Lengths of attractors and transients in neuronal networks with random connectivities

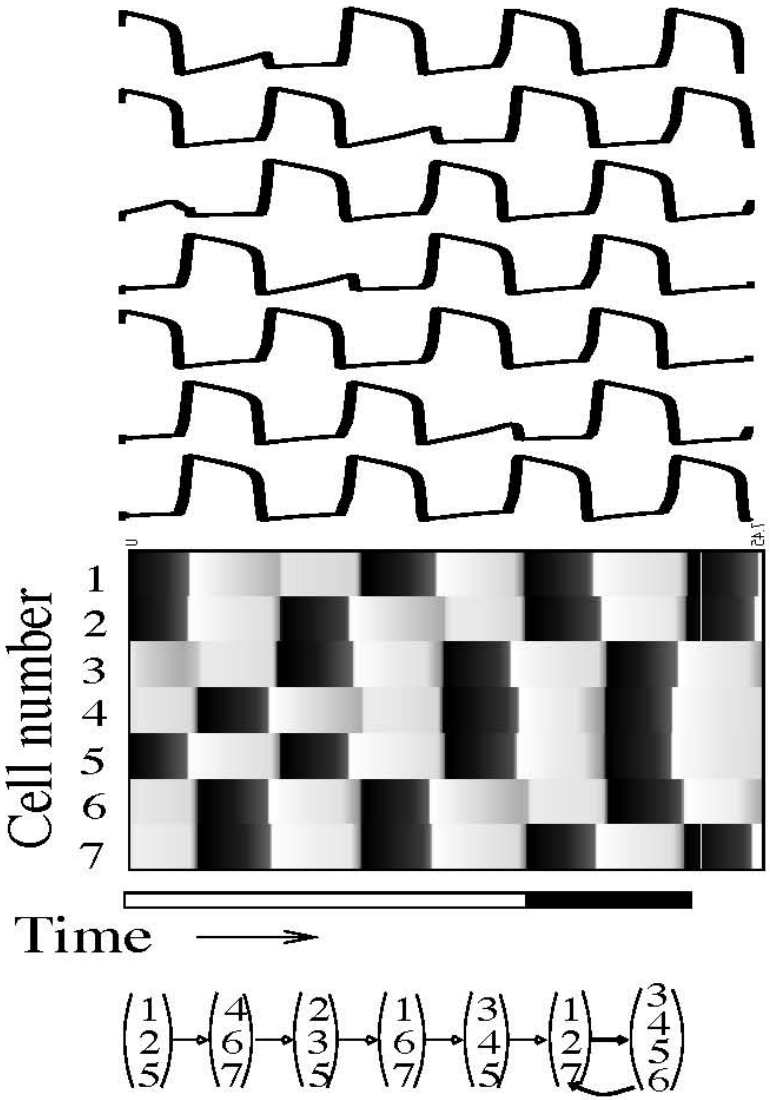
Winfried Just
Department of Mathematics, Ohio University

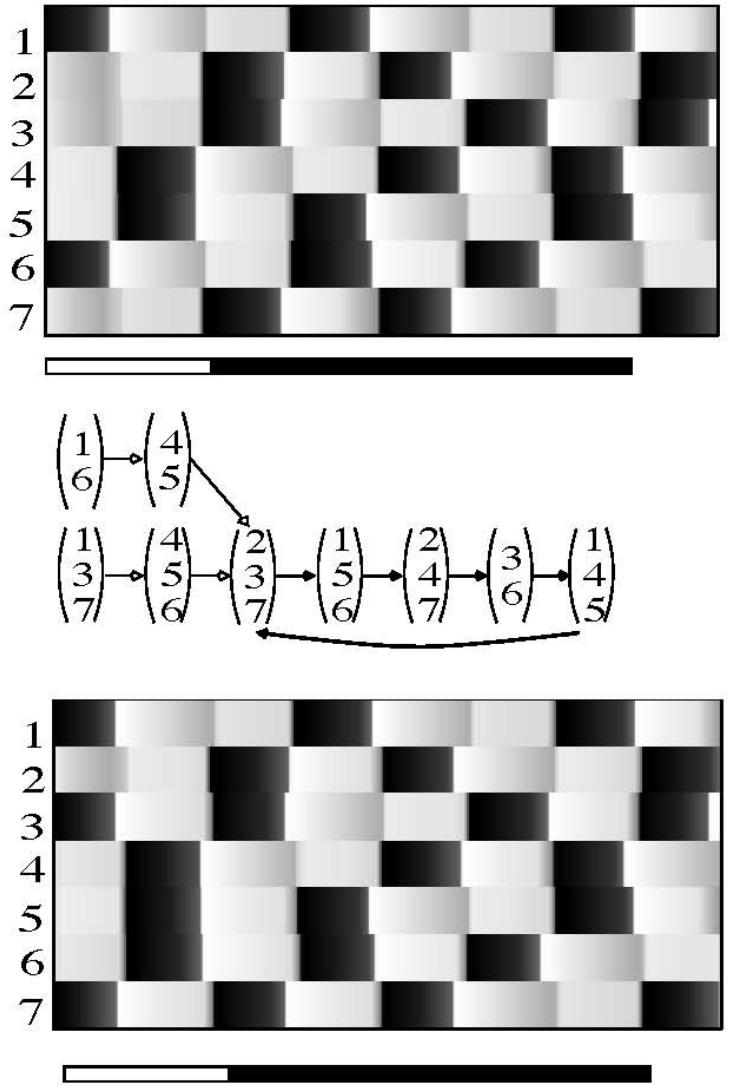
December 3, 2013
Seminarium Matematyki Dyskretnej
Uniwersytet im. Adama Mickiewicza
Poznań, Poland

A problem from mathematical neuroscience

Recordings from certain neuronal tissues (of real organisms) 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.

