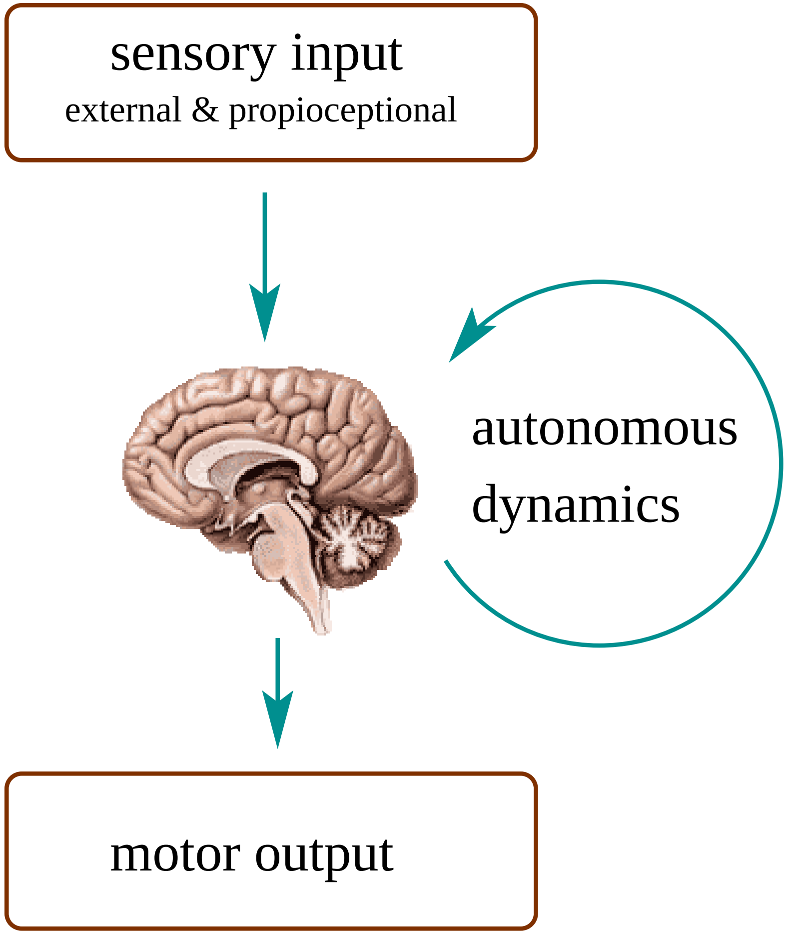
Self-sustained thought processes in a dense associative network

Claudius Gros

Institut für Theoretische Physik,
J.W.Goethe-Universität Frankfurt

brain dynamics



- isolated brain:
  associative thought processes?
- autonomous dynamics:
  self-sustained
  - ▷ epiphenomena?
    by-product of computational activities?
  - ▷ central functional role?

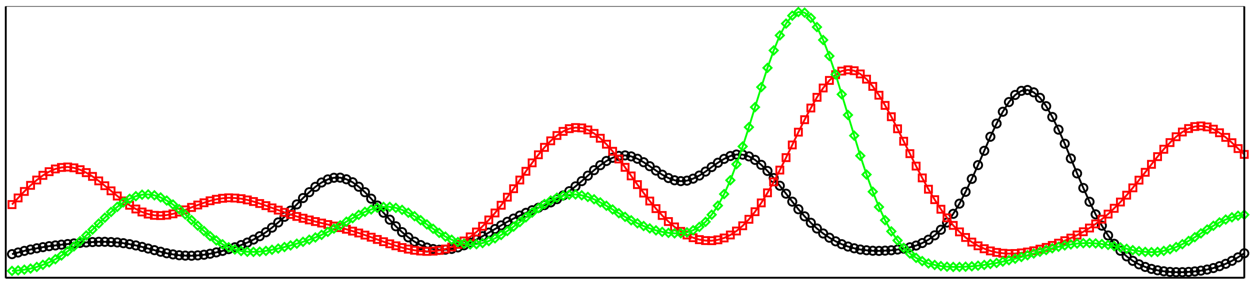
(i) Is the autonomous dynamics central to thinking?
(ii) Which principles govern the autonomous dynamics?

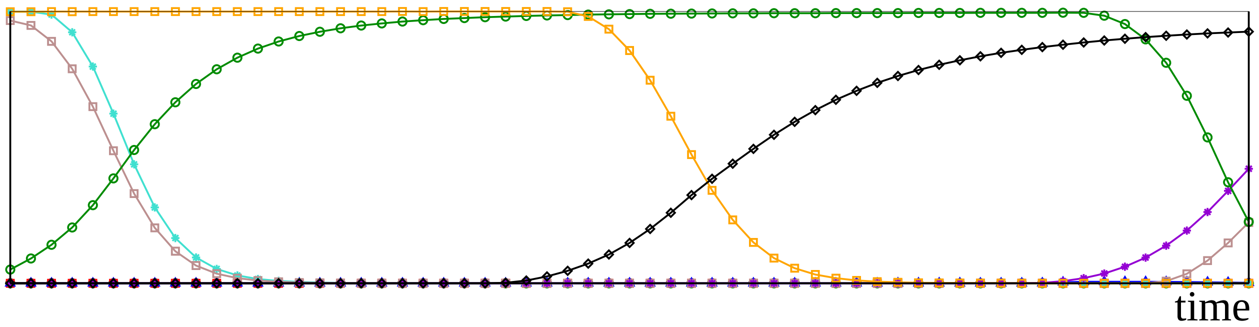
transient attractors

- how does the spontaneous activity look like?
  - ▷ random-like, no apparent meaning
  - ▷ do well defined transient states exist?
    corresponding to a
    well-defined ‘state-of-mind’?

‘critical reentrant events’ [Edelmann & Tononi]

- integrated activity of cell-assemblies





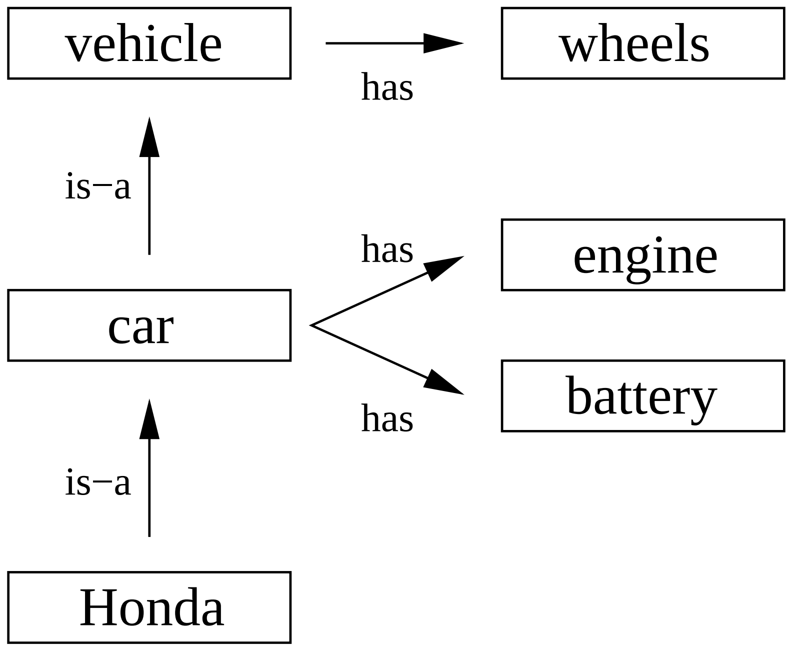
thought processes

- transient ‘states-of-the-mind’
  correspond in the isolated brain to
  memory-states (previously stored/learned)

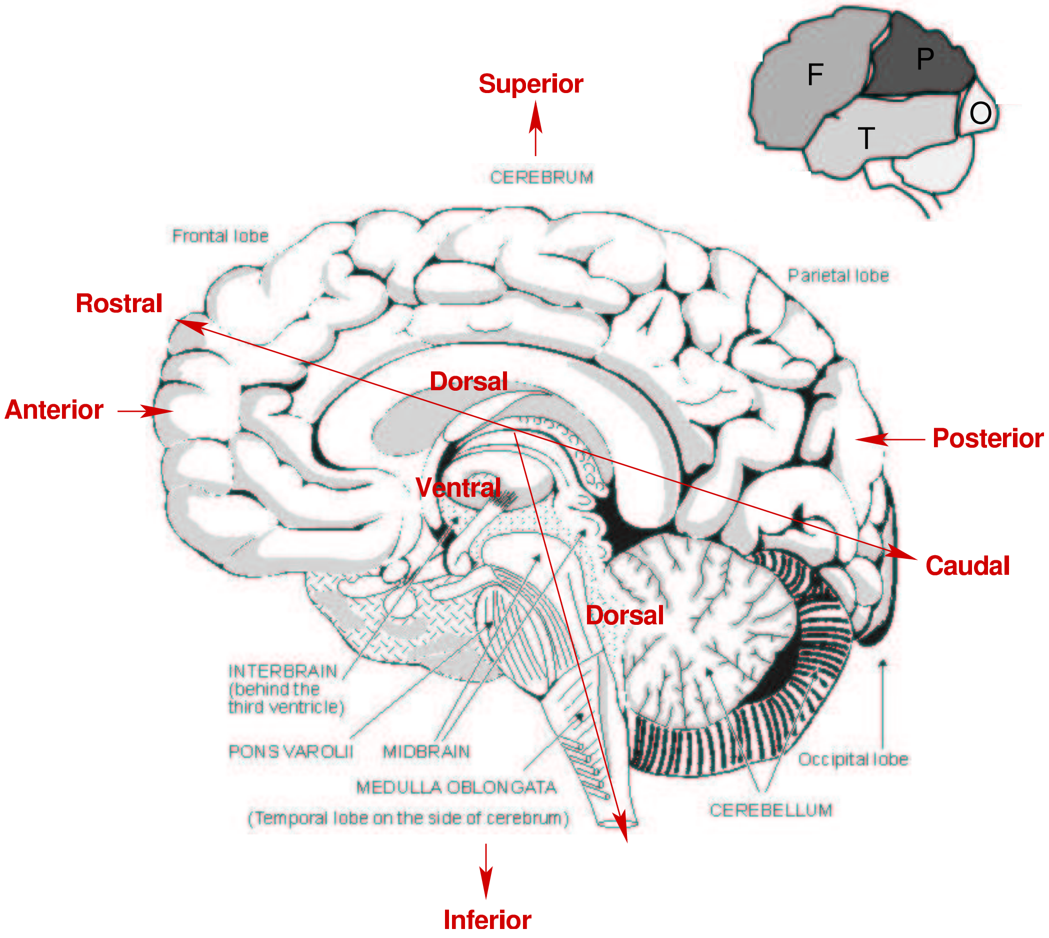
- thought processes
  correspond to a time-sequence of memory-states:
  >> associatively connected <<

