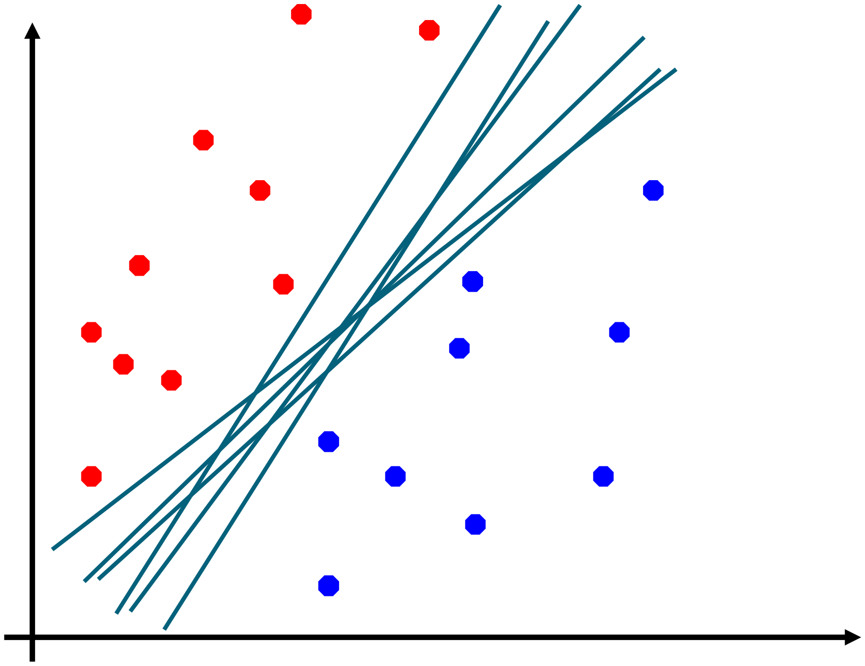
Reservoir Computing

吴思
心理与认知科学学院
IDG/McGovern 脑科学研究所
北大-清华联合生命科学中心
北京大学

The Kernel Trick of
Support Vector Machine

Which One is the Best?

$$y(\mathbf{x}) = sign(\mathbf{w} \cdot \mathbf{x} + b)$$



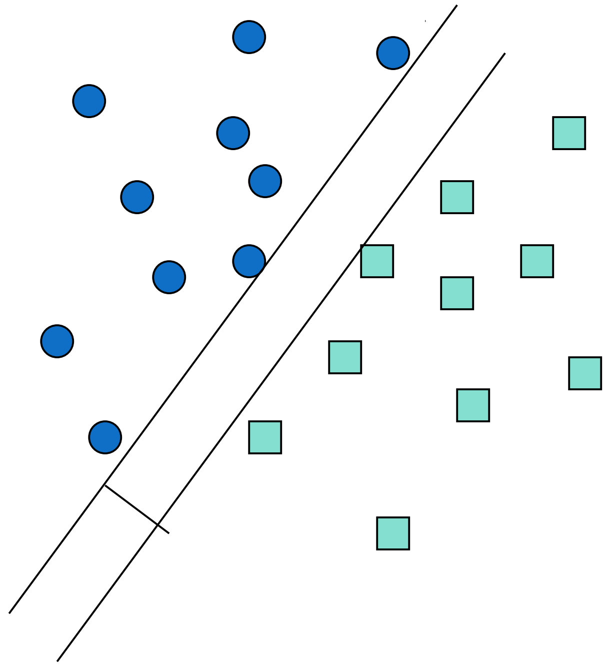
The max-margin classifier

- **The optimization problem:**

Find **w** and b such that

$$d = \frac{2}{\|\mathbf{w}\|}$$ is maximized; and for all $\{(\mathbf{x_i}, y_i)\}$

$\mathbf{w}\cdot\mathbf{x_i} + b \geq 1$ if $y_i = 1$; $\mathbf{w}\cdot\mathbf{x_i} + b \leq -1$ if $y_i = -1$

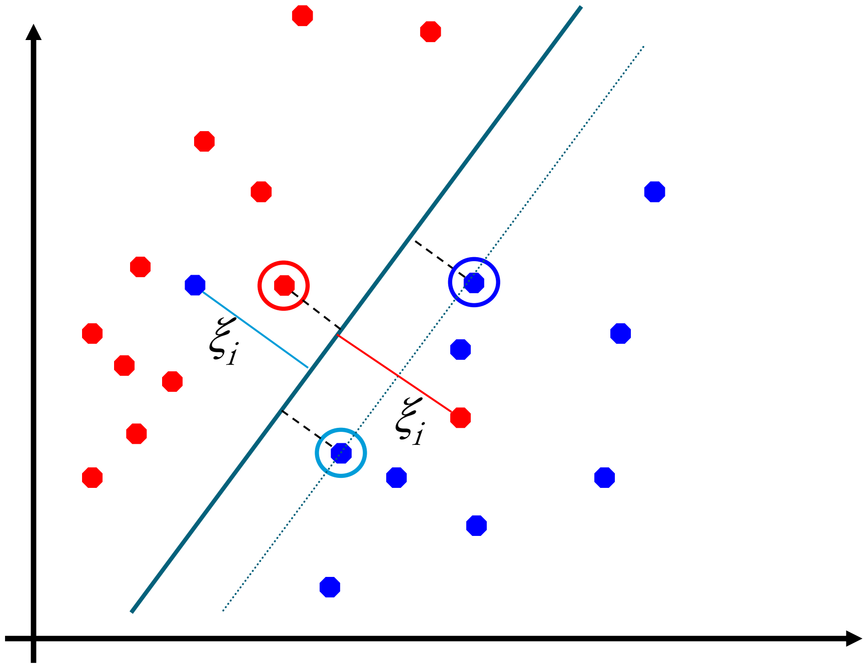


The Soft-Margin Classifier

Minimize $\quad \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_i \xi_i$

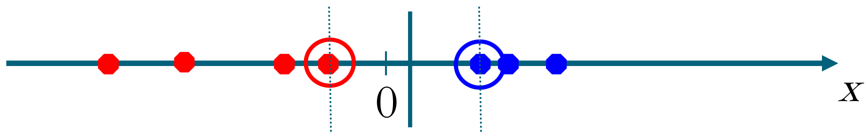
Subject to $\quad y_i(\mathbf{w}\cdot\mathbf{x}_i + b) \geq 1 - \xi_i$

