phase transition at* $\rho(n) \sim \frac{\ln n}{n}$*:*
   - *Above this threshold: a generic initial state belongs to a (fully active) minimal attractor.*
   - *Below this threshold: a generic initial state will not belong to a minimal attractor.*
2. *The second phase transition at* $\rho(n) \sim \frac{c}{n}$*:*
   - *Above this threshold: almost all nodes will belong to the largest autonomous set.*
   - *Below this threshold: the relative size of the largest autonomous set will be close to zero.*

Directions for Further Research

- Another phase transition was detected for $\rho(n) \sim \frac{1}{n}$. Systematically explore what is going on in this region.
- Explore these phenomena for random digraphs other than Erdős-Rényi random digraphs (e.g., scale-free degree distributions).

Moving down to base level

- ecosystems
- populations
- organisms
- organs
- tissues
- cells
- organelles
- molecules

Gene regulation

What's going on in a cell biochemically at any given time is determined by which genes are being expressed at the time.

Gene regulation

What's going on in a cell biochemically at any given time is determined by which genes are being expressed at the time. The possible expression patterns form the states of the gene regulatory network, and everything else in biology is an emergent property of the dynamics of this network.

Gene regulation

What's going on in a cell biochemically at any given time is determined by which genes are being expressed at the time. The possible expression patterns form the states of the gene regulatory network, and everything else in biology is an emergent property of the dynamics of this network.

How can we study this network?

Boolean models of gene regulation

It is difficult to measure actual concentration of gene products (mRNA) with reasonable accuracy. But it is easy to take fuzzy snapshots of mRNA levels at different times even for all genes of an organism simultaneously using microarrays. These snapshots reveal only whether the expression level of a gene is high or low (sort of).

Boolean models of gene regulation

It is difficult to measure actual concentration of gene products (mRNA) with reasonable accuracy. But it is easy to take fuzzy snapshots of mRNA levels at different times even for all genes of an organism simultaneously using microarrays. These snapshots reveal only whether the expression level of a gene is high or low (sort of). One is thus tempted to model gene regulation with Boolean networks, where

- expression levels take only values 0 (low) and 1 (high),
- time proceeds in discrete steps,
- at each time step, all genes are updated simultaneously.

Boolean models of gene regulation

It is difficult to measure actual concentration of gene products (mRNA) with reasonable accuracy. But it is easy to take fuzzy snapshots of mRNA levels at different times even for all genes of an organism simultaneously using microarrays. These snapshots reveal only whether the expression level of a gene is high or low (sort of). One is thus tempted to model gene regulation with Boolean networks, where

- expression levels take only values 0 (low) and 1 (high),
- time proceeds in discrete steps,
- at each time step, all genes are updated simultaneously.

Every one of these assumptions is biologically unrealistic.

Boolean models work

Nevertheless, this approach often works surprisingly well and has already generated a number of real biological insights.

Boolean models work

Nevertheless, this approach often works surprisingly well and has already generated a number of real biological insights.

Why?

Boolean models work

Nevertheless, this approach often works surprisingly well and has already generated a number of real biological insights.

Why?

Can we mathematically explain why this works as well as it does?