How dynamic clustering looks like





How to model dynamic clustering

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.

Why? How can we mathematically explain this phenomenon?

Of course, something like this will occur in many discrete-time dynamical systems, but this does not give an **explanation** as the episodes are built right into the definition of time.

Does the phenomenon occur in biologically realistic ODE models?

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

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

We have a plausible model, but ...

Let us call the model that we just described M.

- The model M does predict dynamic clustering.
- The architecture involves a layer of excitatory neurons and a layer of inhibitory neurons that mediate the firing of the excitatory neurons.
- Individual neurons are modeled by the the Hodgkin-Huxley Equations, which are nonlinear ODEs.
- These are difficult to analyze mathematically even for single neurons, let alone for large networks.

Can we study the dynamics of M by means of a simpler, approximate model N?

Mathematical neuroscience for the rest of us

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.
- Neurons are connected via synapses. Through a given synapse, the presynaptic neuron may send firing input to the postsynaptic neuron.
- 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 class of simple models N of neuronal networks based on these facts.

Discrete dynamical system models $N = < D, \vec{p}, \vec{th} >$

$D = ([n], A_D)$ is a digraph on $[n] = \{1, \ldots, n\}$,
$\vec{p} = (p_1, \ldots, p_n)$, where p_i is the refractory period of neuron i,
$\vec{th} = (th_1, \ldots, th_n)$, where th_i is the firing threshold of neuron i.

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.

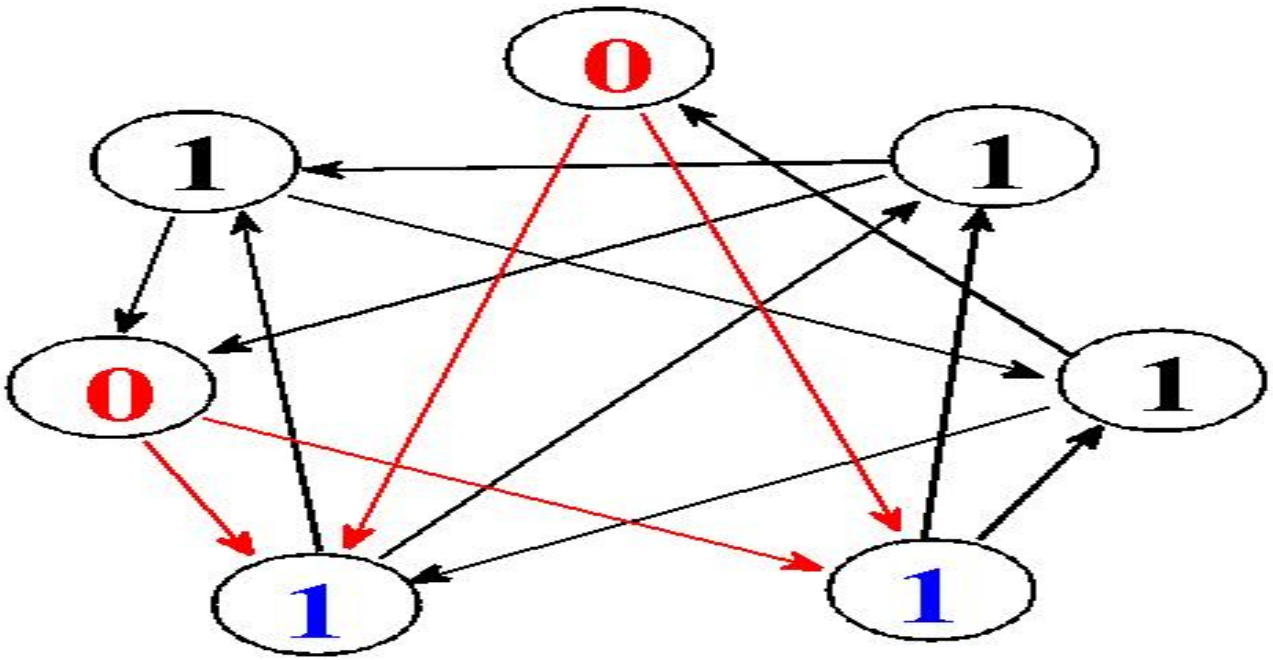
Dynamics of N:

- 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 $s_i(t+1) = p_i$.

If $p_i = 1$ for all i then N is a Boolean system.