$\qquad\qquad\quad \xi_i \geq 0$

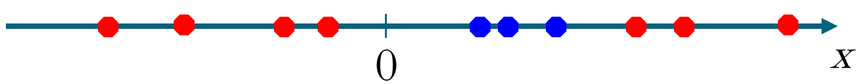


Who really need linear classifiers

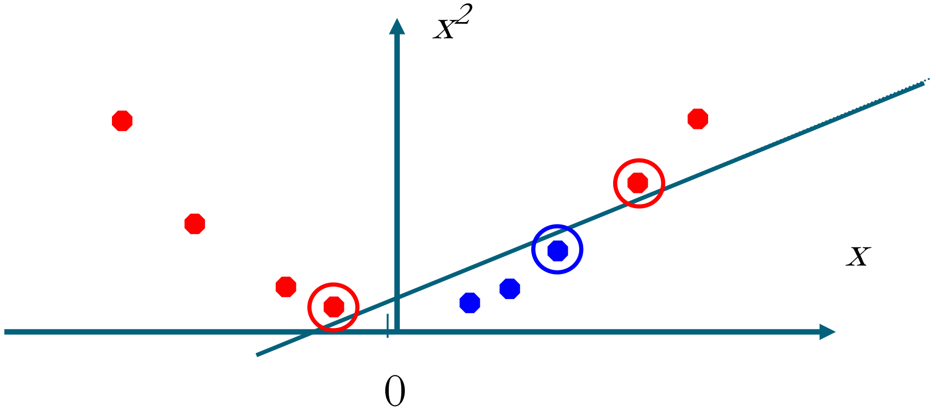
- **Datasets that are linearly separable with some noise, linear SVM work well:**



- **But if the dataset is non-linearly separable?**

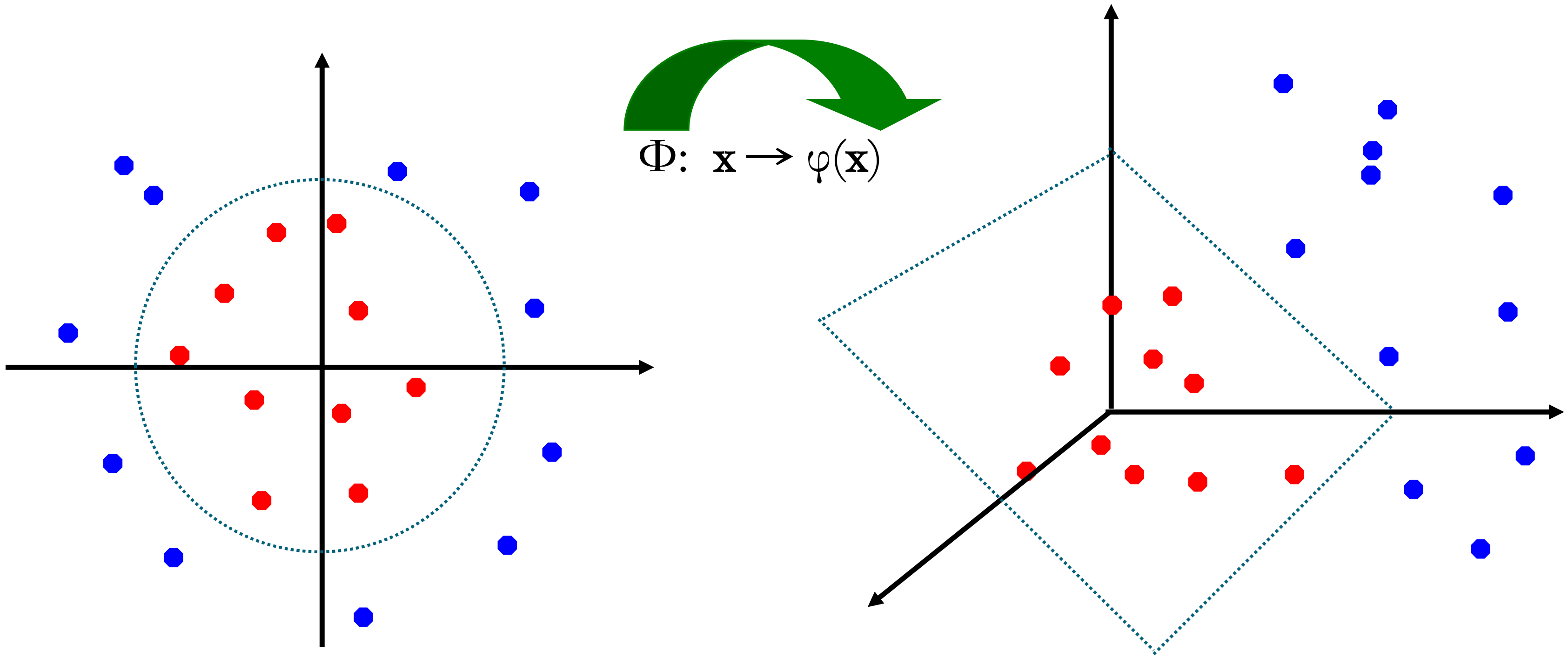


- **How about… mapping data to a higher-dimensional space:**



Non-linear SVMs: Kernel Mapping

- General idea: the original space is mapped to a higher-dimensional space where data becomes linearly separable:



The "Kernel Trick"

- The SVM only relies on the inner-product between vectors $x_i \cdot x_j$
- If every data point is mapped into high-dimensional space via some transformation Φ: $x \rightarrow \varphi(x)$, the inner-product becomes:

$$K(x_i, x_j) = \varphi(x_i) \cdot \varphi(x_j)$$

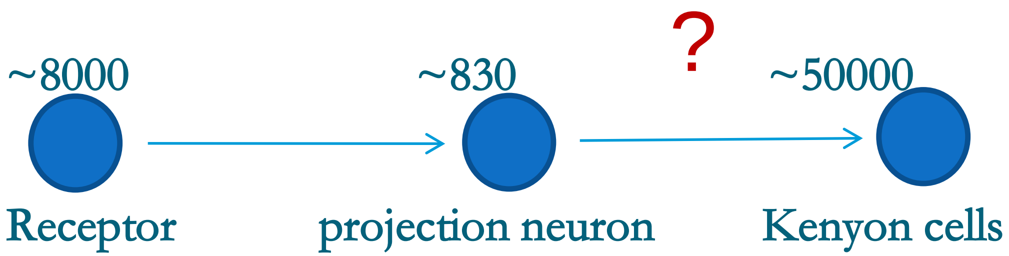
- $K(x_i, x_j)$ is called the kernel function.
- For SVM, we only need specify the kernel $K(x_i, x_j)$, without need to know the corresponding non-linear mapping, $\varphi(x)$.