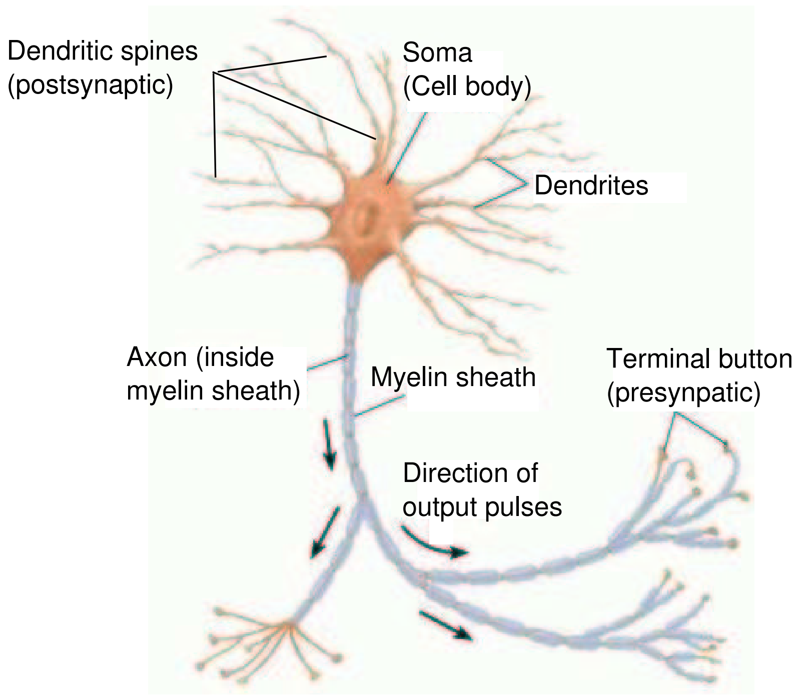
- nature of network:
  two building blocks?
  ▷ stored memories
  ▷ associative links

- nature of memory states:
  isolated or overlapping?
  ▷ one building block only



brain and neurons



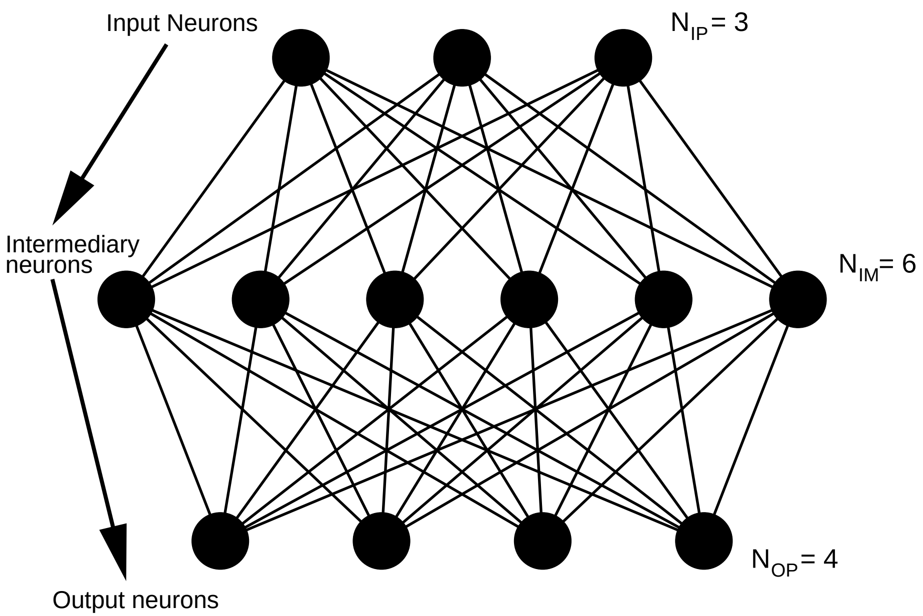


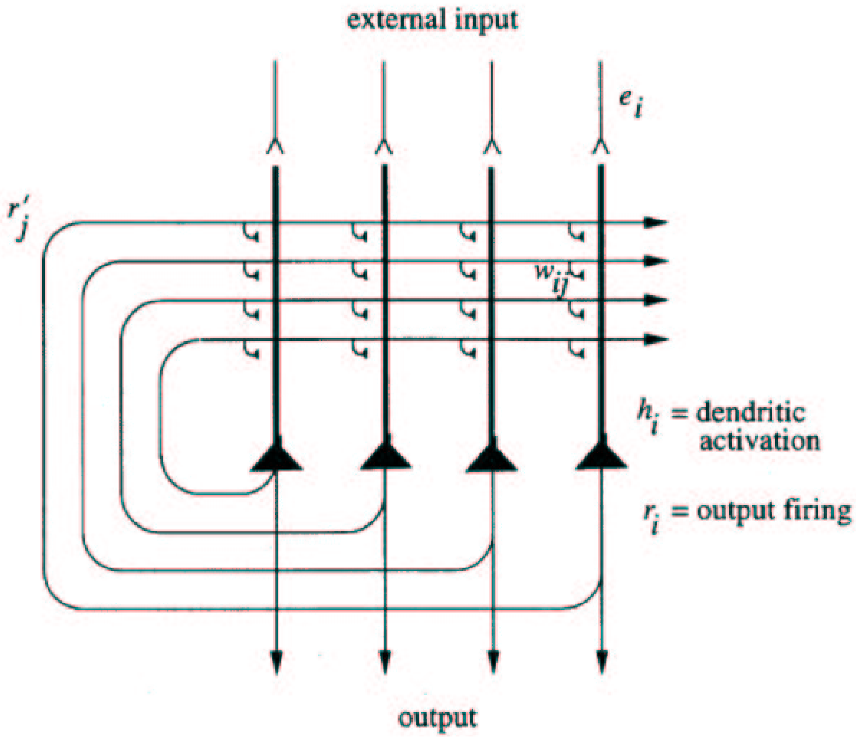
neural networks

- activity of biological neurons: variable $x_i, \; i = 1 \ldots N$ (axon potential)
  - ▷ discrete $x_i = 0, 1$ or $x_i = \pm 1$ (single spikes)
  - ▷ continuous $x_i \in [0, 1]$ (spike-density)
- inter-neural connection: synapses $w_{i,j}$
  - ▷ excitatory $w_{i,j} > 0$
  - ▷ inhibitory $w_{i,j} < 0$
- relation to statistical mechanics $H = \sum_{i,j=1}^{N} x_i w_{i,j} x_j$
- learning by Hebb-rule (long-term potentiation) $\delta w_{i,j} \approx x_i x_j$

inter-neural connections

- feed-forward (layered)
- recurrent (auto-associative)





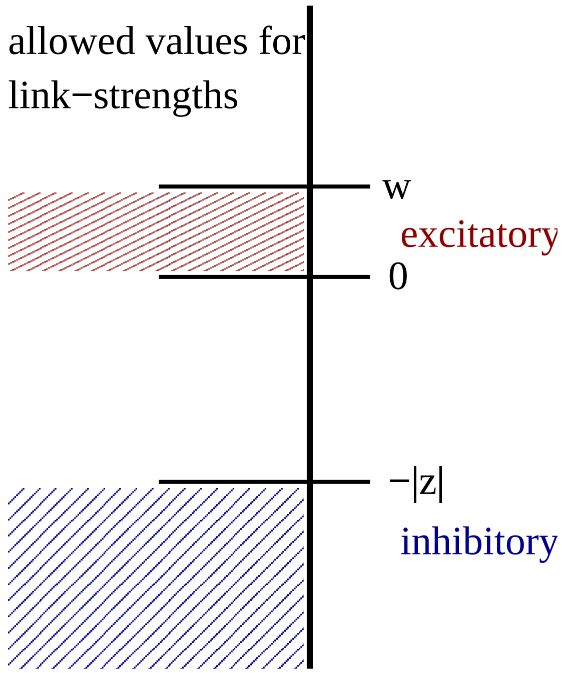
- standard task: input-output mapping

model - short-term relaxational dynamics

- continuous dynamics $x_i = x_i(t) \in [0,1]$
- reservoir $\varphi_i = \varphi_i(t) \in [0,1]$

$$\begin{aligned} \dot{x}_i &= (1-x_i)\Theta(r_i)r_i + x_i[1-\Theta(r_i)]r_i \\ r_i &= g(\varphi_i)\sum_{j=1}^N w_{i,j}x_j + \sum_{j=1}^N z_{i,j}x_j f(\varphi_j)\ , \end{aligned}$$

