Advances in Neural Networks – ISNN 2004: International Symposium on Neural Networks, Dalian, China, August 2004, Proceedings, Part I. Lecture Notes in Computer Science, Volumen 3173. Springer 2004. 20-25

DOI: http://dx.doi.org/10.1007/978-3-540-28647-9_4

Mutual Information and Topology 2: Symmetric Network

Kostadin Koroutchev **, David Dominguez * **
Eduardo Serrano ** and Francisco B. Rodríguez **

EPS, Universidad Autonoma de Madrid,
Cantoblanco, Madrid, 28049, Spain
{kostadin.koruchev,david.dominguez,eduardo.serrano,
francisco.rodriguez}@ii.uam.es

Abstract. Following Gardner [1], we calculate the information capacity and other phase transition related parameters for a symmetric Hebb network with small word topology in mean-field approximation. It was found that the topology dependence can be described by very small number of parameters, namely the probability of existence of loops with given length. In the case of small world topology, closed algebraic set of equations with only three parameters was found that is easily to be solved.

1 Introduction

There are 10^{11} neurons in a human brain, each neuron is connected with some 10^3 other neurons and the mean radius of connection of each neuron is about 1 mm, although some neurons has extent of decimeters [2]. It seems that, at least in the cortex, there is no critical finite-size subnetwork of neurons. From this considerations it is clear that the topology of this real neural network is far from fully connected, uniformly sparse connected or scale free.

All neural networks (NN) process information and therefore the description of the network in terms of information capacity seems to be the most adequate one. For a living system it is clear that better information processing provides evolutionary advantage.

One can expect, that in order to survive, at least to the some extend, the biological systems are optimal in sense of usage of their resources. Fixing the number of neurons, the main characteristic that can vary in a real NN is the topology of the network, that give rise to the intriguing question:

How does the topology of the network interfere on the information characteristics of the network? The first part of this communication [3] investigates asymmetric small world networks via simulation. Also more complete resume of the existing publications is given. This article tries to answer partially to the previous question, using as a model symmetrical Hebb NN with small world (SW) topology [4] at zero temperature limit.

* DD thanks a Ramon y Cajal grant from MCyT
** Supported by the Spanish MCyT BFI-2003-07276 and TIC 2002-572-C02.

If one considers a network of N neurons, there exists a huge number of $2^{N(N-1)/2}$ different topologies. If all neurons are connected in the same manner, the number of possible topologies is still of order of $2^{(N-1)}$. That makes impossible the exact topology description of the network because the parameters to describe it are too many. Here we show that actually, the dependence of the topology of a symmetric NN can be described by only few real numbers, with clear topological meaning.

The paper is organized as follows: In the following Section, the system is defined and replica-trick solution is given. A zero temperature limit is performed. The topological dependence is found to be expressed by topological coefficients a_k, that are the probabilities of cycle with given length $(k+2)$. Monte-Carlo procedure of finding the topological dependence is described, that makes possible to calculate the dependence on the topology for virtually any topology. It is shown that the method converges fast enough. In Section 3, a small world topology is explored. It is shown that the dependence on the SW topology parameters is trivial. A resume of the results is given in the last section.

2 Replica Trick Equations

We consider a symmetrical Hebb network with N binary neurons each one in state $\sigma_i,\ i \in \{1...N\}$. The topology of the network is described by its connectivity matrix $\mathcal{C} = \{c_{ij} | c_{ij} \in \{0,1\}, c_{ij} = c_{ji}, c_{ii} = 0\}$.

The static thermodynamics of symmetric Hebb neural network in thermodynamic limit ($N \to \infty$) is exactly solvable for arbitrary topology, provided that each neuron is equally connected to the others, and site independent solution for the fixed point exists. Small world topology [4], the case studied in details here, is of that type. The energy of the symmetric network in state $\boldsymbol{\sigma} \equiv \{\sigma_i\}$:

$$H_s = -\frac{1}{2} \sum_{i,j} c_{ij} K_{ij} \sigma_i \sigma_j;\ K_{ij} \equiv \frac{1}{\gamma N} \sum_{\mu=1}^{P} \xi_i^\mu \xi_j^\mu \tag{1}$$

defines completely the equilibrium thermodynamics of the system. In the above equation P is the number of memorized patterns $\boldsymbol{\xi} \equiv \xi_i^\mu$ and γ is the average connectivity of the network, that is the fraction of existing links or the largest eigenvalue of $\mathcal{C}/N$. K_{ij} is defined in the the complementary work [3].