Key ideas of SVM

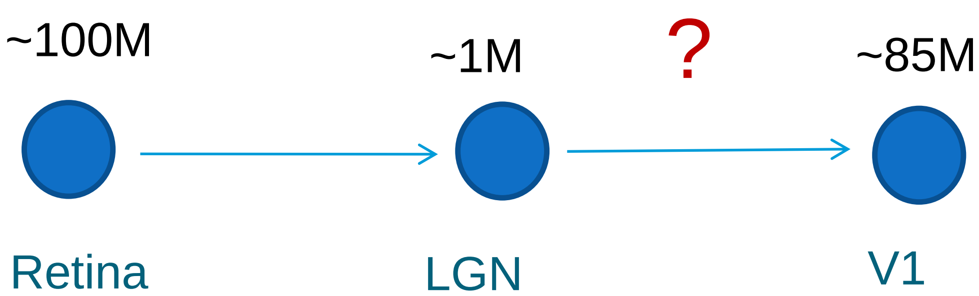
- A hyper-plane classifier
- A hyper-plane that maximizes the margin between two classes of data points
- Using a kernel function to map the original data into a high-dimensional space
- Soft-margin to accommodate noises

"Kernel Trick" in Neural Information Processing?

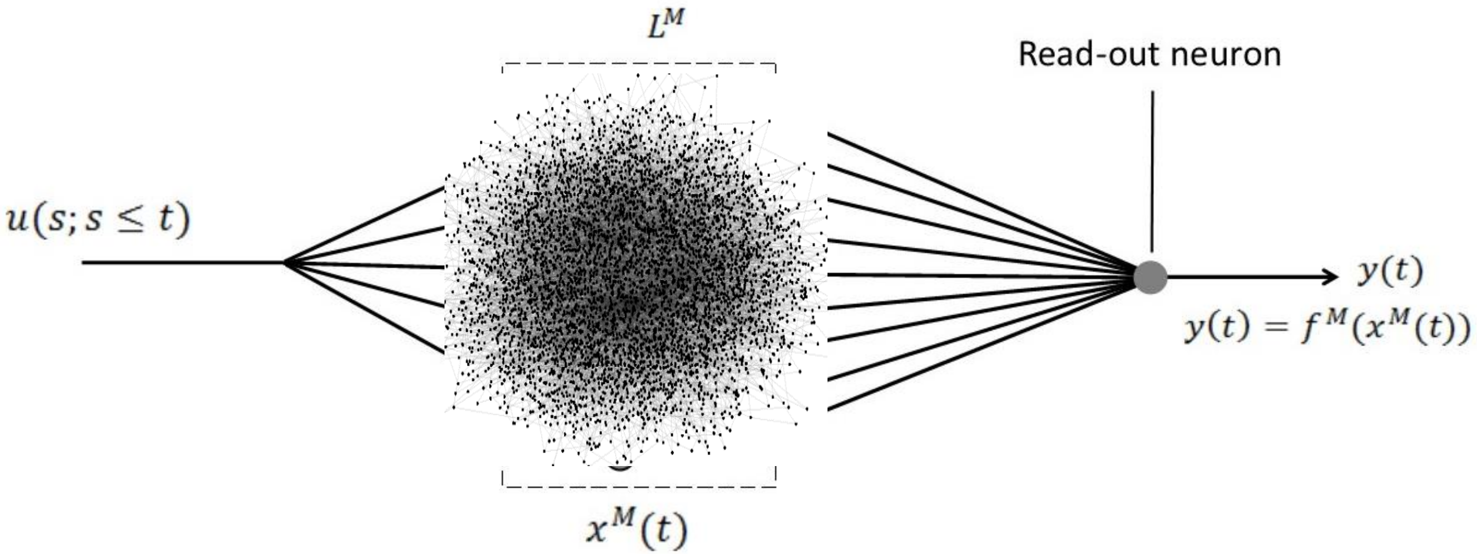
- Olfactory system of Locust



- Visual system of primates

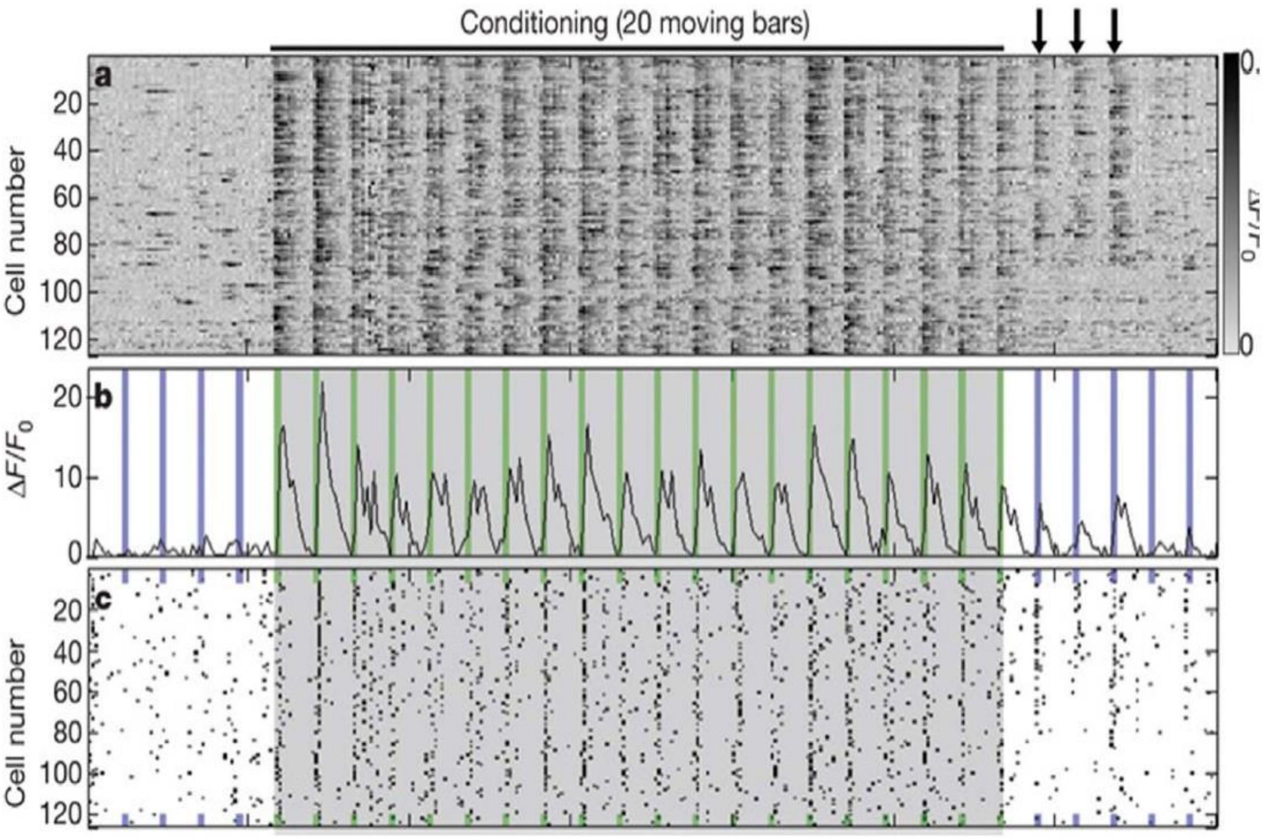


Reservoir networks



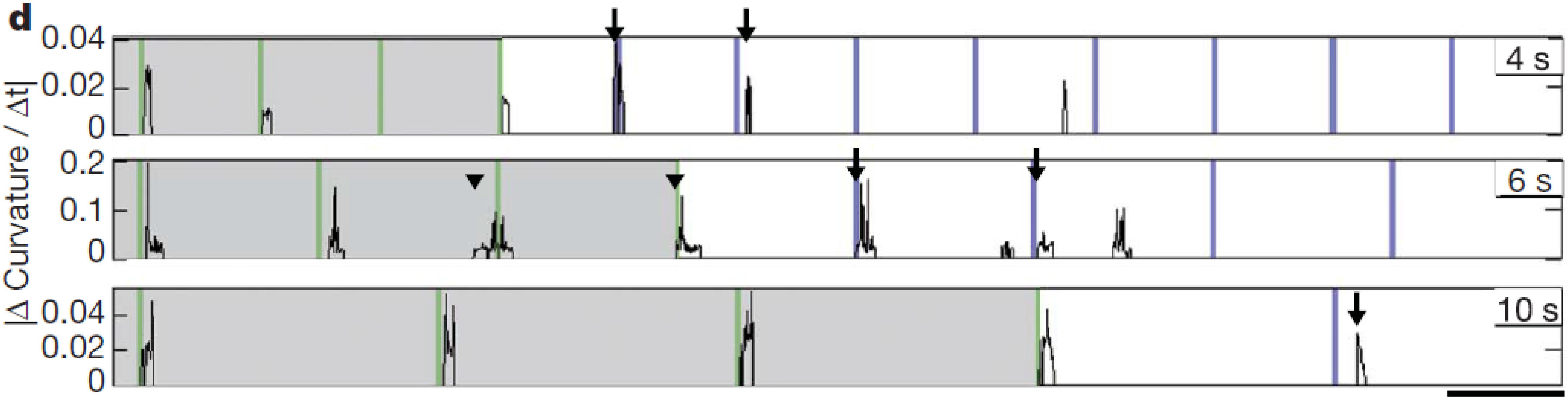
- Liquid state machine