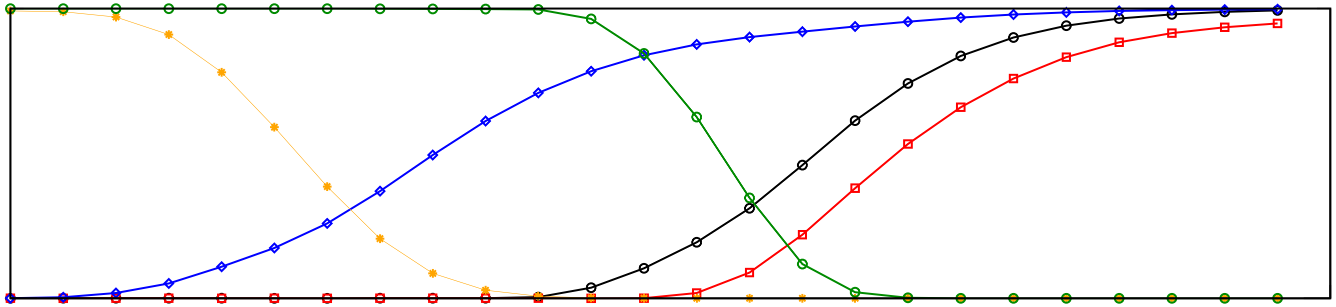
- two sets of connections $w_{i,j}z_{i,j} = 0$
  - ▷ excitatory $w_{i,j} \geq 0$
  - ▷ inhibitory $z_{i,j} \leq 0$
- growth-rates r_i
  - ▷ attractor network for $g(\varphi) = f(\varphi) \equiv 1$
  - ▷ relaxational dynamics



activity centers (AC) and memory states

The x_i are activity centers - the output of primary or secondary input processing units.
Memory states correspond to the simultaneous and stable activation of a finite number of activity centers.

- ‘winners-take-all’ network
  - ▷ winners (active): $x_i(t) \longrightarrow 1$
  - ▷ losers (inactive): $x_i(t) \longrightarrow 0$

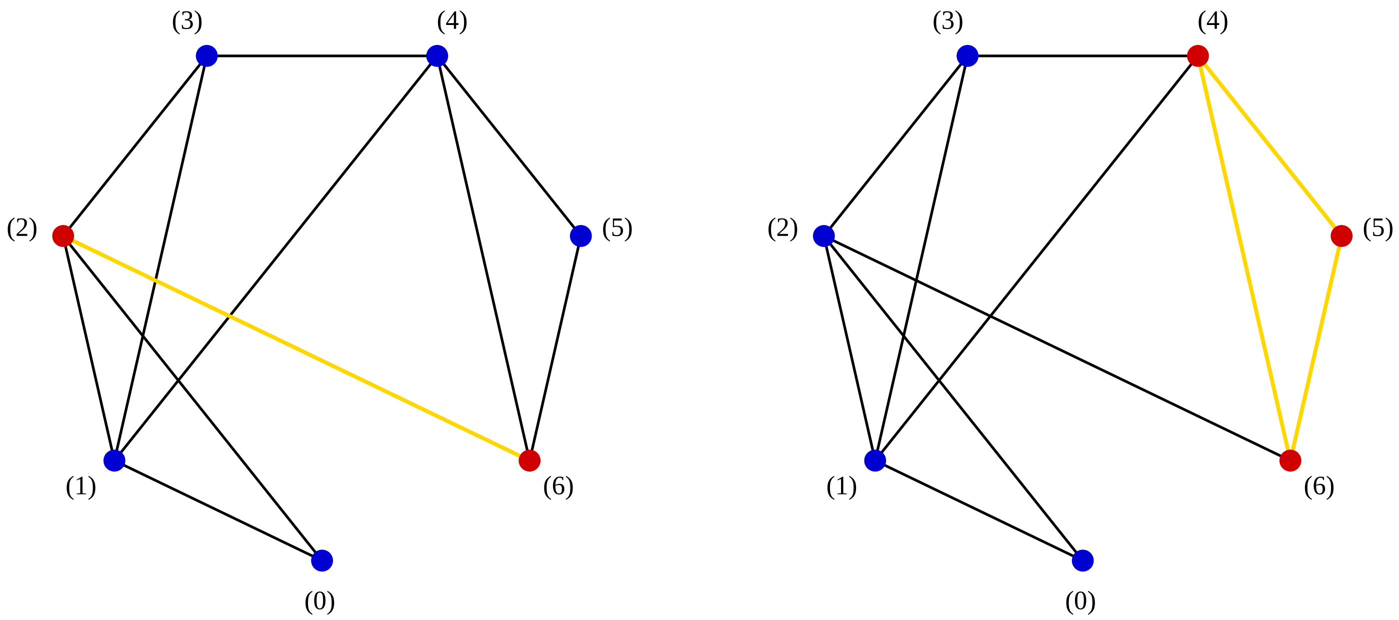


  - ▷ decision-making capabilities

stable memory states

- illustration: lines are $w_{i,j} > 0$
- stable memory states:

$$(0,1,2), \quad (1,2,3), \quad (1,3,4), \quad (4,5,6), \quad (2,6).$$



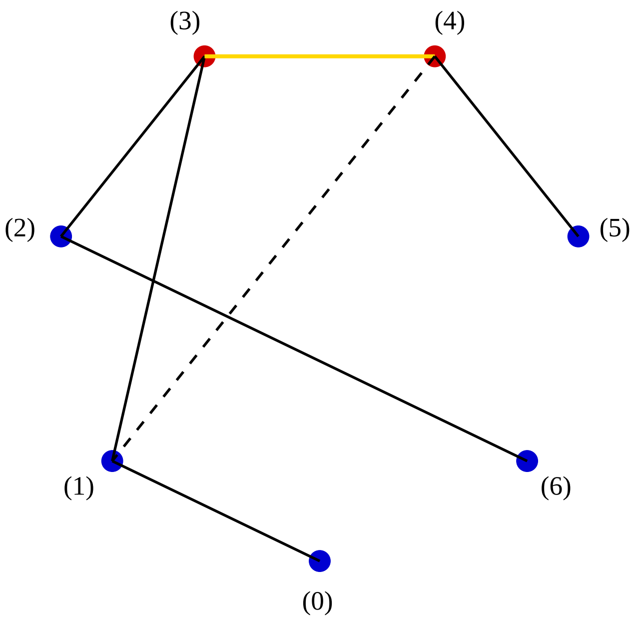
memory states $\Leftrightarrow$ fully connected clusters

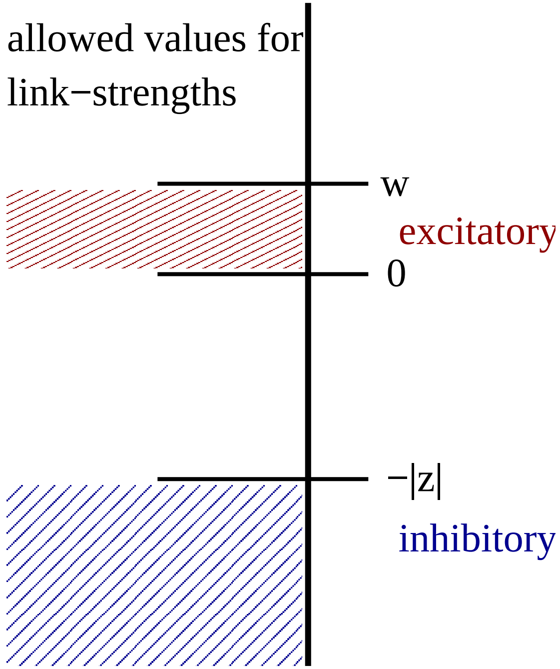
stabilization of memory states

- homogeneous case:

  $w_{i,j} = w > 0, \quad z_{i,j} = z = -|z|$

  ▷ condition: $|z| > (Z-1)w$



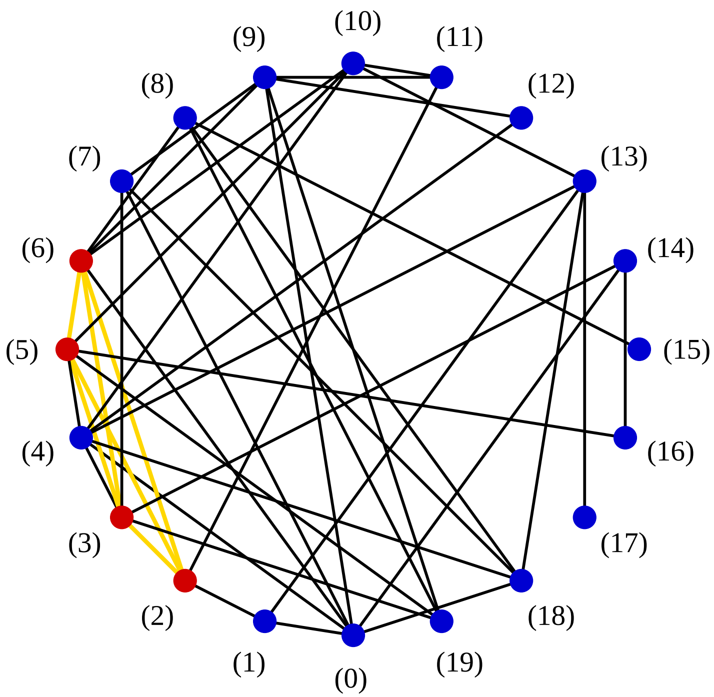


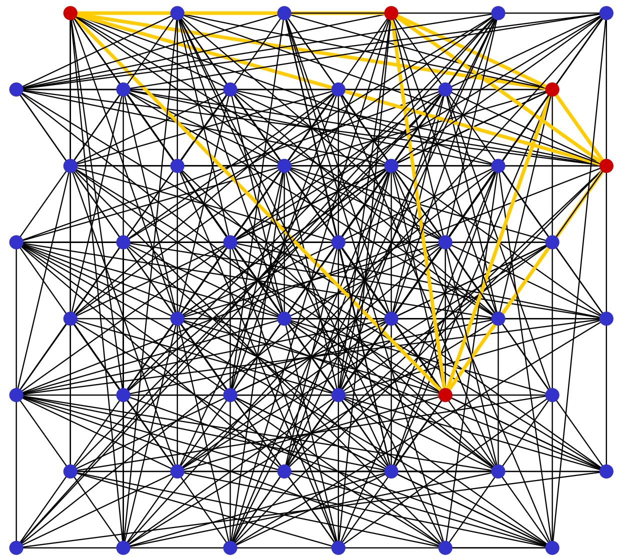
dense and homogeneous networks

memory states and associations between memories are identical !

stable memory states:

	2-AC	3-AC	4-AC	5-AC	total
20-sites	19	10	1	0	30
48-sites	2	166	66	2	236



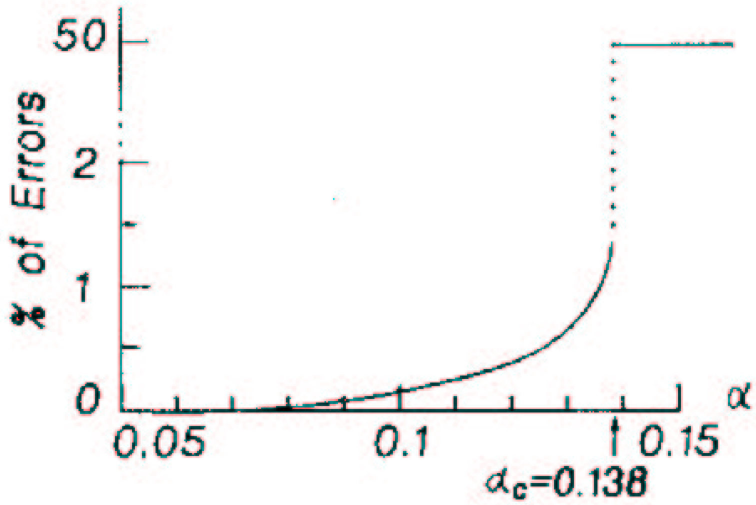


storage capacity - standard implementation

- ‘normal’ neural networks: average activity $\overline{x_i(t)} \approx 0.5$

- maximal number of storable patterns N_p extensive

$$N_p = \alpha N, \qquad \alpha \approx 0.138$$



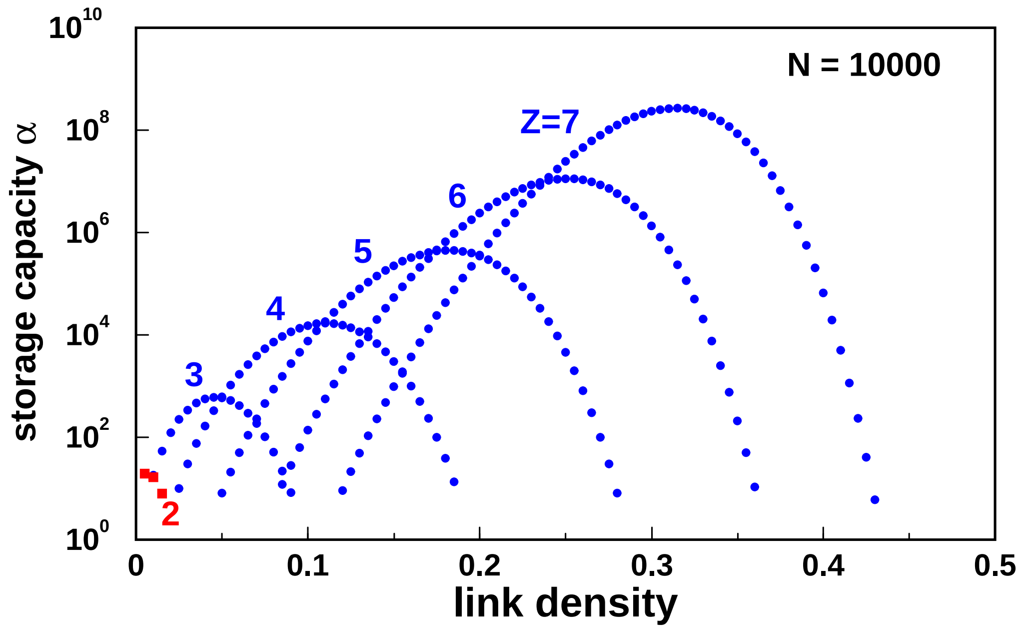
- ‘sparse’ neural network: average activity $\overline{x_i(t)} \ll 1$

- statistical storage capacity: $N_p = \alpha N, \qquad \alpha \sim N$

finite-size memory states

- here: finite-number Z of active neurons
  - ▷ $\rho_l = 2N_l/N(N-1)$: density of excitatory links N_l
  - ▷ spurious state:
    Z - excitatory links to one other site
  - ▷ N_Z : number of stable Z-center MS

$$N_Z = \binom{N}{Z} \rho_l^{\binom{N}{Z}} \left(1-\rho_l^Z\right)^{N-Z}$$



- ▷ gigantic storage capacity

autonomous long-term dynamics

$$\begin{aligned} \dot{x}_i &= (1-x_i)\Theta(r_i)r_i + x_i[1-\Theta(r_i)]r_i \\ r_i &= g(\varphi_i)\sum_{j=1}^{N} w_{i,j}x_j + \sum_{j=1}^{N} z_{i,j}x_j f(\varphi_j) \end{aligned}$$

- reservoir dynamics: slow
  - ▷ deletion-rate: $\Gamma_\varphi^- \sim 10^{-2}..10^{-1}$
  - ▷ filling-rate: $\Gamma_\varphi^+ \sim 10^{-2}..10^{-1}$

$$\dot{\varphi}_i = \Gamma_\varphi^+ \Theta(x_c - x_i)(1-\varphi_i) - \Gamma_\varphi^- \Theta(x_i - x_c)\varphi_i$$

- coupling via reservoir functions $f(\varphi)$, $g(\varphi)$

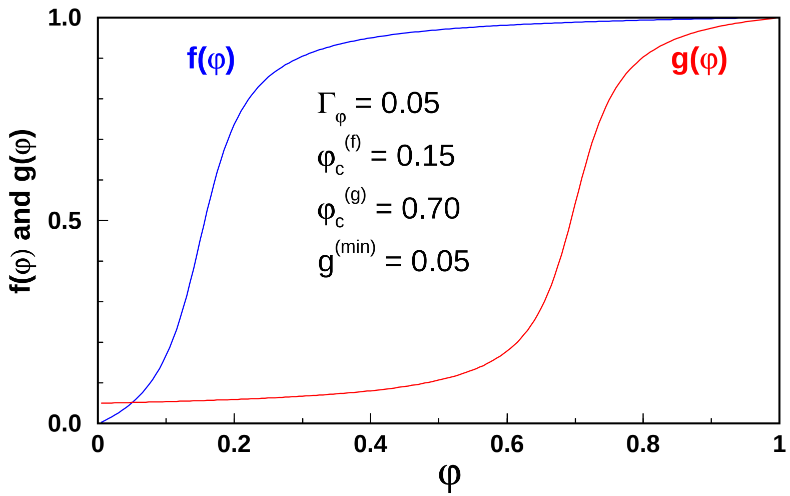
reservoir functions

$$
\begin{aligned}
\dot{x}_i &= (1-x_i)\Theta(r_i)r_i + x_i[1-\Theta(r_i)]r_i \\
r_i &= g(\varphi_i)\sum_{j=1}^{N} w_{i,j}x_j + \sum_{j=1}^{N} z_{i,j}x_j f(\varphi_j)
\end{aligned}
$$