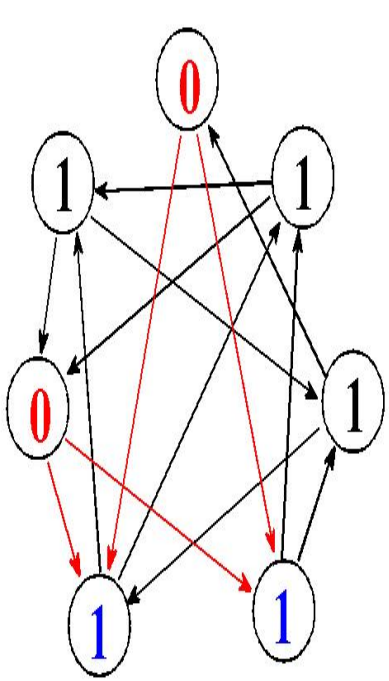
An example

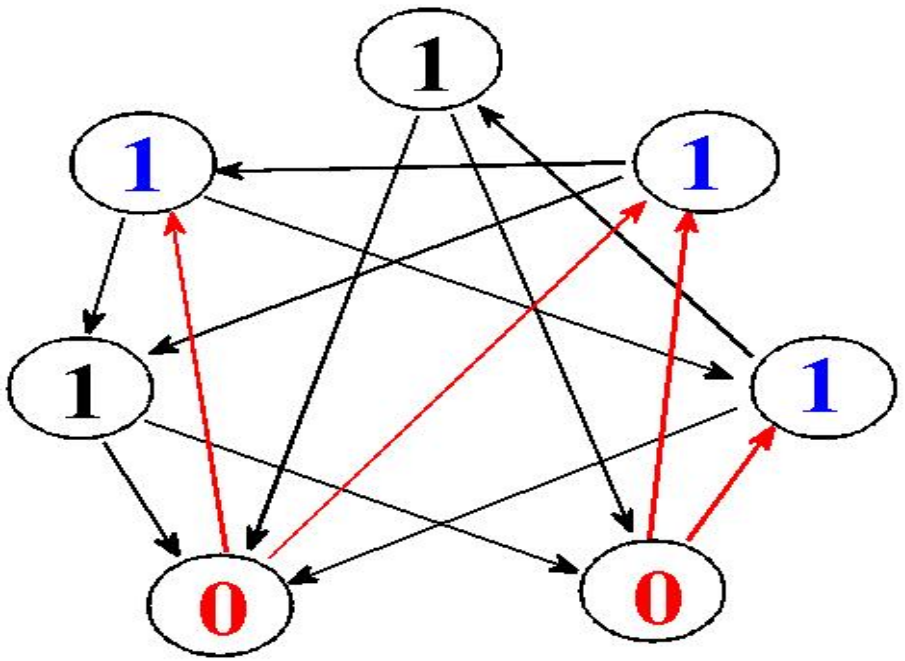
Assume refractory periods $\vec{p} = \vec{1}$ and firing thresholds $\vec{th} = \vec{1}$.



An example

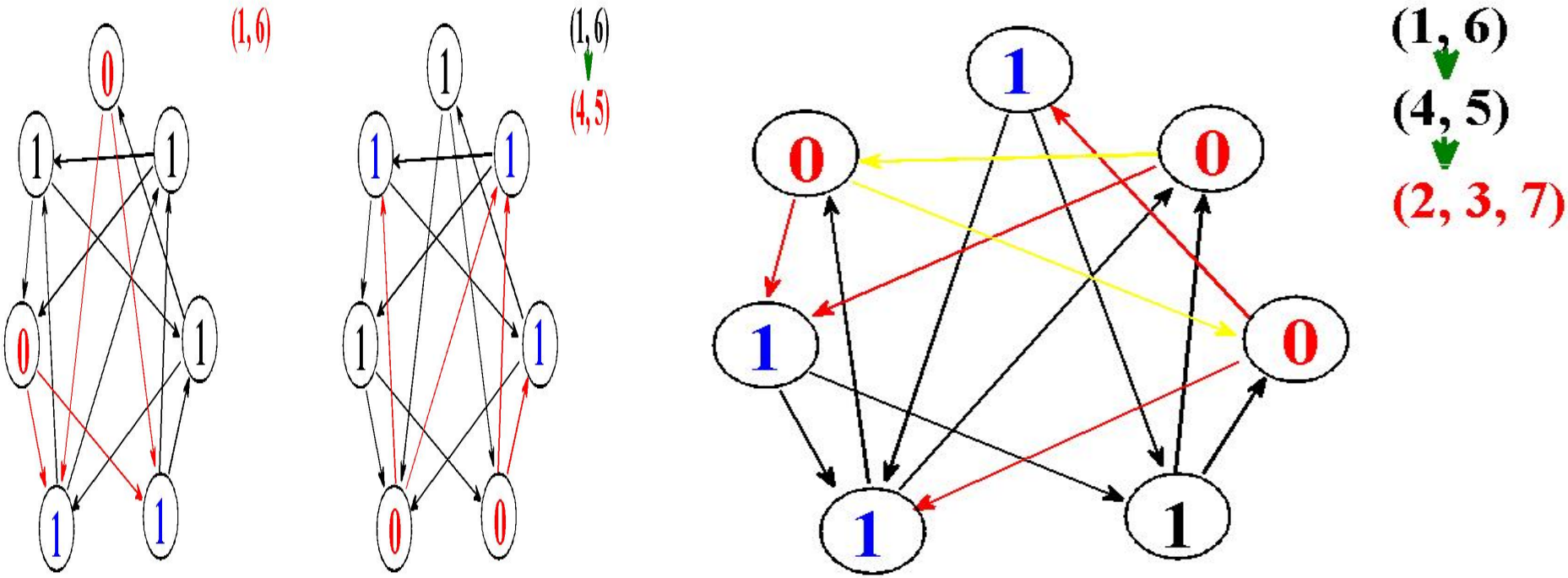
Assume refractory periods $\vec{p} = \vec{1}$ and firing thresholds $\vec{th} = \vec{1}$.





An example

Assume refractory periods $\vec{p} = \vec{1}$ and firing thresholds $\vec{th} = \vec{1}$.



Reducing neuronal networks to discrete dynamics,

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

Theorem

For each ODE model M of neuronal networks as described above, if the intrinsic and synaptic properties of the cells are chosen appropriately, the dynamics of M will exhibit dynamic clustering. Moreover, there exists a discrete model $N =< D, \vec{p}, \vec{th} >$ that correctly predicts, for a large region U of the state space of M and all times t, which neurons will fire during which episodes.