- Echo state machine

Long-duration Rhythmic Synchronous Firings in Zebrafish

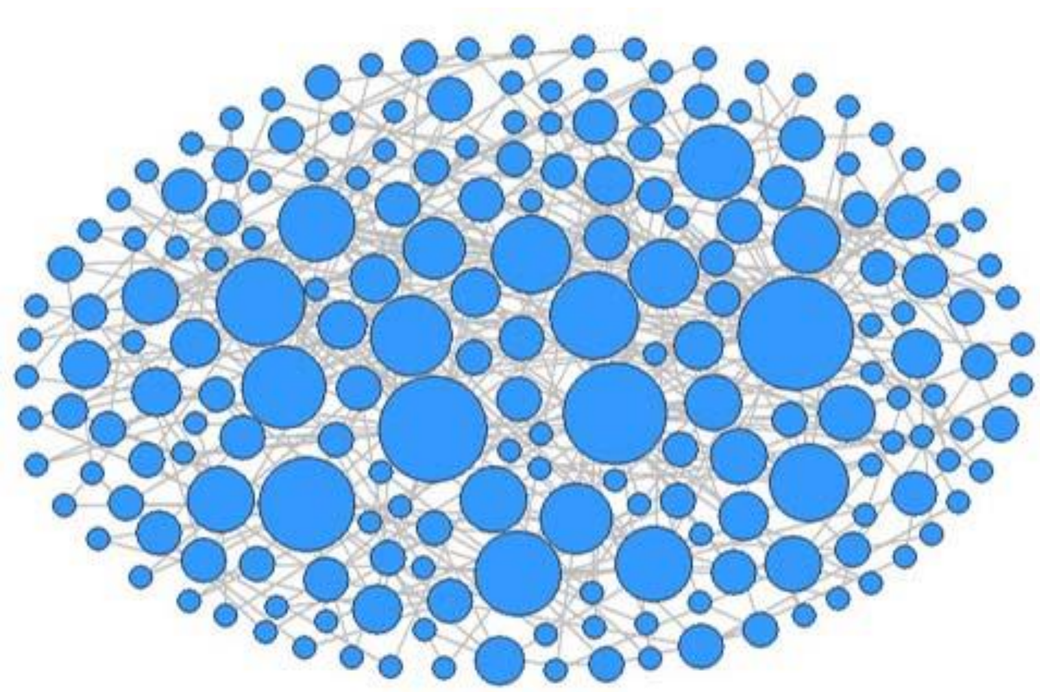


Sumbre et al. Nature, 2009

Substrate of Perceptual Memory



Scale-Free Network



Diameter of a neuron is proportional to its connectivity.
N=200, <K>=4

$$P(K) \sim K^{-\gamma}$$

$$K > 6: \text{ hub neuron}$$

$$K \leq 6: \text{ low-degree neuron}$$

Dynamics of Single Neurons

- **Dynamics of neurons**

$$\frac{du_i}{dt} = -\frac{1}{\varepsilon} u_i (u_i - 1)\left(u_i - \frac{v_i + b}{a} \right) + D \sum_{j=1}^{N} M_{ij} (u_j - u_i)$$