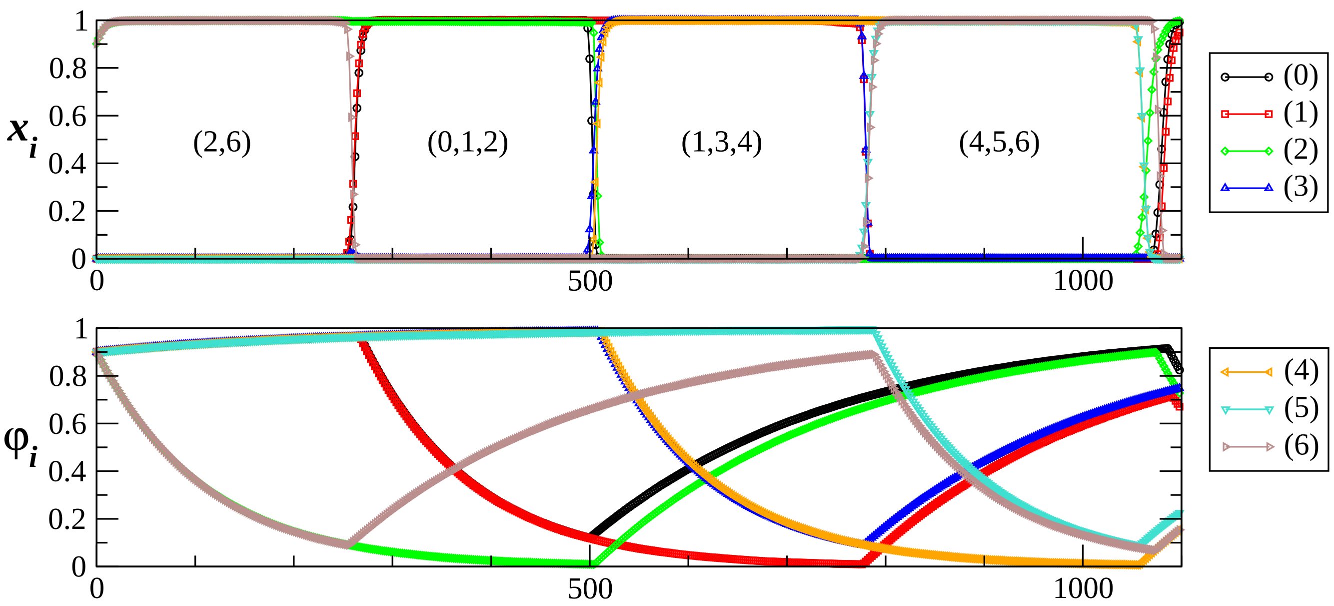
▷ neuron can inhibit other neurons if $\varphi > \varphi_c^{(f)} = 0.15$

▷ neuron can be activated by other neurons if $\varphi > \varphi_c^{(g)} = 0.8$

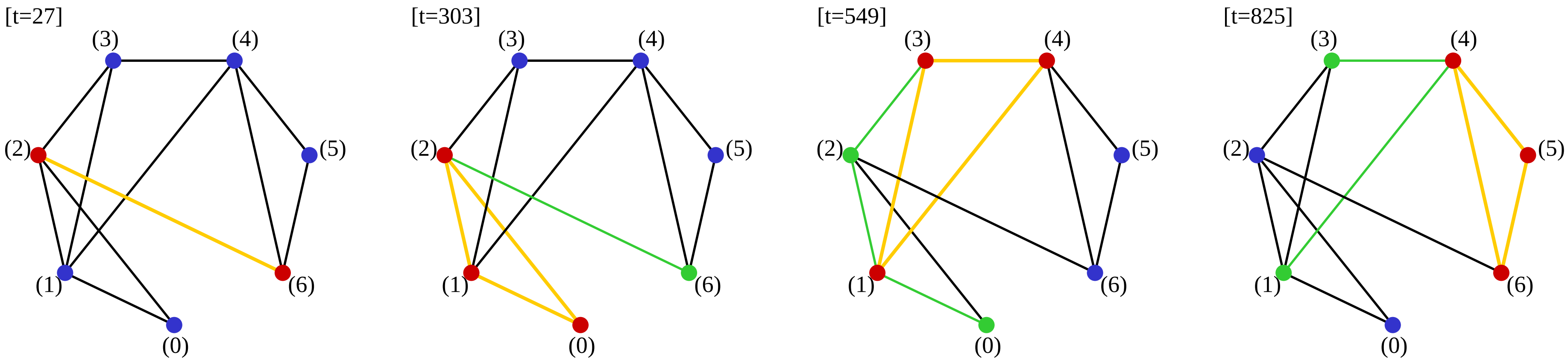
$$
f(\varphi) = \Theta(\varphi - \varphi_c^{(f)}) \qquad g(\varphi) = \Theta(\varphi - \varphi_c^{(g)})
$$



a simple thought process



The thought process $(2,6) \to (0,1,2) \to (1,3,4) \to (4,5,6)$

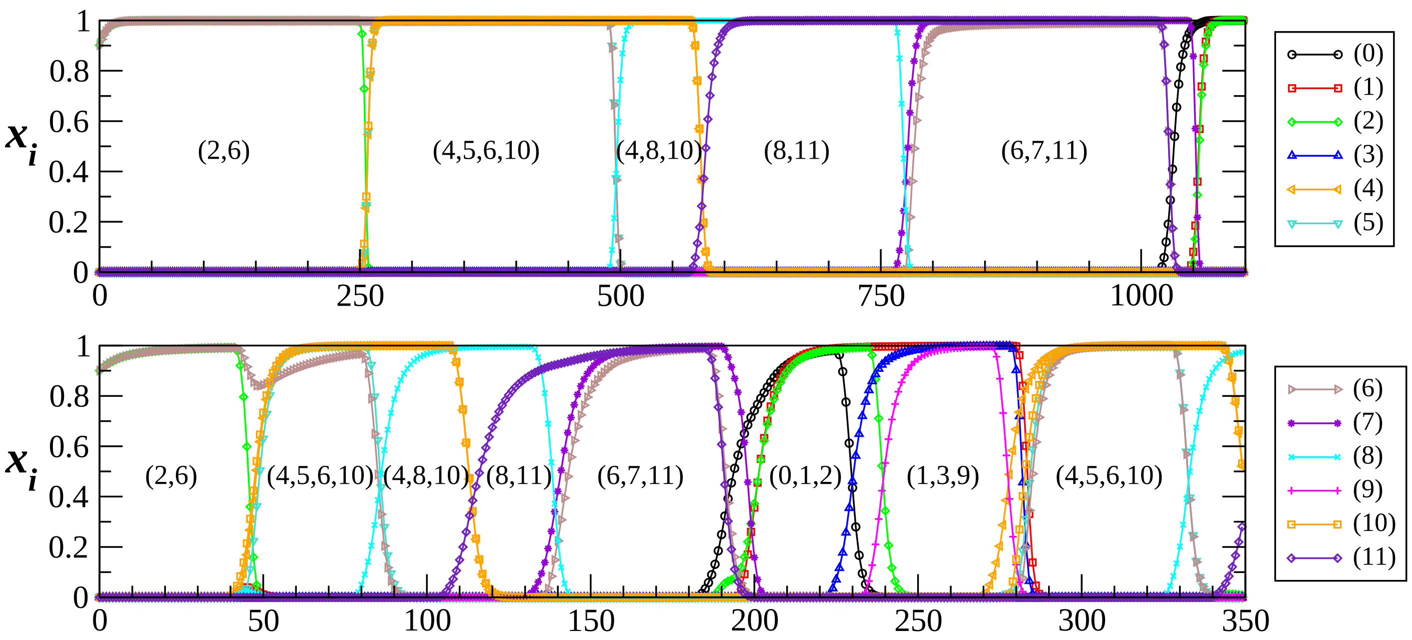


transient attractors

- parameter sets

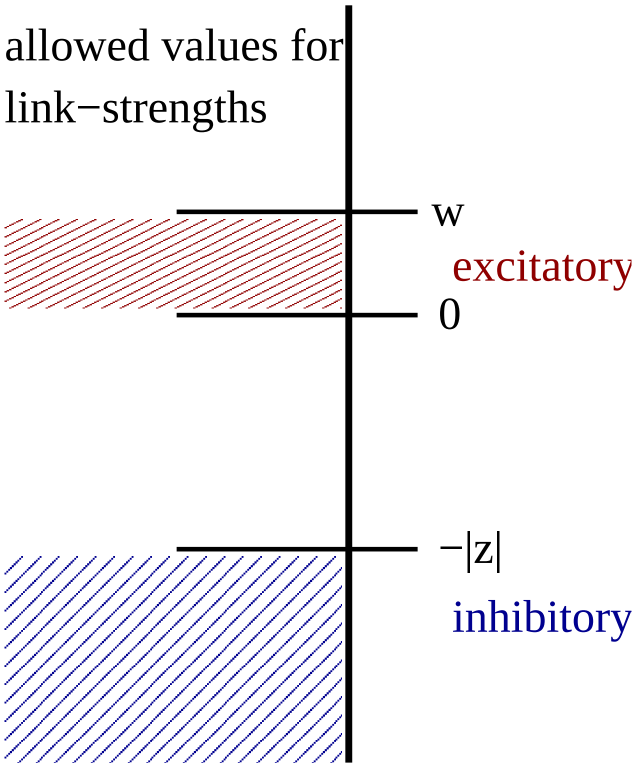
w	z	x_c	Γ_φ^+	Γ_φ^-	$\varphi_c^{(f)}$	$\varphi_c^{(g)}$	Γ_φ	$g^{(min)}$
0.15	-1.0	0.85	0.004	0.009	0.15	0.7	0.05	0.00
0.15	-1.0	0.50	0.005	0.020	0.15	0.7	1.00	0.10