Following Gardner [1], using replica-symmetry trick, in thermodynamic limit ($N \to \infty$) one obtains:

$$m^\mu = \left\langle\!\left\langle \int \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \xi^\mu \tanh \beta\gamma(\sqrt{\alpha r} z + \mathbf{m}.\xi) \right\rangle\!\right\rangle_\xi \tag{2}$$

$$q = \left\langle\!\left\langle \int \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \tanh^2 \beta\gamma(\sqrt{\alpha r} z + \mathbf{m}.\xi) \right\rangle\!\right\rangle_\xi \tag{3}$$

$$r = \mathrm{Tr}_{ij} \left[\frac{q}{\gamma} \left(1 - \frac{\beta \mathcal{C}(1-q)}{N} \right)^{-2} \left(\frac{\mathcal{C}}{N} \right)^2 \right], \tag{4}$$

where $\alpha \equiv P/(\gamma N)$, $\beta = 1/T$ is the inverse temperature, m^μ are the order parameters, q is the mean overlap of σ_i between the replicas and r is the mean overlap of m^μ between the replicas.

The small world topology can be defined as :

$$Prob(c_{ij} = 1) \equiv \omega\gamma + (1 - \omega)\theta(\gamma - ((i - j + N + \gamma N/2) \ \mathbf{mod} \ N)/N)), \quad (5)$$

where ω is the small world parameter defined as p in [4] and $\theta(.)$ is the θ-Heaviside function.

At $T \to 0$, assuming that only one $m^\mu \equiv m$ differs from zero, that is the system is in ferromagnetic state, and keeping the quantity $G \equiv \gamma\beta(1 - q)$ finite, in the equation (2) the $\tanh(.)$ converge to sign(.), in the next equation $\tanh^2(.)$ converges to $1 - \delta(.)$ and in (4) the expression can be expanded in series by G, giving:

$$m = \mathrm{erf}(m/\sqrt{2r\alpha}) \quad (6)$$

$$G = \sqrt{2/(\pi r\alpha)}e^{-m^2/2r\alpha} \quad (7)$$

$$r = \sum_{k=0}^{\infty}(k + 1)a_k G^k, \quad (8)$$

where $a_k \equiv \gamma \mathrm{Tr}((\mathcal{C}/\gamma N)^{k+2})$. Note that a_k is the probability of existence of cycle of length $k + 2$ in the connectivity graph.

The only equation explicitly dependent on the topology of the system is the equation (8) and the only topology describing characteristics important for the network equilibrium are a_k. As $k \to \infty$, a_k tends to γ. Actually, for the small world topology $\forall k > 40, \ |a_k - \gamma| < 10^{-2}$, except for extremely small γ, ω.

The coefficients a_k can be calculated using a modification of the Monte-Carlo method proposed by [1], see Fig. 2,*Left*. Namely, let us consider k as a number of iteration and let us regard a particle that at each iteration changes its position among the network connectivity graph with probabilities defined by the topology of the graph (small world). If the particle at step k is in node $j(k)$, we say that the coordinate of the particle is $x(k) \equiv j(k)/N$. When $N \to \infty$, $x(k)$ is continuous variable. If n is uniformly distributed random variable between -1/2 and 1/2, then in step $k + 1$ the coordinate of the particle changes according to the rule:

$x(k + 1) = x(k) + \gamma n$ **mod** 1 with probability $1 - \omega$ and

$x(k + 1) = x(k) + 1n$ **mod** 1 with probability ω.

If the particle starts at moment $k = 0$ form $x(0) = 0$ and is in position $x(k)$ at step k then the probability to have a loop of length $k + 1$, according to the small world topology is (assuming $x(k) \in [-1/2, 1/2]$):

$$\hat{a}_{k-1} = \theta(\gamma/2 - |x(k)|)(1 - \omega) + \omega\gamma \quad (9)$$

One can estimate easily the speed of the convergence of $\langle \hat{a}_k \rangle$ to a_k. As a_k decrease with k, it is sufficient to estimate the error when k is large. But when k is large, the coordinate at the step k is uniformly distributed in $[0, 1)$ and therefore the process of selecting $\hat{a}_k$ is Bernoulli process that with probability γ gives value