The theorem essentially tells us that as long as M is a biologically sufficiently realistic model of a given neuronal network, then so is the corresponding model N.

The discrete models N are much more tractable than the ODE models M. In particular, they permit us to study the dependence of the dynamics on the network connectivity D.

Why do we want to study this question for random connectivities?

Amazing fact: There exists a little roundworm, *Caenorhabditis elegans,* with 302 neurons, for which each single synapse has been mapped!

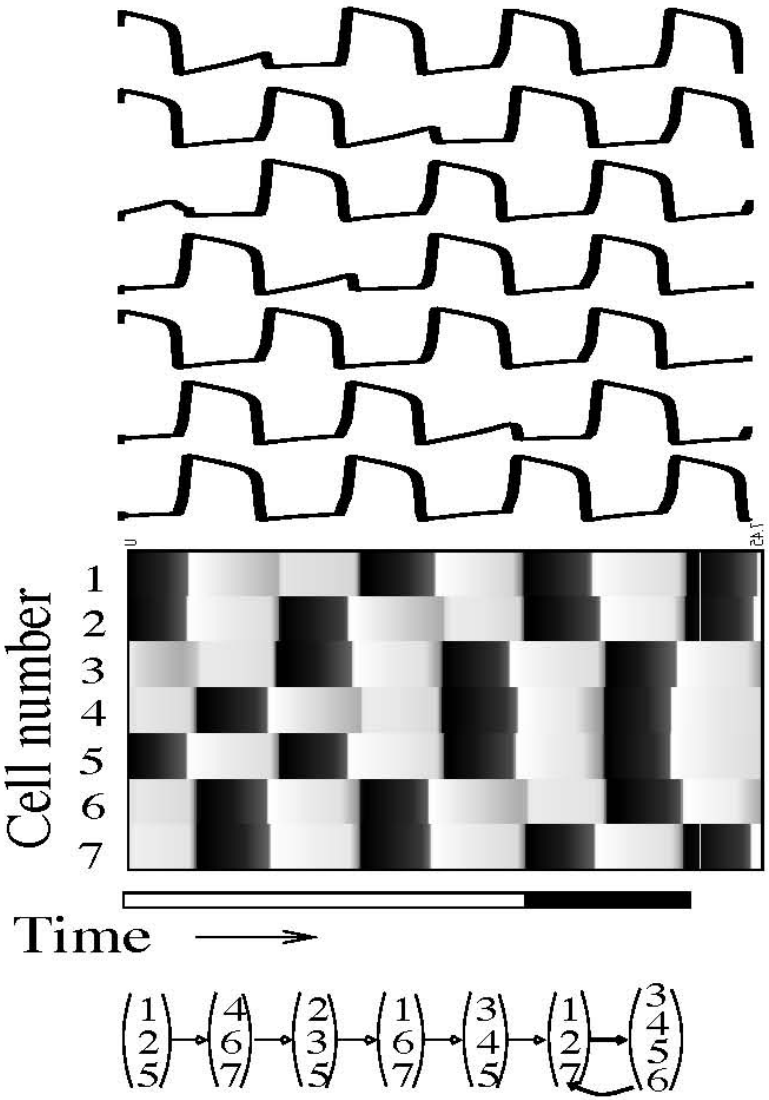
For higher organisms though, our knowledge of the actual neuronal wiring is only very fragmentary. We may, however, have some information about global network parameters such as the degree distribution. For example, there are about 10^{12} neurons and 10^{15} synaptic connections in the human brain, which gives a mean degree of about 1000 for the network.

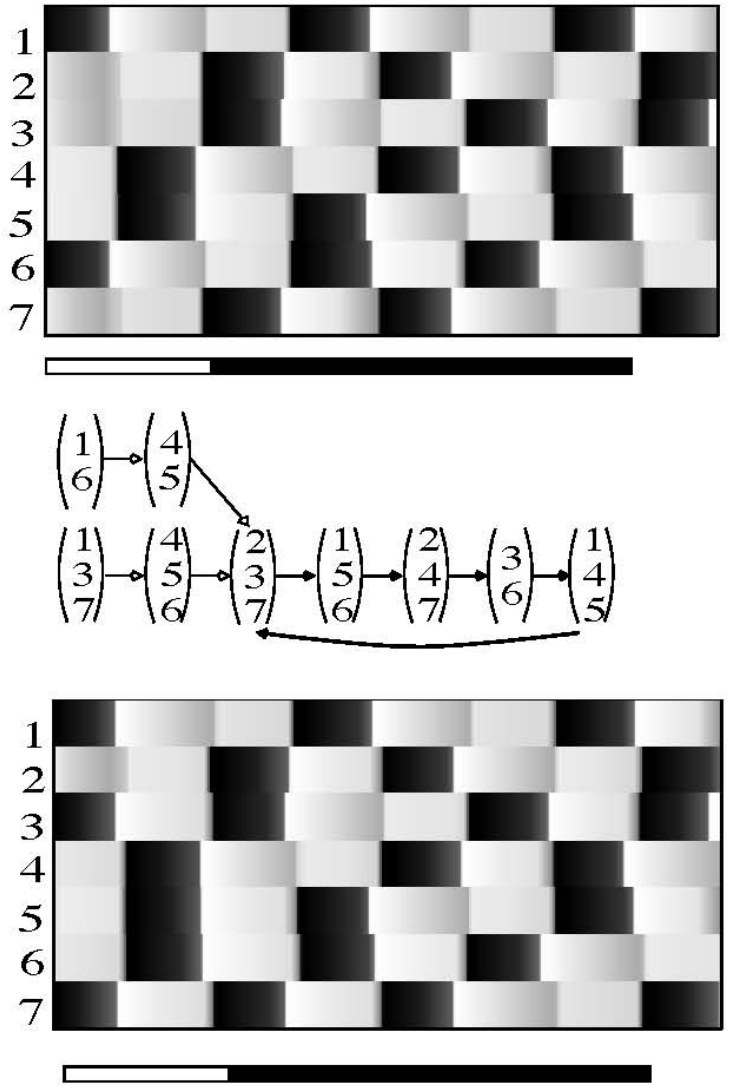
The architecture of actual neuronal networks has been shaped by evolution and to some extent by learning, both of which are **stochastic processes.** Thus it is reasonable to assume that the actual architecture exhibits features that are reasonably typical for a relevant probability distribution on digraphs.