$$\frac{dv_i}{dt} = f(u_i) - v_i .$$

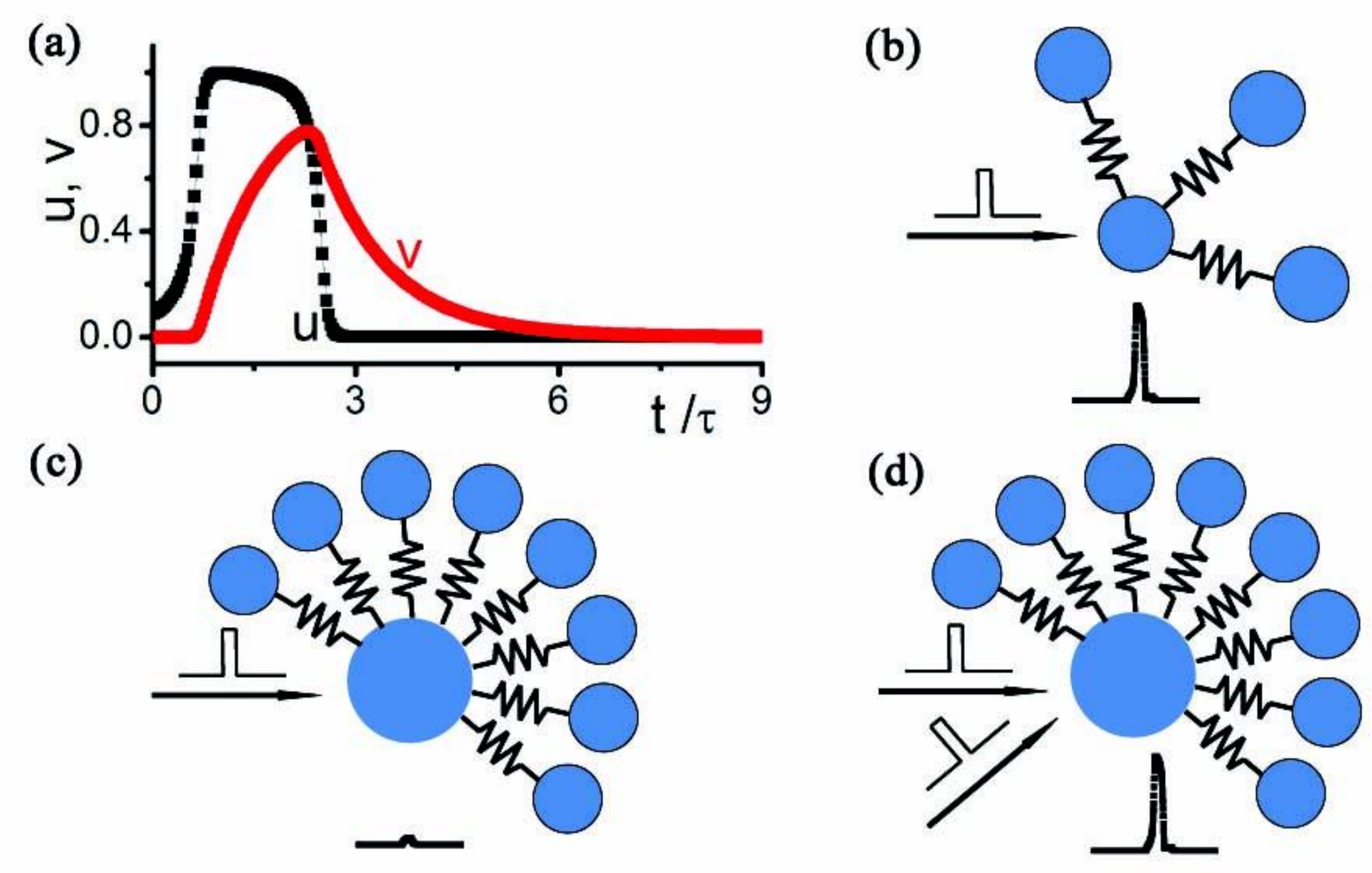
$$f(u_i) = \begin{cases} 0, & u_i < \frac{1}{3} \\ 1 - 6.75 u_i (u_i - 1)^2, & \frac{1}{3} \le u_i \le 1 \\ 1, & u_i > 1 \end{cases}$$