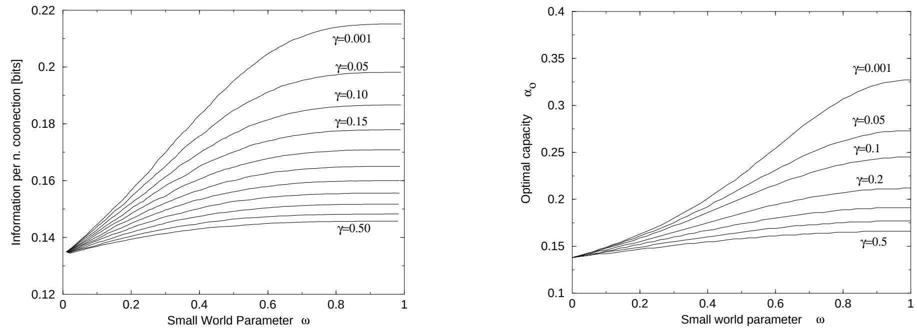

Fig. 1. *Left:* The optimal information capacity per connection $i(\alpha_o; \omega, \gamma)$ as a function of the network topology. Small world parameter ω for different dilutions γ. *Right:* The optimal capacity α_o. No minima nor maxima are observed.

of $(1-\omega+\omega\gamma)$ and with probability $(1-\gamma)$ gives value $\omega\gamma$. Therefore, the error of the calculation behaves as $1/\sqrt{MCtrials}$. As a practical consequence, using Monte Carlo method as few as 10000 steps provide good enough (up to 2 digits) precision for $a_k \approx \langle \hat{a}_k \rangle_{MCtrials}$.

The calculation of $a_k = a_k(\mathcal{C})$ closes the system of Eqs.(6-8) giving the possibility to solve them in respect of (G, m) for every (ω, γ, α).

The mutual information per neural connection (information capacity) is given by:

$$i(m, \alpha) \equiv \alpha \left(1 + \frac{1-m}{2} \log_2 \frac{1-m}{2} + \frac{1+m}{2} \log_2 \frac{1+m}{2} \right) [\text{bits}]. \qquad (10)$$

The interesting values of m, α are those of the phase transition $\alpha_c : m(\omega, \gamma, \alpha_c) = 0; \forall \alpha < \alpha_c \quad m(\omega, \gamma, \alpha) > 0$ and those with maximal information capacity $\alpha_o : (di(m, \alpha)/d\alpha)(\alpha_o) = 0, m \neq 0$. The mean field theory provides good approximation far from the phase transition point. If $\alpha_c \approx \alpha_o$ then significant differences of the mean-field prediction can be expected.

3 Small World Topology

Fig. 1 show $i(\alpha_o)$ and α_o. as a function of the topology parameters (γ, ω). In all the cases monotonicity and smooth behavior in respect to the topological parameters is observed. Therefore, a small world topology has no significant impact on the network performance. Particularly, there is no maxima nor minima of the capacity $i(\alpha_o)$ or α_o as a function of the network topology for any intermediate values of (γ, ω) (except the trivial ones $c \to 0$, $\omega \to 1$). In contrast, using simulations it was shown that in the asymmetrical case there exists maxima of the information capacity [3].

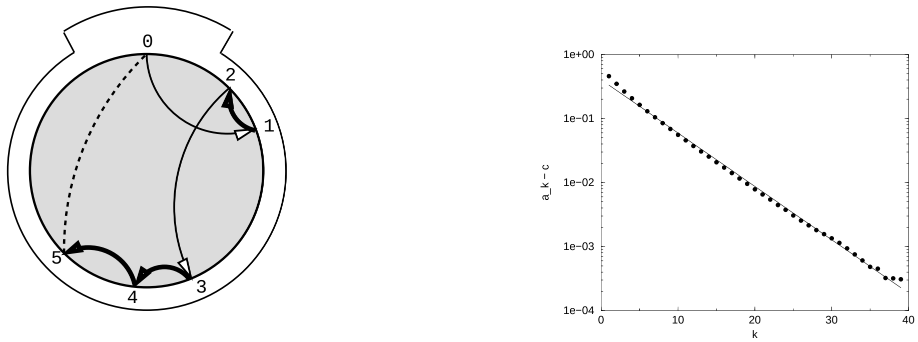

Fig. 2. *Left:* Calculus of a_k by random walk in SW topology. The network with its connection is represented by the gray circle. The height of the surrounding circle represents the probability of connection of the neuron at position 0. The “particle” moves at distance shorter then $\gamma/2$ with high probability (bold lines) and at large distances with smaller probability (thin lines). After 5 steps, the probability of existence of circle with length 6 (a_4) is given by the probability of existence of connection between x_5 and x_0 (dotted line). *Right:* A typical behavior of the topology dependent coefficients $a_k(k) - \gamma$ on the index k, for SW topology with $\gamma = 0.001$ and $\omega = 0.15$ together with the regression fit.