Basics of network dynamics

- The trajectory of initial state $\vec{s}(0)$ is the sequence $(\vec{s}(0), \vec{s}(1), \ldots, \vec{s}(t), \ldots)$
- States that are visited infinitely often by the trajectory are called persistent states. Since the sate space is finite, every trajectory must eventually reach a persistent state. The set of these persistent states is called the attractor of the trajectory.
- Transient states are visited only once. Their sequence is an initial segment of the trajectory, called its transient (part).
- The vector $\vec{p} = (p_1, \ldots, p_n)$ is the unique steady state of a network N. The set $\{\vec{p}\}$ is the only steady state attractor. All other attractors, if such exist, are called periodic attractors.

Examples of transients and periodic attractors





The basic setup

Just W and Ahn S, in progress

- Let π be a function that assigns to each positive integer n a probability $\pi(n)$.
- Randomly draw an Erdős-Rényi digraph D on $[n]$ where each potential arc is included with probability $\pi(n)$.
- Fix $1 \leq p_* \leq p^*$ and $1 \leq th_* \leq th^*$.
- Randomly draw $\vec{p}$ and $\vec{th}$ from the uniform distribution of all n-dimensional vectors with $p_* \leq p_i \leq p^*$ and $th_* \leq th_i \leq th^*$ for all $i \in [n]$.
- Randomly draw an initial condition $\vec{s}(0)$ in the chosen network.
- Let α be the length of the attractor and let τ be the length of the transient of the trajectory of $\vec{s}(0)$.

In the remainder of this talk we will for simplicity make the standing assumption that $p^* = th^* = 1$.

Scaling laws for α and τ

We are interested in how the medians and all fixed percentiles of α and τ scale as $n \to \infty$.

Why percentiles?

If the means of α or τ do scale differently from the percentiles, then this must be due to rare outliers. The majority of studies on network dynamics in mathematical biology relies on simulations. These are unlikely to detect extremely rare outliers. Thus theoretical results on the scaling of fixed percentiles will in general be better predictors of simulation results than theoretical results on the means.

Why are these scaling laws relevant?

Several of the neuronal tissues in which dynamic clustering has been observed are part of olfactory systems. There is an ongoing debate among neuroscientists whether odors are coded in the attractors or in the transients of neuronal dynamics. The first scenario requires sufficiently many different (long) attractors, the second requires sufficiently long transients.

Classes of Boolean systems can be roughly categorized as those exhibiting predominantly ordered dynamics and those exhibiting predominantly chaotic dynamics. The former are characterized (among other hallmarks) by relatively short transients and attractors; the latter by relatively long ones. The difference between "short" and "long" often corresponds to polynomial *vs.* exponential scaling with system size n. The capability of the system to perform complex computations appears to require that its dynamics falls into the critical regime, right at the boundary between order and chaos.

Acyclic connectivities

Proposition

Assume D is acyclic. Then

(i) $\alpha = 1$

(ii) $\tau + 1$ *does not exceed the maximal length* γ *of a directed path in* D.

Proof: If node i fires at time $T > 0$, then there must exist a sequence of nodes $(i = i_T, i_{T-1}, \ldots, i_0)$ with

- $s_{i_t}(t) = 0$ for all $t \leq T$,
- $< i_t, i_{t+1} > \in A_D$ for all $t < T$.

If $\alpha > 1$, we have such T with $T \geq n$. If D is acyclic, this sequence must contain pairwise distinct nodes. $\square$

For $\pi(n) = \frac{c}{n}$ with $c < 1$, this simple observation immediately implies that at least some percentiles of α scale like 1, and the same percentiles of τ scale like $O(\log n)$.

The effect of one directed cycle

Proposition

Assume D is a directed cycle of length L. Then α is a divisor of L.

Let us call D supersimple if it is either acyclic or contains exactly one directed cycle C and for every node j outside of C there exists at most one directed path from j to C and at most one directed path from C to j.

Lemma

Assume D is supersimple and contains a directed cycle of length L. Then

(i) α is a divisor of L.

(ii) The percentiles of τ scale like $\Theta(\gamma)$.

α and τ in terms of upstream components

If $\pi(n) = \frac{c}{n}$ for $c < 1$, then with probability arbitrarily close to 1 as $n \to \infty$, all upstream components of D will be supersimple.

The internal dynamics of the upstream component $UC(i)$ of node i will not be influenced by the remainder of the system.
Let α_i, τ_i denote the length of the attractor and the transient in the internal dynamics of $UC(i)$. Then

$\alpha = lcm\{\alpha_i : \ i \in [n]\}$ and $\tau = max\{\tau_i : \ i \in [n]\}$.