$M_{ij} = 1$ interaction

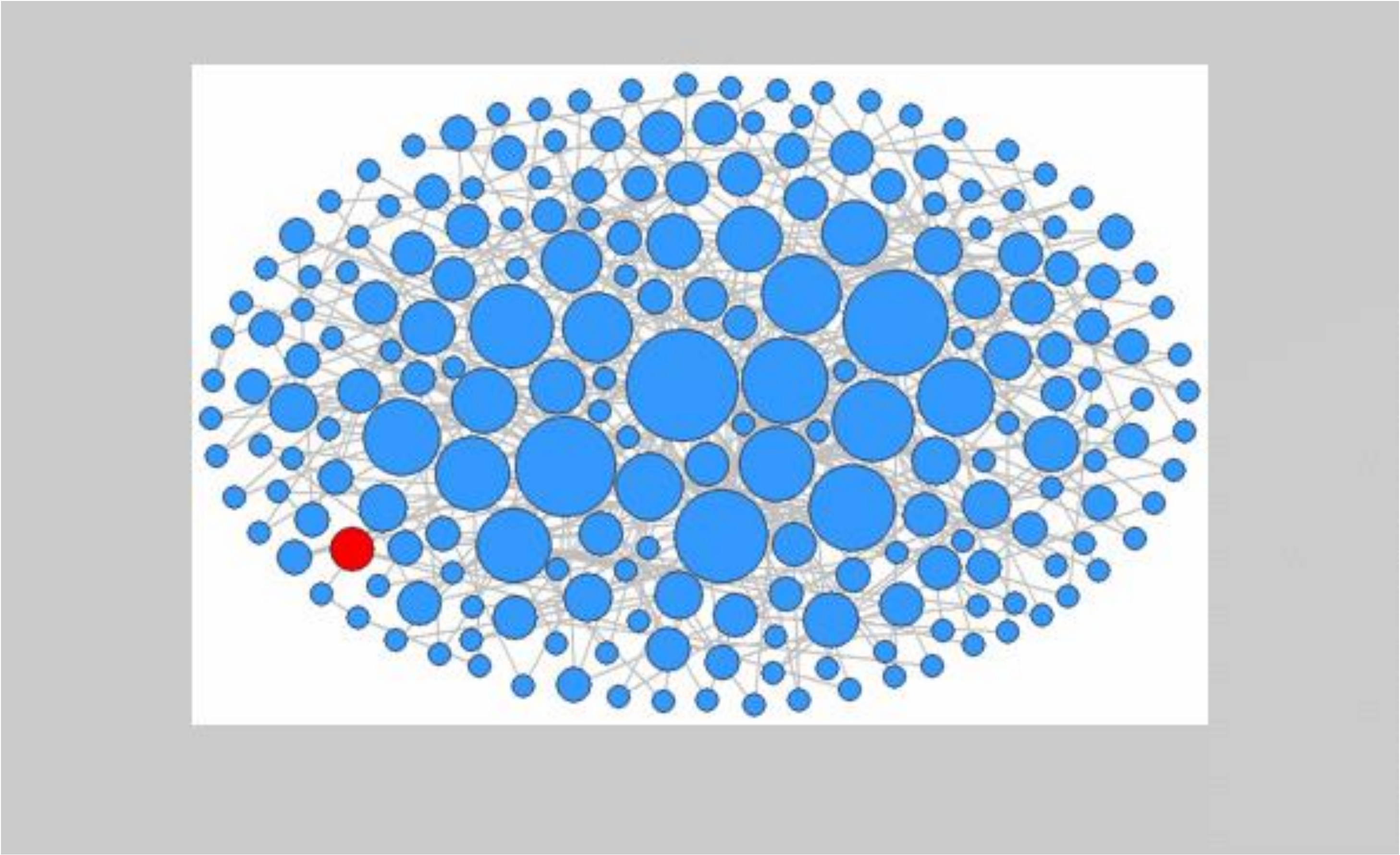
$M_{ij} = 0$ no interaction

- Scaled chemical synapse: higher connectivity, weaker efficacy
- Electrical synapse