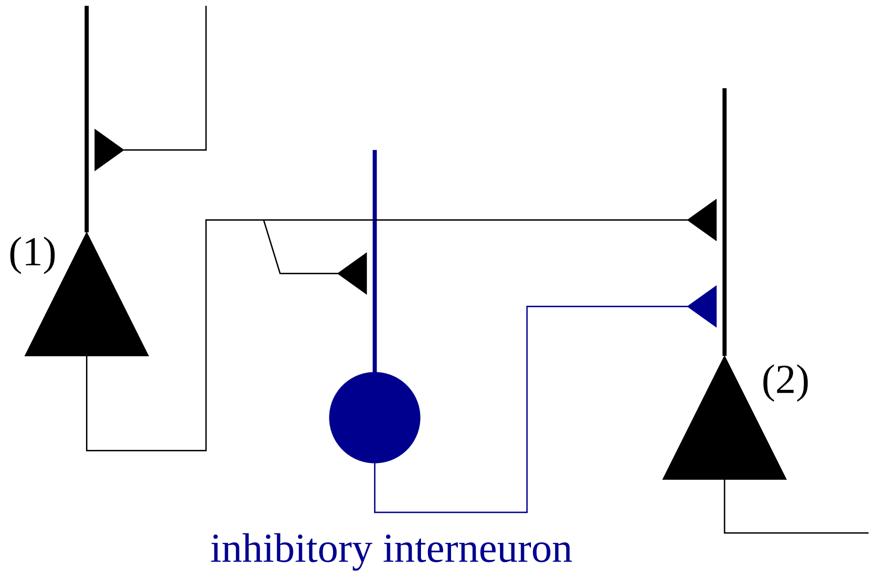
- full control over transient dynamics



effective neurons

- discontinuous link matrices $w_{i,j} z_{i,j} \equiv 0$
  - ▷ excitatory neurons: $w_{i,j} \geq 0$
  - ▷ inhibitory interneurons: $z_{i,j} \leq 0$





> integrating-out the interneurons may lead to effective neurons with discontinuous synapsing strength

biological considerations

- sparsification (eye,CA3): underlying principle for data compression
  - ▷ local representation of data ('grandmother cells')
  - ▷ here: finite number of active centers

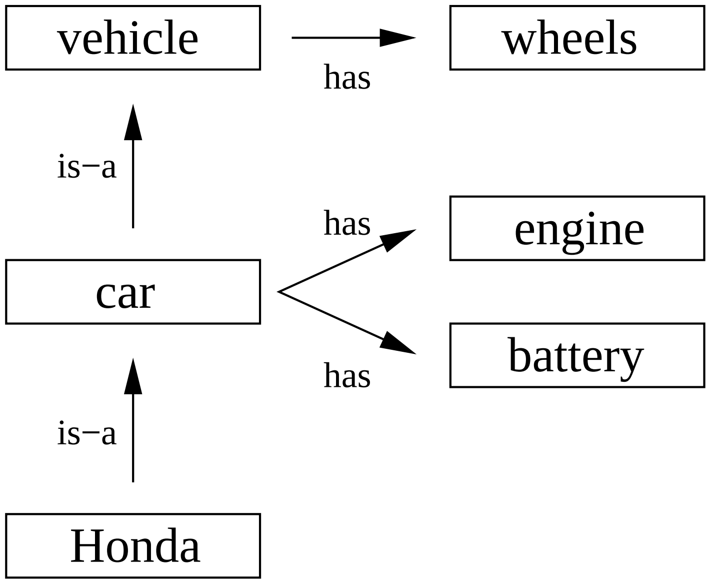
- synchronized firing of winning coalition

  [von der Malsburg, Singer]
  - ▷ only possible with spiking neurons, but
  - ▷ possible interpretation within effective model building

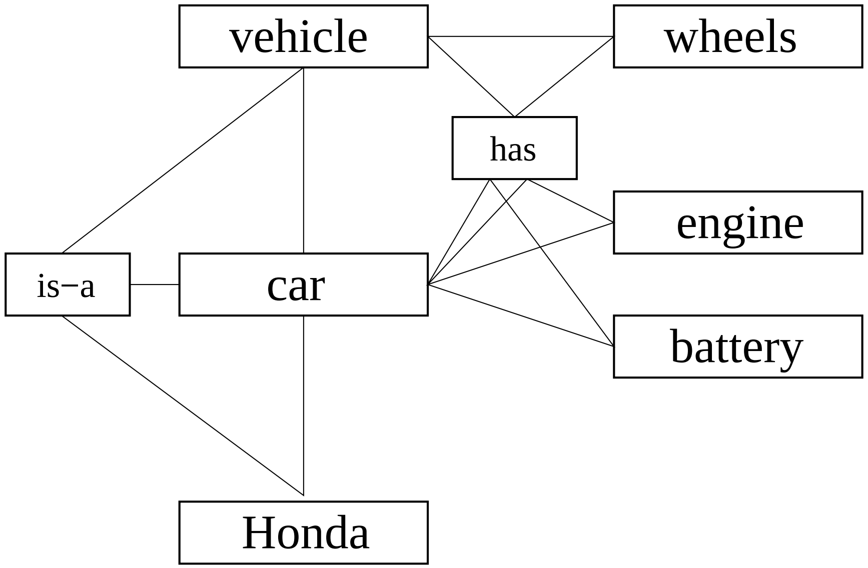
> The reservoirs $\varphi_i(t)$ correspond to the synchronization rate of center (i) relative to the synchronized firing of winning cluster.
> $\Rightarrow$ decay rate $\Gamma_\varphi^- \,\hat{=}$ dephasing rate

semantic vs. associative network

semantic network



associative network



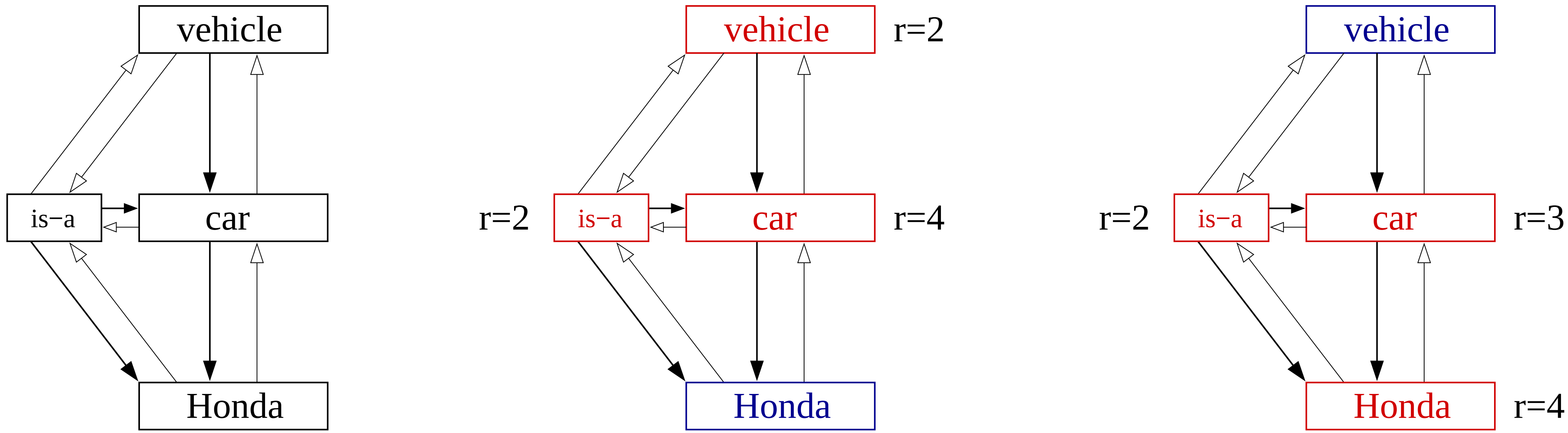
- internal (Brain) data representation
  in form of semantic networks?
- corresponding associative networks need
  hierarchical object representation

hierarchical object representation

- strength of excitatory links: weak / **strong**
  - ▷ needs gap in synapsing strength:
    small negative links not allowed!
- assymmetric: $w_{i,j} \neq w_{j,i}$

stable memory states

assymmetric network | car is a vehicle | Honda is a car



hierarchy generation via distinct growth rates r_i

the wordNet lexikon

[http://wordnet.princeton.edu]

Hypernym

Y is a hypernym of X if X is a (kind of) Y

Sense 1

uncle – (the brother of your father or mother; the husband of your aunt)

=> kinsman – (a male relative)

Sense 2

uncle – (a source of help and advice and encouragement; "he played uncle to lonely students")

=> benefactor, helper – (a person who helps people or institutions (especially with financial help)

- (small world) semantic network

[Sigman & Cecchi, PNAS 2002]