These observations imply

Theorem

Assume $\pi(n) = \frac{c}{n}$ with $c < 1$. Then

(i) Each fixed percentile of α scales like $O(1)$.

(ii) Each fixed percentile of τ scales like $\Theta(\log n)$.

Thus the subcritical case exhibits hallmarks of highly ordered dynamics.

Approaching the critical window from below

Theorem

Assume $\pi(n) = \frac{1-n^{-\beta}}{n}$, where $0 < \beta < 1/4$. Then with probability arbitrarily close to 1 as $n \to \infty$

(i) Every upstream component of D will be supersimple.

(ii) γ, and hence τ, scales like $O((\log n)n^{\beta})$.

(iii) γ, and hence τ, scales like $\Omega(n^{\beta})$.

(iv) $\alpha \leq e^{\sqrt{n \ln n}+o(1)}$ and thus scales subexponentially.

(v) $\alpha \geq e^{\Omega(\log n \log\log n)}$ and hence scales faster than any polynomial function.

Thus we observe hallmarks of the critical regime for the dynamics.

The proof of point (iv)

Theorem

Assume $\pi(n) = \frac{1-n^{-\beta}}{n}$, *where* $0 < \beta < 1/4$. *Then with probability arbitrarily close to 1 as* $n \to \infty$

(iv) $\alpha \leq e^{\sqrt{n \ln n}+o(1)}$ *and thus scales subexponentially.*

Proof: Let $\{L_1, \ldots, L_r\}$ denote the set of all lengths of directed cycles in D. If every upstream component of D is supersimple, then

$\alpha \leq lcm\{L_1, \ldots, L_r\}$,

and, moreover, the directed cycles of D are pairwise disjoint.

It follows that $\alpha \leq g(n)$, where $g(n) \sim e^{\sqrt{n \ln n}+o(1)}$ is Landau's function. $\square$

The proof of (v)

Theorem

Assume $\pi(n) = \frac{1-n^{-\beta}}{n}$, *where* $0 < \beta < 1/4$. *Then with probability arbitrarily close to 1 as* $n \to \infty$

(v) $\alpha \geq e^{\Omega(\log n \log\log n)}$ *and hence scales faster than any polynomial function.*

Outline of the Proof: First we show that for every given probability $q < 1$ and positive integer k there exist a positive integer $K = K(q,k)$ and a positive real κ such that for all sufficiently large n, with probability $> q$, the digraph D will contain a set of more than $\frac{k}{\kappa}$ directed cycles of lengths $L_i = K_i P_i$, where $K_i < K < n^{\kappa} < P_i$ and the numbers P_i are distinct primes. Such a set can be obtained from Euler's result on reciprocals of primes by the second moment method.

Then $\prod P_i > n^k$ and this implies $lcm\{L_1, \ldots, L_r\} > n^k$, but so far we are only guaranteed that $\alpha \leq lcm\{L_1, \ldots, L_r\}$.

The proof of (v), continued

Note that α_i in the upstream component of the cycle of length K_iP_i either is a multiple of P_i, or satisfies $\alpha_i \leq K_i$.

To complete the proof, we showed that the probability of the inequality $\alpha_i \leq K_i$ becomes negligible for large n.

This seems intuitively obvious, as $L_i = K_iP_i$ with $K_i < K \ll P_i$ implies that there are vastly more attractors of length $\geq P_i$ than attractors of length $\leq K_i$ in this upstream component.

But the distribution of the sizes of their basins of attraction (number of initial conditions from which the attractor is reached) is not uniform. This poses some technical difficulties that required a more elaborate argument. $\square$

What happens right inside the critical window?

One would conjecture that for $\pi(n) = \frac{1}{n}$ both α and τ scale even faster. Simulations studies indicate as much.

However, our arguments so far relied on having almost perfect control over the dynamics, as both in the subcritical case and at the lower end of the critical window the important structures, directed cycles and long directed paths, are neatly segregated into separate upstream components. Higher up in the critical window we lose all such control.

Thus it seems very challenging to develop good tools for exploring the dynamics of our system deep inside the critical window.

What happens above the critical window?

When $\pi(n) = \frac{c}{n}$ for some constant $c > 1$, we regain a certain amount of control.

In this case we can assume that there exists a giant strongly connected component GC. If we remove it together with all nodes downstream of it, the remaining digraph will exhibit the same features as in the subcritical case: small and supersimple upstream components.

This essentially leaves us with investigating what happens inside the giant component, and it seems that together with the giant component there appears a new feature that will simplify the dynamics.

Definition

A node i is eventually minimally cycling if there are only finitely many times t with $s_i(t) = s_i(t+1) = p_i$.

The supercritical case: The first minimally cycling node in the giant component

Proposition

There exists $1 \leq c_{crit} \leq 2$ *such that for* $\pi(n) = \frac{c}{n}$ *with* $c > c_{crit}$, *with probability approaching 1 as* $n \to \infty$,

(i) The giant strongly connected component will contain an eventually minimally cycling node.

(ii) The smallest time t_{first} *at which some node in the giant strongly connected component becomes minimally cycling is bounded by a constant that depends on* c *but not on* n.

Proof: We show that $c = 2$ works here.

If D contains a directed cycle C of even length such that the initial state has alternating zeros and ones along this cycle, then each node in C will be (eventually) minimally cycling, starting from time $t = 0$.

The supercritical case: The first minimally cycling node in the giant component

Proposition

If $\pi(n) = \frac{c}{n}$ with $c > 2$, with probability approaching 1 as $n \to \infty$, the giant strongly connected component will contain a minimally cycling node.