Difficulty of Activating Hub Neurons

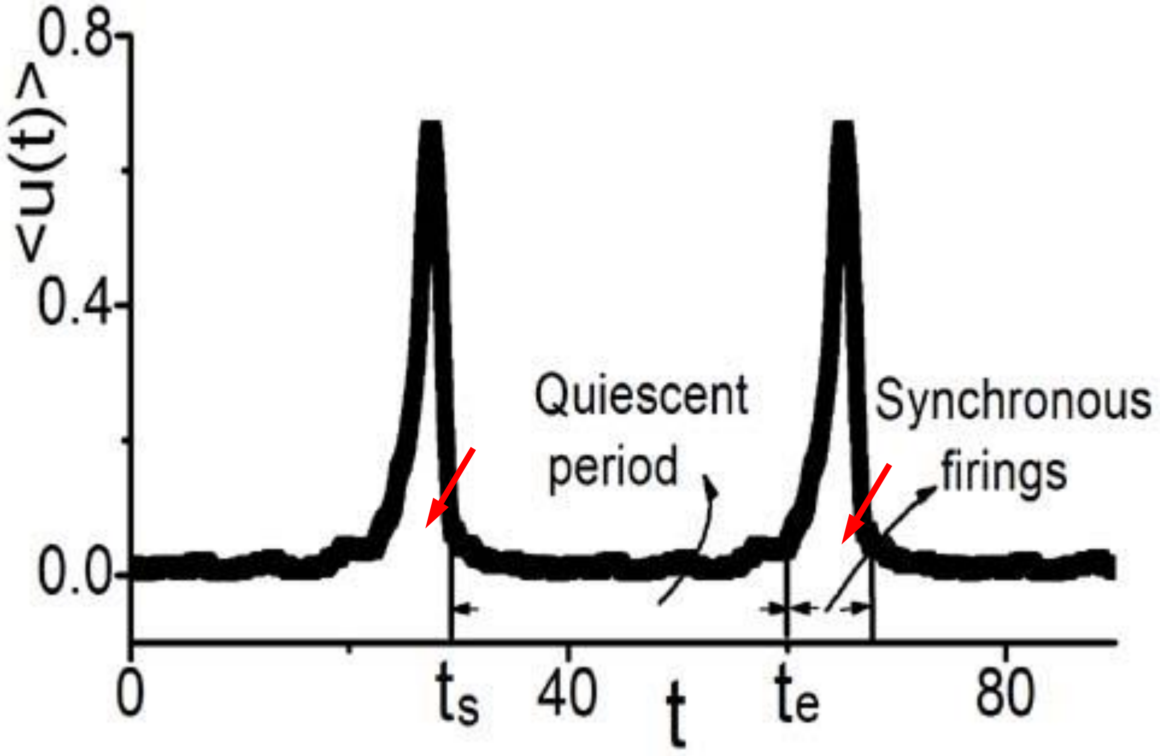


Rhythmic Synchronous Firing in a Scale-Free Network



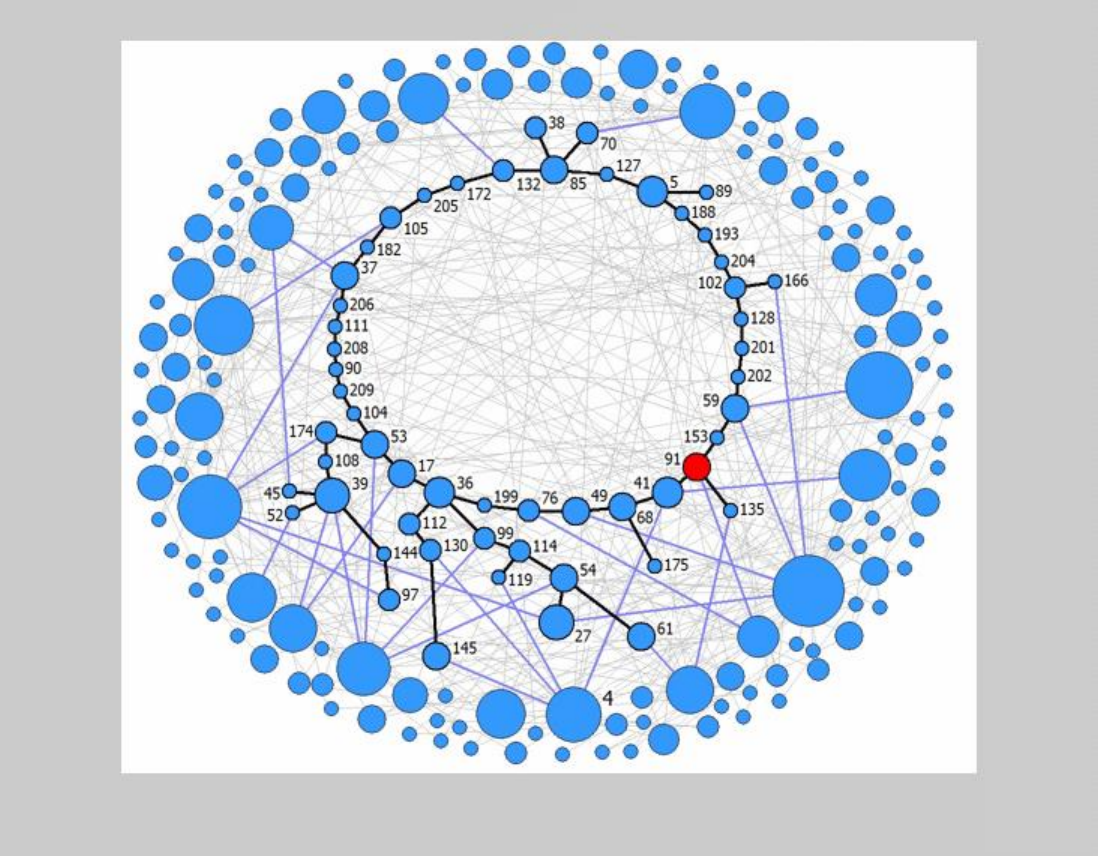
Mi et al., PNAS 2013

Population Activity



$$< u(t) >= \frac{1}{N}\sum_{i=1}^{N} u_i(t)$$

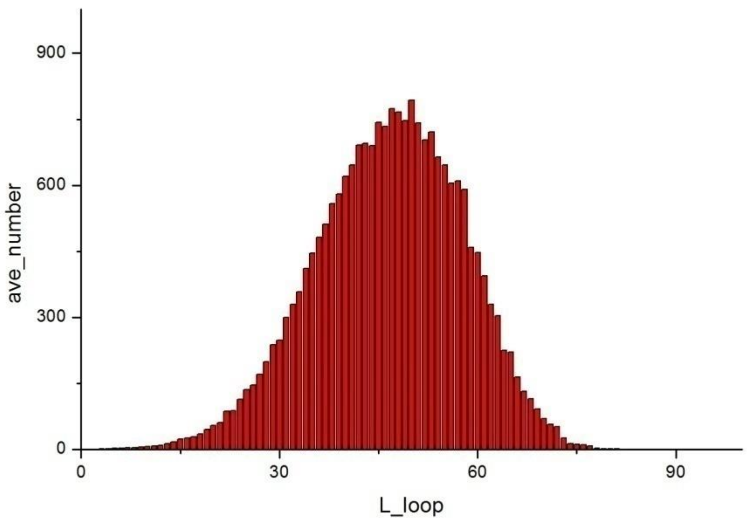
The Mechanism

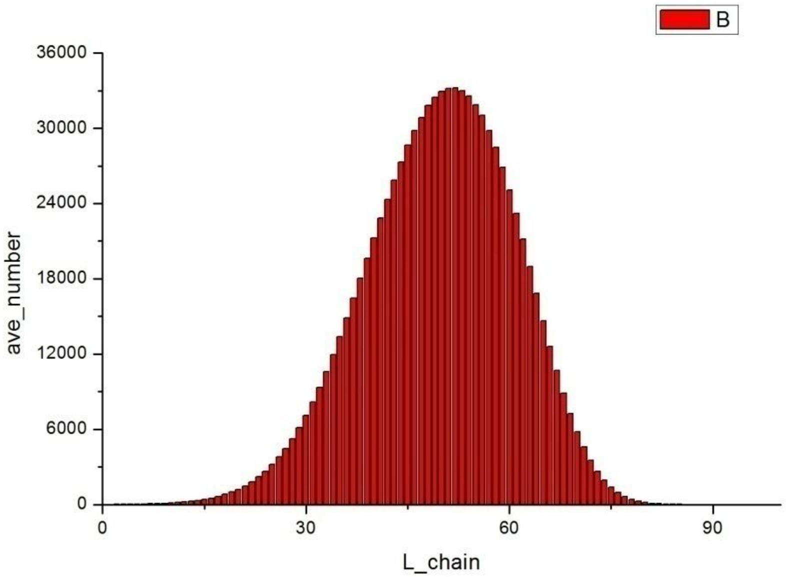


<u>Hub Neurons</u> trigger synchronous firing;

<u>Loop formed by low-degree neurons</u> define the rhythm

Facts of Scale-Free Networks (1)