  - 79689 nouns
  - 16 semantic cagetories

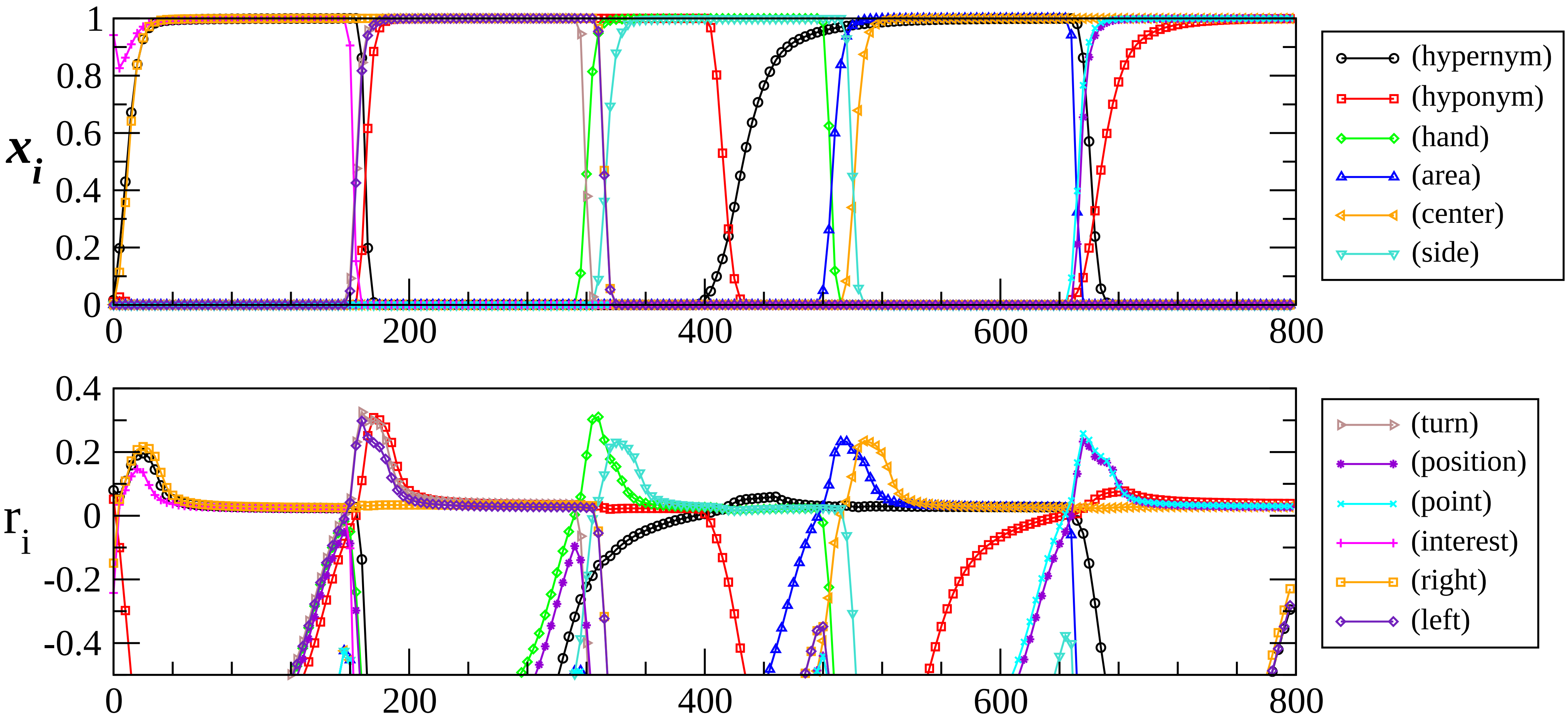
semantic categories within wordNet

Antonym	contrary meaning	1
Hypernym	this is a kind of...	1
Hyponym	...is a kind of this	1
Holonym	this is a part of...	3
Meronym	parts of this	(member,part,substance)
Domain	...is within the domain	3
Domain term	is the domain of...	(category, region, usage)

- implementation as an associative network
  - ▷ 3435/414 most common nouns
  - ▷ 10943/1116 definitions result in stable clusters:

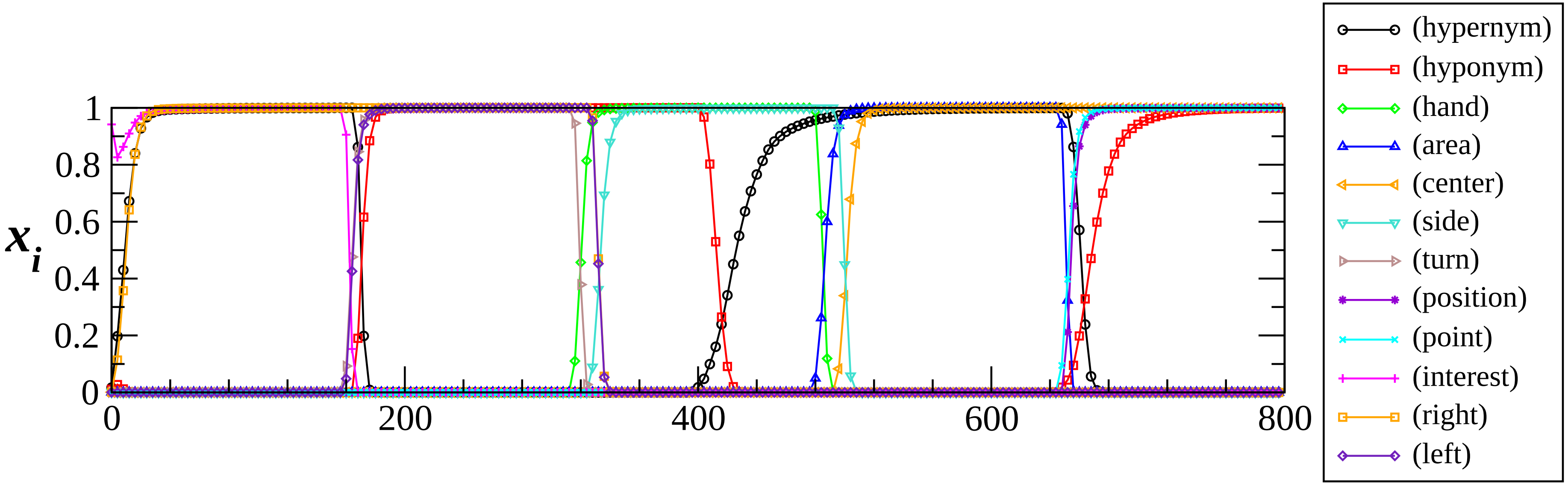
Z	3	4	5	6	N
N_Z	11543	1135	40	2	3447
	1129	104	7	0	423

a N=423 WordNet autonomous process



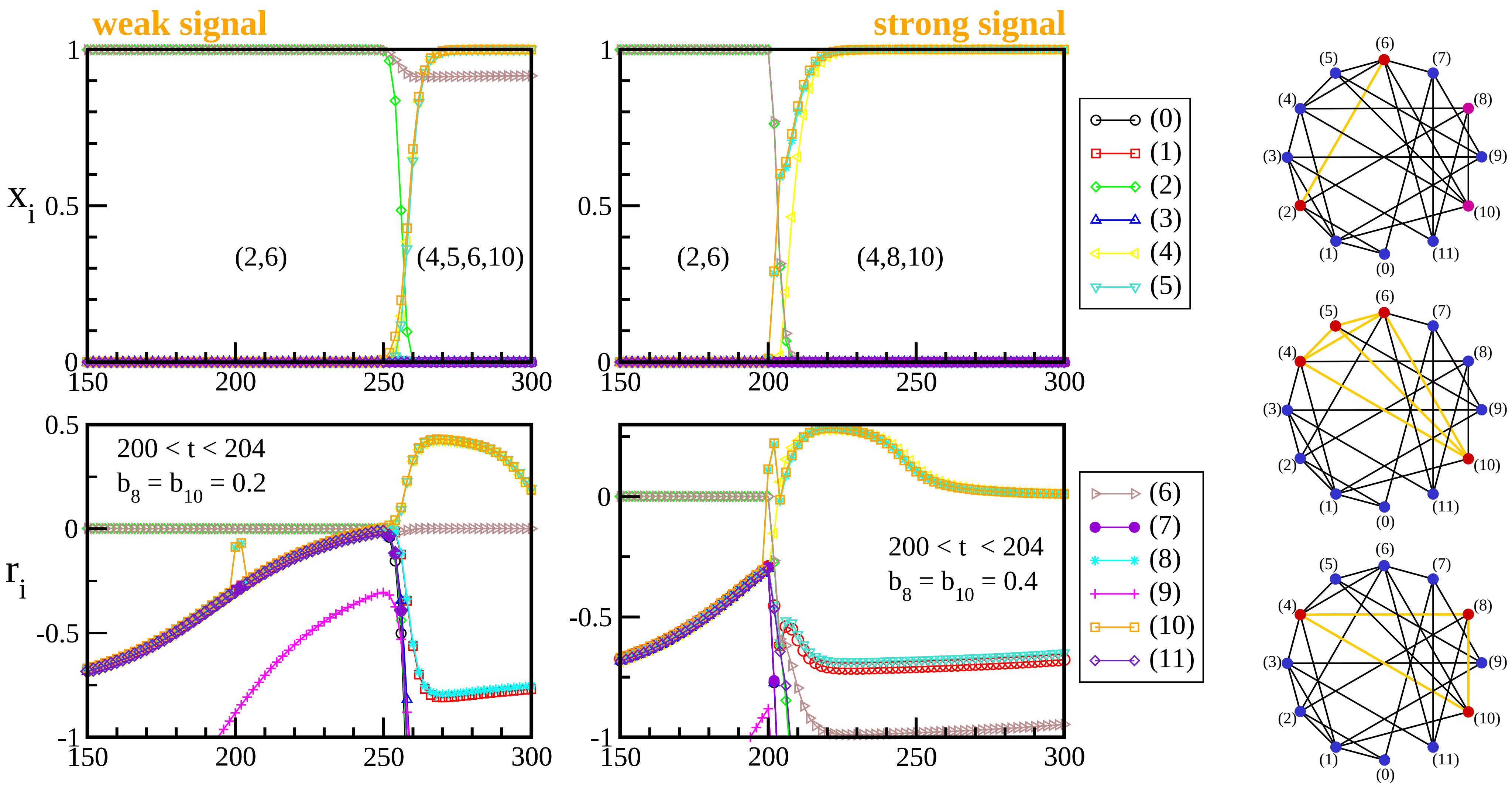
(only active centers)