Proof (continued): For any given even L, the expected number of such directed cycles of length L approaches $\frac{c^L}{L2^{L-1}}$ as $n \to \infty$.

A standard second-moment argument now shows that for any given constant A, with probability approaching 1 as $n \to \infty$, the digraph D will contain such directed cycles of length $L > A$.

Since the probability that the part of D outside of the giant strongly connected component will contain directed cycles of length $\geq A$ becomes arbitrarily small for sufficiently large A, the result follows. $\square$

The supercritical case: All nodes in the giant component are minimally cycling

Lemma

Suppose node j is eventually minimally cycling.

(i) If node i is such that there exists a directed path in D from j to i, then i is also eventually minimally cycling.

(ii) Let τ_{all} denote the time it takes from the moment τ_{first} when the first node of the giant strongly connected component of D becomes minimally cycling until all nodes in DC, the set of nodes downstream of the giant strongly connected component become eventually minimally cycling.
Then all percentiles of τ_{all} scale like $O(2^{diam(DC)})$, where $diam(DC)$ denotes the maximum length of the shortest directed path from $j \in GC$ to $i \in DC$.

Attractors and transients in the supercritical case

For $c > c_{crit}$ we get the following picture:

- $\prod \alpha_i$, taken over i that is **not** downstream of of the giant strongly connected component, scales like a constant.
- All nodes that are downstream of the giant strongly connected component are minimally cycling in the attractor, which adds a factor of most 2 to α.
- $\max \tau_i$, taken over i that are **not** downstream of of the giant strongly connected component, scales like $\Theta(\log n)$.
- For i that are downstream of of the giant strongly connected component, $\max \tau_i = \tau_{first} + \tau_{all}$, which scales like $O(n^k)$ for some $k = k(c)$.

Thus, as in the supercritical case, all percentiles of α scale like a constant, and all percentiles of τ scale polynomially and like $\Omega(\log n)$.

Open problems

Problem 1: Find the exact scaling law for the length τ of the transient in the supercritical case, or at least narrow the gap between $\Omega(\log n)$ and $O(n^{k(c)})$.

Problem 2: Assume $\pi(n) = \frac{1-n^{-\beta}}{n}$, where $0 < \beta < 1/4$. Find the exact scaling law for the length τ of the transient.

At this time we know that it is between $\Omega(n^{\beta})$ and $O((\log n)n^{\beta})$.

We also know that in this case the scaling law must be the same as the one for the maximum length of a directed path in D.

Open problems, continued

Problem 3: Assume $\pi(n) = \frac{1-n^{-\beta}}{n}$, where $0 < \beta < 1/4$. Find the exact scaling law for the length α of the attractor.

At this time we know that it is between $e^{\Omega(\log n \log\log n)}$ and Landau's function $g(n) \sim e^{\sqrt{n \ln n}+o(1)}$.

Problem 4: Does there exist, for any n, a network $< D, \vec{1}, \vec{1} >$ on $[n]$ that contains any attractor of length $\alpha > g(n)$?

This question is entirely deterministic, but it is conceivable that it is easier to show the existence of such examples probabilistically rather than by an explicit construction.

Open problems, continued

Problem 5: Does there exist $\pi(n)$ such that $\tau(n)$ scales faster than any polynomial?

At this time we don't even know whether there exists $\pi(n)$ where $\tau(n)$ scales like $\Omega(n)$.

We have deterministic examples for which $\tau(n)$ roughly scales like Landau's function $g(n) \sim e^{\sqrt{n \ln n}+o(1)}$.

If there exists a sequence of probabilities that gives a positive answer, it must be in the critical window $\frac{1-n^{-1/4+o(1)}}{n} \leq \pi(n) \leq \frac{c_{crit}+o(1)}{n}$.

Problem 6: More generally, explore what happens for $\pi(n)$ in the critical window.

This seems quite challenging and appears to require new methods.

Open problems, continued

Problem 7: Investigate α and τ for the case when $th_* > 1$, that is, when all firing thresholds exceed 1.

We have some rudimentary results, but a full characterization will require new methods. In particular, if $UC(i)$ is simple, then $\alpha_i = 1$. In other words, the existence of periodic attractors requires more complicated structures inside D than just directed cycles.

Problem 8: Investigate the behavior of α and τ for other types of random connectivities.

Empirical results indicate that the degree distributions in actual neuronal networks may be closer to scale-free than to normal. Thus results on Erdős-Rényi random connectivities may not be directly applicable to neuroscience.

But we had to start our investigations somewhere.

Open problems, continued

Problem 9: Investigate how the Hamming distance $H(\vec{s}(t), \vec{s}^*(t)) = \#\{i : s_i(t) \neq s_i^*(t)\}$ evolves.

If the Hamming distances between a randomly chosen initial condition $\vec{s}(0)$ and a small perturbation $\vec{s}^*(0)$ of it tends to significantly increase, then we have decorrelation.

Decorrelation indicates sensitive dependence on initial conditions and is a chaos-like property. Neuroscientists believe that some form of decorrelation is needed if odors are to be coded in transients.

Problem 10: Try to generalize our results to systems with other types of rules.

For example, the result that $\alpha = 1$ for acyclic D generalizes to any type of network dynamics where the updating rules do not allow sustained oscillations of a node under constant external input.

Open problems, continued

Back to our paper.

Problem 11: What is c_{crit}, really?

We implicitly defined c_{crit} as the minimum value such that for every constant $c > c_{crit}$ and $\pi(n) = \frac{c}{n}$, with probability approaching 1 as $n \to \infty$, an eventually minimally cycling node will appear in the giant component at time τ_{first}, where τ_{first} is supposed to scale like a constant.