The mean field approximation is valid, because for all values of the parameters of SW, except the case of fully connected network, it was calculated that $1 - \alpha_o/\alpha_c > 2.5\%$.

A typical behavior of the coefficients a_k, $k > 0$ in the equation (8) is shown in Fig. 2,*Right*. One can see that a_k is roughly exponentially falling with the index k toward γ, that is

$$a_k \approx (a_1 - \gamma)e^{-(k-1)\eta} + \gamma. \tag{11}$$

By definition, a_1 is the probability of having 3-cycle, e.g. the clusterization index of the graph.

From this observation it follows that in very good approximation the behavior of the network is described by only 3 numbers: the average connectivity γ, the clustering coefficient a_1 and η.

Substituting the values of a_k in the expression of r (8), one obtains:

$$r = 1 + \gamma[(1 - G)^{-2} - 1] + (a_1 - \gamma)[(1 - e^{-\eta}G)^{-2} - 1], \tag{12}$$

According to the numerical fitting, the parameters a_1 and η depend on γ and ω as:

$$\eta = -0.078 + 0.322\gamma - 0.985\omega + 2.07\gamma^2 - 1.65\gamma\omega - 0.73\omega^2 \tag{13}$$

$$a_1 - \gamma = \exp(-0.413 - 1.083\gamma - 0.982\omega + 0.31\gamma^2 - 6.43\gamma\omega - 3.27\omega^2). \tag{14}$$

Of course, one can find more precise numerical approximations for $\eta, a_1 - \gamma$, using more sophisticated approximations. Although the approximation is very rough, the results for the area of SW behavior of the graph are exact of up to 2 digits.

The equations (6, 7, 12, 13) provide closed system that solves the problem of small world topology network. This system can be solved by successive approximations.

4 Summary and Future Work

The effect of the topology on the performance of a symmetrical Hebb NN at zero temperature was found to be dependent only on small number of real parameters a_k, that are the probabilities of having cycle of length $k+2$ in the connectivity graph. The first a_k, $a_0 = 1$, because the network is symmetric, the second a_1 is the clustering coefficient of the network and the limit value of a_k by large k is the average connectivity of the network γ. Monte-Carlo procedure was designed to find this probabilities for arbitrary topology.

Empirical study of a_k shows that for a small world topology these coefficients converge as exponent to its limit value, that makes possible to express analytically the behavior of SW NN using directly its definition parameters.

Although the replica-symmetry solution for SW is stable, it is not clear if it is so for different topologies. Curiously, although lacking theoretical background, the comparison between the simple solution given here and the simulation for a wide variety of topologies, even asymmetric, is almost perfect. It would be interesting to find similar, theoretically grounded, framework for calculating the behavior of general connectivity distribution and asymmetrical networks.

The particular choice of small world topology seems to have little impact on the behavior (11) of a_k. It seams that the this behavior is universal with large number of rapidly dropping with the distance connectivities. If the mean interaction radius of each neuron is much smaller then N it is easy to show that the equation (11) holds exactly.

However, the exact behavior of η and the range of the validity of the approximation (11) deserves future attention. It would be interesting to find how the parameters just described and the mean minimal path length are connected. It is clear that η increases with the drop of the mean minimal path length.

References

1. Canning, A. and Gardner, E. *Partially Connected Models of Neural Networks,* J. Phys. A, **21**,3275-3284, 1988
2. Braitengerg, V. and Schuez, A. Statistics and Geometry of Neural Connectivity, Springer, 1998
3. Dominguez, D., Korutchev, K.,Serrano, E. and Rodriguez F.B. *Mutual Information and Neural Network Topology.* This Publication
4. Watts, D.J. and Strogatz, S.H. *Collective Dynamics of Small-World Networks,* Nature, **393**:440-442, 1998