a N=423 WordNet thought process



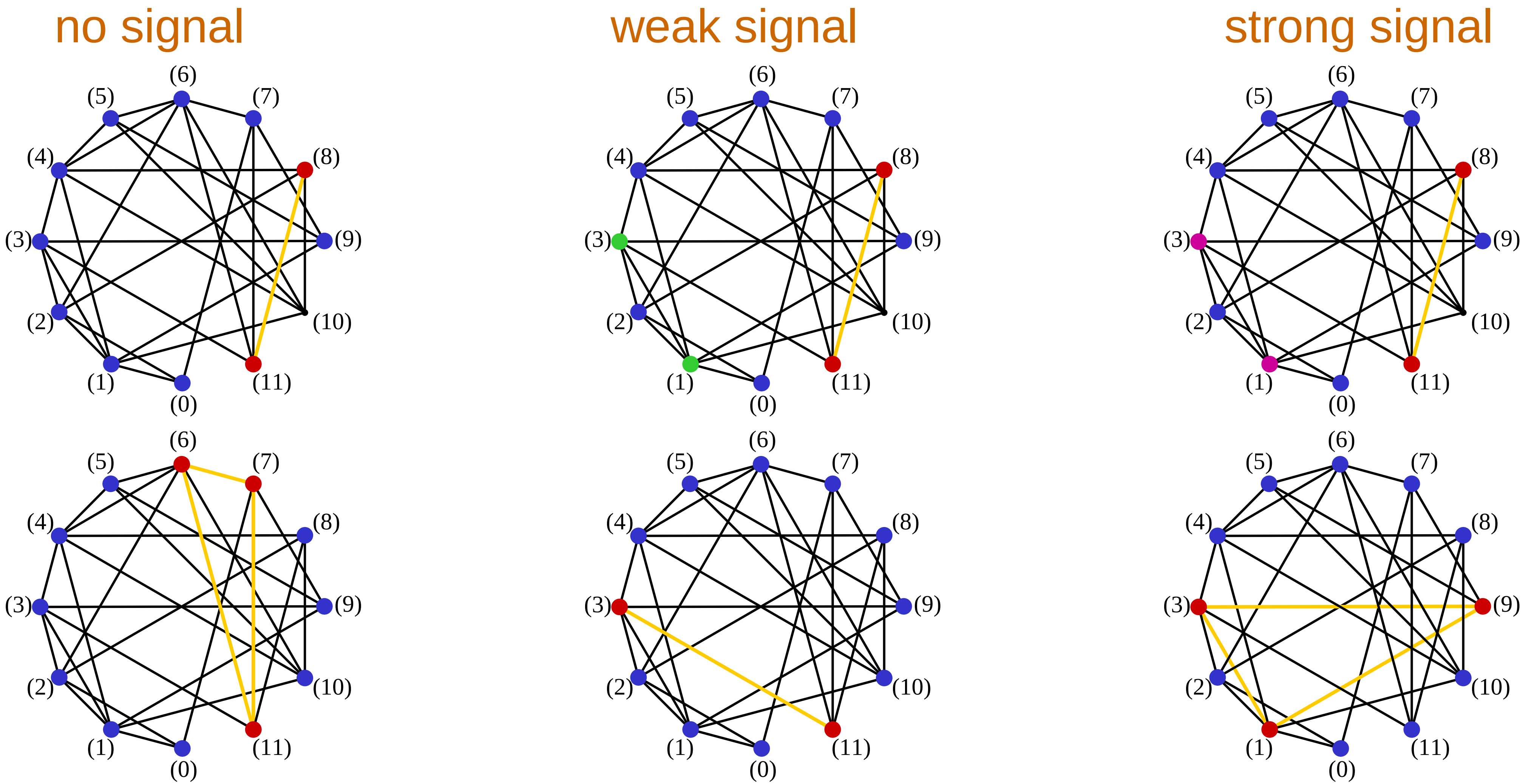
cluster	time	comment
(right,hypernym,interest)	30	hypernym: this is a kind of...
(turn,hyponym,right,left)	190	(turn,hyponym,right), (turn,hyponym,left)
(side,hyponym,hand)	360	(left,hypernym,hand)
(hand,hypernym,side)	450	
(center,hypernym,area)	520	(side,hypernym,area)
(center,position,point)	620	new cluster

response to stimuli



- discrimination between weak and strong preprocessed signals – recognition

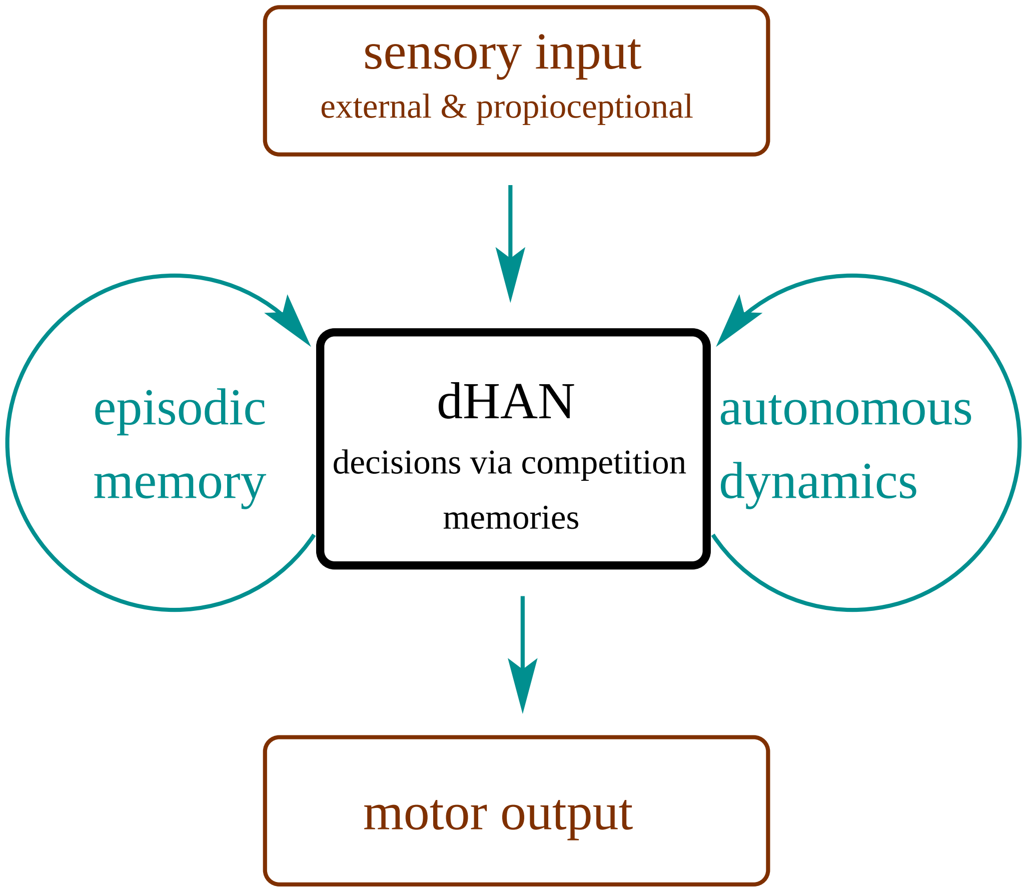
inspiration vs. recognition



effect of dynamical attention-field

decisions by competition

- rudimentary cognitive system
  - ▷ single task at one time
  - ▷ winners-take-all core allows for decision-taking



- missing components
  - ▷ diffusive control
  - ▷ specialized moduls for cognitive tasks?
  - ▷ ‘unconscious’ secondary processes
  - ▷ higher-level hierachies

principles

- There exists self-induced brain activity and it is of central functional importance for higher cognitive processes.
- There exist well-defined transient states (attractors).
- The spontaneous activity would occur also in the isolated brain.
- The transient attractors correspond in the isolated brain to stored memory states.
- A thought process corresponds in the isolated brain to the self-induced generation of a time-sequence of associatively connected memory states.
- The associative network consists of overlapping memory states and is homogeneous - there don't exist distinct connecting units.
- Memory states have a dual functionality: They serve as transient attractors during a thought process and provide the necessary associative connections.