- **Article:** Neural Networks and physical systems with emergent collective computational abilities
- **Authors:** J. J. Hopfeild

1 Main idea

In this article, Hopfield proposed a mechanism for a network system to exhibit computational abilities, such as contend-addressable memory and categories of generalization.

If a system has contend-addressable memory and we input enough partial information, it can give us the completed information of the memory. If we use language in physics to describe this ability, we could imagine that we use many coordinates to describe a system. There are some local stable points which could be denoted as $X_a, X_b \dots$. They store the information of memory. The starting points near the local stable states bring "partial" information of the state. As the system evolves, it may finally reach one of those states and thus

2 Model

In a neuron network, neurons are represented by vertices and the edges describe the interaction between them. A neuron i could have two state: $V_i = 0$ and $V_I = 1$.

The connection between neurons i and j is defined as T_{ij}. Each neuron updates its state in random time interval but with an average rate W. The rejustment of states satisfies the following rule:

$$V_i \to 0 \quad if \quad \sum_{j \neq i} T_{ij} V_j < U_i \tag{2.0.1}$$

$$V_i \to 1 \quad if \quad \sum_{j \neq i} T_{ij} V_j > U_i \tag{2.0.2}$$

If there are n sets of memory that we want to store. Each of them could be represented by a vector V^s in phase space. The storage prescription is described as:

$$T_{ij} = \sum_s (2V_i^s - 1)(2V_j^s - 1) \quad for \quad i \neq j \tag{2.0.3}$$

$$T_{ii} = 0 \tag{2.0.4}$$

Then we can get:

$$\sum_j T_{ij} V_j^{s'} = \sum_s (2V_i^s - 1)[\sum_j V_j^{s'} (2V_j^s - 1)] \tag{2.0.5}$$

The mean value of the bracketed term:

$$\begin{array}{l} if \quad s \neq s' \\ < \sum_j V_j^{s'}(2V_j^s - 1) >_s = \sum_j V_j^{s'} < (2V_j^s - 1) >_s = 0 \\ if \quad s = s' \\ < \sum_j V_j^{s'}(2V_j^s - 1) >_s = \sum_j < V_j^s(2V_j^s - 1) >_s = \frac{N}{2} \end{array}$$

This yields:

$$< \sum_j T_{ij} V_j^{s'} > \approx (2V_i^{s'} - 1)\frac{N}{2} \tag{2.0.6}$$