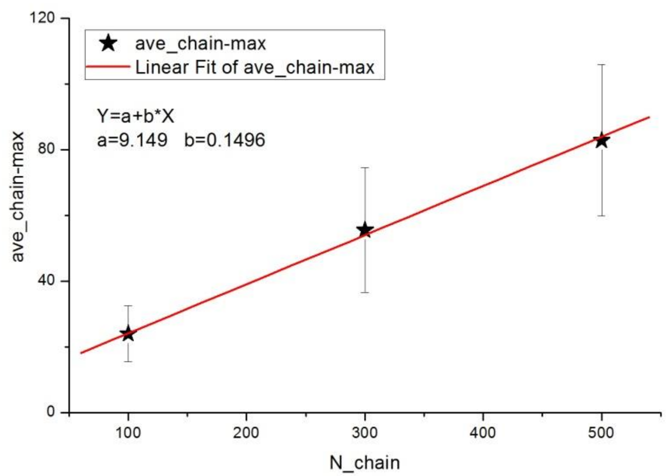
Histogram of loops

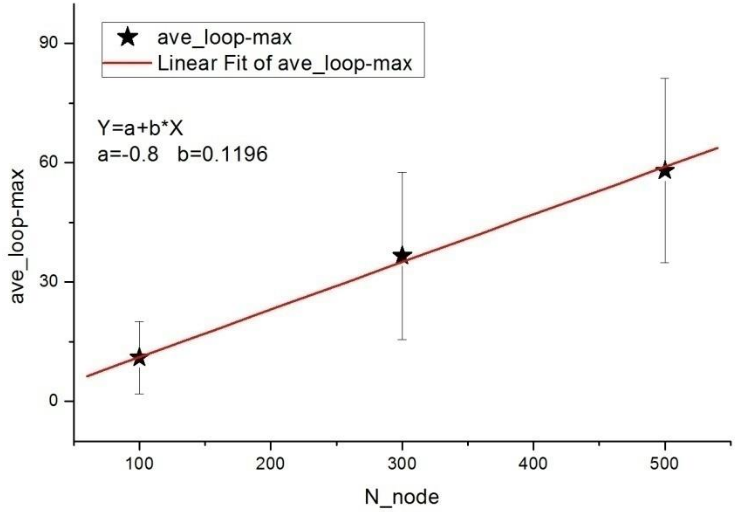
Histogram of chains

After hub neurons are removed.

Facts of Scale-Free Networks (2)

The maximum loop or chain length vs. the network size





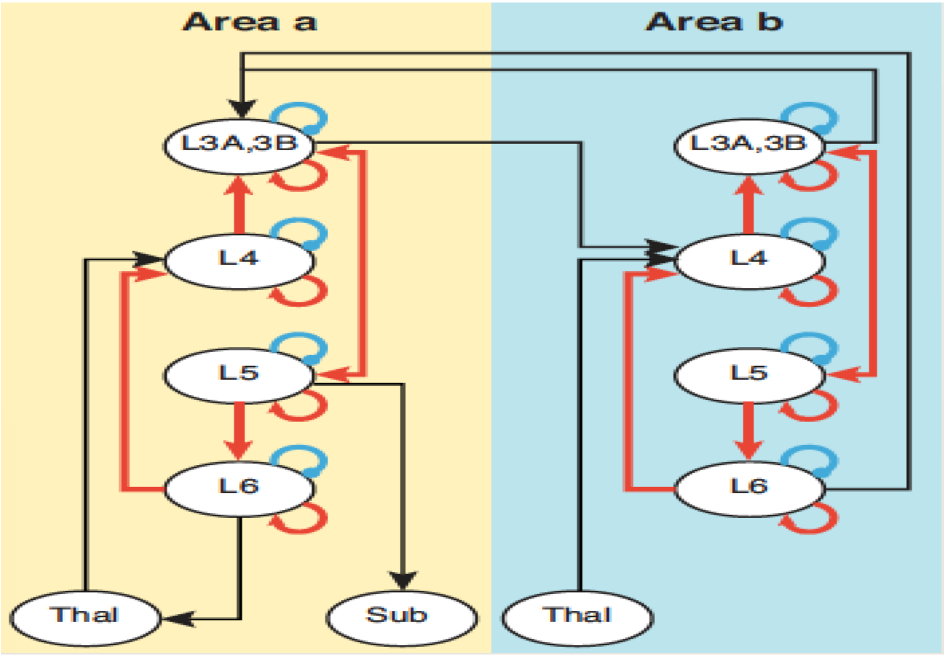
For N=(1~10)*10^4,   loop=1000~10,000

For t=10ms, T=10~100s

Scale-free Network: a general property

➢ Scale-free topology achieves a balance between connection cost & communication efficiency

➢ Reservoir network
—Over-complete resources
—Fast & simple

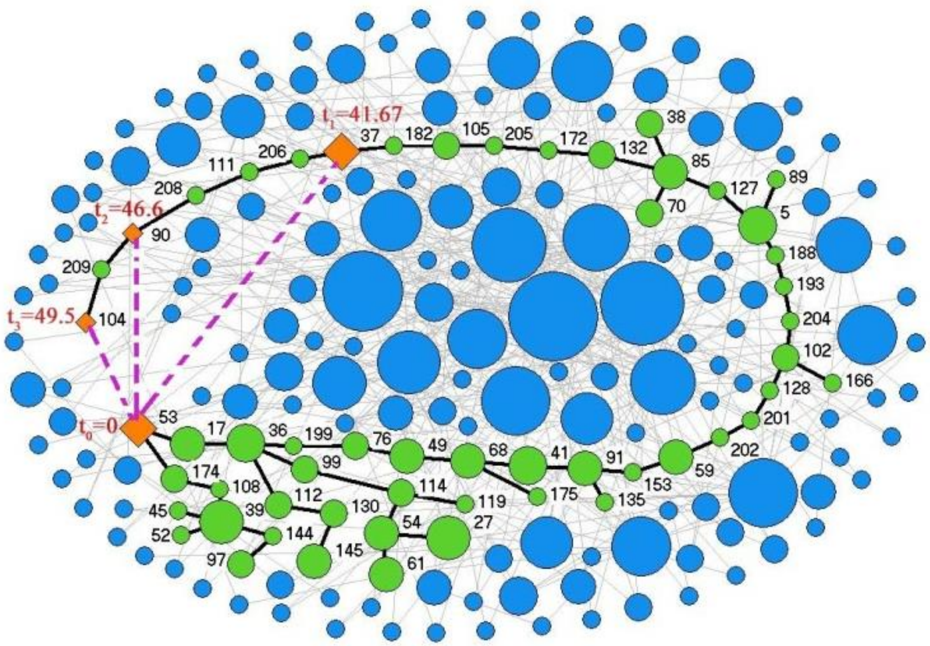


Layer 2/3:
Hubal
Neurons
Layer 5:
Low-degree
Neurons

How to acquire the right loop

◆Matching the rhythm of input with a loop of proper size from a repertoire.

- ➢Long loop/chain holds memory trace
- ➢Hebbian learning establish association



Why Reservoir Computing?

- **Rapid response**, rather than **Efficient utilization**, is crucial for animal survival.

- **Reservoir computing** provides a simple yet efficient way to process information.

- **Spatial-temporal information** of the inputs folded into **pattern of spatial activity** in a large-size network.