We showed that $1 \leq c_{crit} \leq 2$.

We conjecture that $c_{crit} = 1$.

Simulation results indicate as much

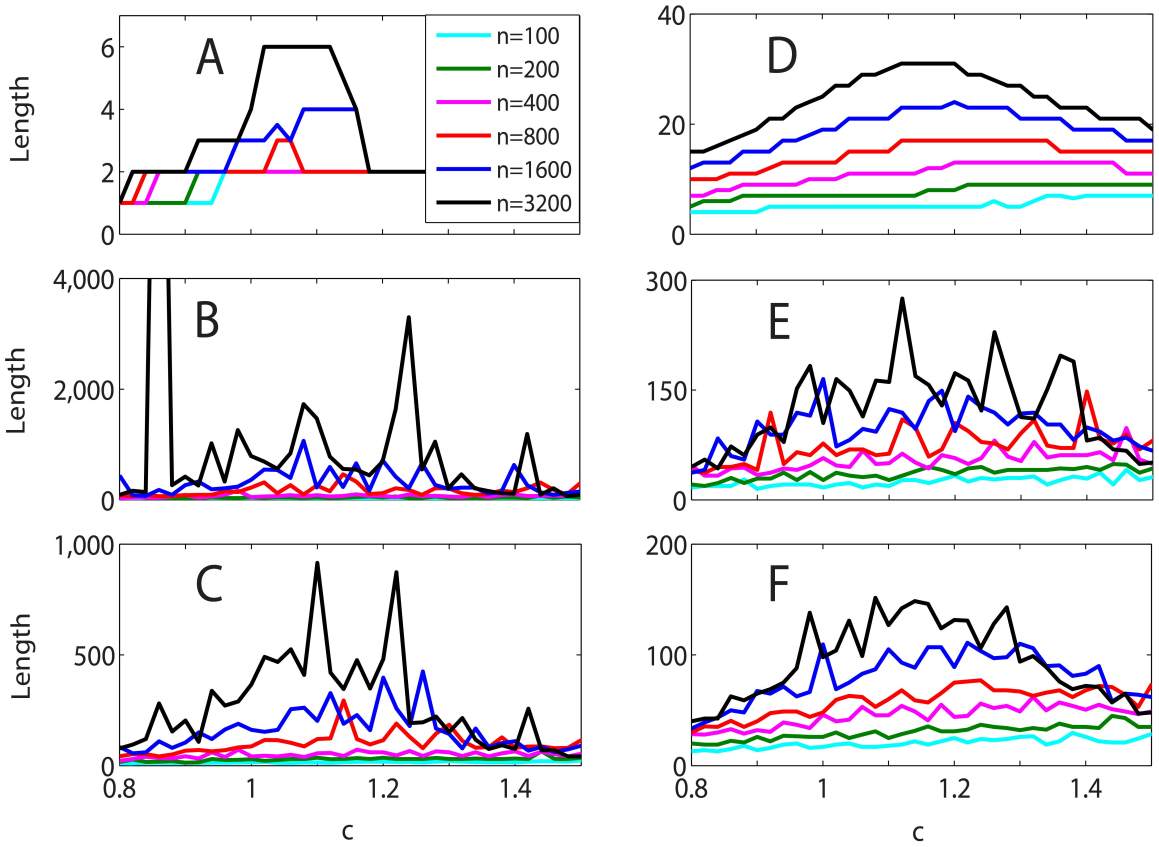


Figure: Lengths of the attractors and the transients for $\vec{p} = t\vec{h} = \vec{1}$. (A-C) Median, maximum, and 99.9th percentile of α. (D-F) Median, maximum, and 99.9th percentile of τ.

Modified Erdős-Rényi (di)-graphs

Consider the following procedure for producing a (di)-graph:

- Consider an algorithm A that takes as input a (di)graph D on $[n]$ with some labeling of the vertices with a fixed set of labels, and outputs another labeled (di)graph $A(D)$ on $[n]$.
- The algorithm decides whether or not $< i, j >$ is an arc (edge) of $A(D)$ only based on the structure and labels of subgraph induced by all nodes that can be reached from i or j via a (directed) path of length $\leq N$, where N is fixed and does not depend on n.
- Let D be an Erdős-Rényi (di)graph.
- Generate the labels independently, with specified probabilities of assigning a given label.

This defines a family of distributions $A(D)$.

What methods can be used to study such distributions?

A specific instance

- Let D be an Erdős-Rényi digraph with $\pi(n) = \frac{c}{n}$ for some $c > 1$.
- Generate labels by drawing an initial state $\vec{s}(0)$.
- For every $\varepsilon > 0$ there exist $N = N(\varepsilon, c)$ and $\delta = \delta(\varepsilon, c)$ as well as an algorithm A as on the previous slide so that with probability approaching 1 as $n \to \infty$:
  (a) The labeling of $A(D)$ will be the state $\vec{s}(N)$ and a coloring of $A(D)$.
  (b) A does not add new arcs, so that $A(D)$ is a subdigraph of D.
  (c) The mean degree of the subdigraph $B(D)$ of $A(D)$ induced by the giant strongly connected component of D is at least $1 + \delta$.
  (d) The proportion of nodes of $B(D)$ with indegree zero is at most ε.

Problem 12: Can we deduce that $B(D)$ must contain a directed cycle?

The answer would be trivially "yes" for the undirected case. $B(D)$ cannot contain a directed cycle of odd length by point (a).