available at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/brainres

BRAIN RESEARCH

Research Report

The emergence of connectivity in neuronal networks: From bootstrap percolation to auto-associative memory

*Tatyana S. Turova**

Mathematical Center, University of Lund, Box 118, Lund S-221 00, Sweden

ARTICLE INFO

Article history:
Accepted 24 July 2011
Available online 30 July 2011

Keywords:
Integrate-and-fire network
Random graph
Storage capacity
Percolation

ABSTRACT

We consider a random synaptic pruning in an initially highly interconnected network. It is proved that a random network can maintain a self-sustained activity level for some parameters. For such a set of parameters a pruning is constructed so that in the resulting network each neuron/node has almost equal numbers of in- and out-connections. It is also shown that the set of parameters which admits a self-sustained activity level is rather small within the whole space of possible parameters. It is pointed out here that the threshold of connectivity for an auto-associative memory in a Hopfield model on a random graph coincides with the threshold for the bootstrap percolation on the same random graph. It is argued that this coincidence reflects the relations between the auto-associative memory mechanism and the properties of the underlying random network structure.

This article is part of a Special Issue entitled "Neural Coding".

1. Introduction

In many models of neural networks a large part of a network consists of, say N neurons each of which has C inputs and outputs, with C being much smaller than N. This applies for example to auto-associative and competitive networks found in the brain (see Rolls, 2008). C is often in the range $0.1N$–$0.01N$. Hence, compared to a fully connected network, we are dealing with *diluted connectivity*. Indeed, it is known that a substantial proportion of the initial connections of the densely interconnected neural networks is removed within the first several months or years of life Bourgeois and Rakic (1993), Huttenlocher (1979) and Huttenlocher et al. (1982).

It is also known that the proportion of active neurons in a network for any one stimulus or event (the sparseness of the representation) is in the order of 0.1, C in the cortex is in the order of 10,000, and the number of neurons firing in the densely interconnected patch of cortex for any one event (i.e. at any one time) is of order 100–1000 see Rolls (2008), which in mathematical terms can be treated as a constant, assuming that $N \to \infty$.

Given this evidence of the architecture of the cerebral cortex, a natural question arises: can a network with almost equal small (compare to the size of the network) numbers of in- and out-connections be a result of some selective synaptic pruning in a highly interconnected network?

To get a satisfactory answer consistent with the biology we have to specify a class of allowed dynamics involved in setting up or pruning the connections between the neurons in the network.

One natural scenario for a *connectivity algorithm* is the following. Assume that initially transmission is possible from each to any other of the neurons in the network. Then an external stimulus activates a (small) part of the network, causing a random trajectory of the propagation of the impulses in the network. As soon as an action potential has successfully

* Fax: +46 462224623.
E-mail address: tatyana@maths.lth.se.

0006-8993/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.brainres.2011.07.050

caused firing of the receiving neuron, the weight of the particular connection involved, if the causing neuron is excitatory, is increased by some constant. This is analogous to a Hebb rule for associative learning. It is noted in the context of this synaptic weight modification that there is some evidence that the synaptic weights may follow a discrete distribution Montgomery and Madison (2004).

The activation of the same or a different sets of neurons is repeated for a number of times, each time amplifying the strength of the connections being used from the excitatory neurons to produce the neuronal firing. After a course of learning those connections which have not been activated are weakened, and eventually are pruned.

Simulation of such a scenario see Iglesias et al. (2005) and Iglesias and Villa (2007, 2008) gave a somewhat surprising result, that typically (depending on the size of the inputs and the strengths of the connections), either that the initial impulse activates the entire network, or that, on the contrary, the transmission of impulses stops rather soon, so that the average outgoing number of synapses from a neuron is zero.

Here we argue that an architecture with almost equal numbers of in- and out-connections, which is very close to the synfire chains studied by Bienenstock (1995) and Hertz and Prügel-Bennett (1996), can emerge in a class of stochastic networks given that certain conditions are fulfilled. Furthermore, we show that the set of parameters under which the amount of excitation remains constant while the set of firing neurons is changing in time, is rather small within the whole space of possible parameters. This may account for difficulties in finding the critical parameters by simulations by Iglesias and Villa (2007), Iglesias and Villa (2008) and Iglesias et al. (2005).

Turova and Villa (2007) introduced a mathematical framework to study these questions. Conceptually their study was close to the ideas of Bienenstock (1995), who argued in favor of using spatio-temporal patterns for encoding information in the brain. It was also related to the analysis of activity in synfire chains by Hertz and Prügel-Bennett (1996). The model of Turova and Villa (2007) relies essentially on the known properties of random graphs and percolation. (It is worth recalling that the use of percolation models in neuronal science was predicted already by Harris (1963) who in turn referred to the study of Eccles (1952).) Independently in the same time a number of computational results were reported within a similar paradigm of "neuropercolation" see, e.g., Kozma et al. (2005) and Puljic and Kozma (2005).

The Turova and Villa (2007) model gave rise to a number of mathematical problems investigated and solved in the last few years by Vallier (2007) and Janson et al. (2010). In particular, the recent results of Janson et al. (2010) prove a sharp phase transition in a network with excitatory connections only. We shall argue below that this may also be viewed as a theoretical explanation of the phenomena observed in the simulations of Iglesias and Villa (2007, 2008).

We shall use here the basic model of evolution of random graphs suggested by Turova and Villa (2007) for investigating networks with embedded spike timing-dependent plasticity. With the help of this model we shall construct a process of selective pruning which produces a desired architecture with almost equal numbers of in- and out-connections.

It is interesting to compare our training algorithm with the algorithm used in the Hopfield network in Hopfield (1982). A similar comparison was considered earlier by Hertz and Prügel-Bennett (1996). We shall discuss below how our network can be used in further neural activity modeling, in particular, in connection with Hopfield networks on random matrices. The latter model has been extensively studied after the works of Sompolinsky (1986) and of Derrida, Gardner and Zippelius (1987). For rigorous results in this direction consult Bovier and Picco (1998) and Talagrand (1998).

2. Analysis of the dynamics of activation

2.1. *Bootstrap percolation*

We shall begin by describing a simple model for the spread of activation in a network with random connections, which consists entirely of excitatory neurons. This is bootstrap percolation. Being simple it permits a complete rigorous analysis that will reveal some of the reasons for the difficulties that Iglesias and Villa (2007, 2008) had in simulating their model.

Bootstrap percolation on a given (directed or undirected) graph proceeds as follows:

> *At each point in time any node of an n-neuron network can only be in one of 2 states: active or inactive. Time is measured in discrete units. Assume, that initially there are exactly* $A(0) \ll n$ *active vertices (or neurons in neural network language). At each step, those vertices which have not been active, but which have connections to at least* $r \geq 2$ *active vertices become active as well (in the case of directed network we count only incoming edges). At the same time those active vertices remain active, i.e., we have propagation only when new vertices are excited.*

Our interest is in the total number of activated vertices up to time t; denoted by $A(t)$. If no new vertices are activated at step t, the process stops, i.e. $A(t')$ remains constant for $t' \geq t$. Clearly, the process will not run more than n steps. We want to know the final size of the active set, i.e., $A(n)$.

The bootstrap percolation model has a long history. Chalupa et al. (1979) introduced the model with a random initial set on a regular infinite tree. Afterwards, it was studied rigorously on various regular graphs: e.g., Aizenman and Lebowitz (1988), Cerf and Manzo (2002), (d-dimensional grid), Holroyd (2003) (2-dimensional grid), Balogh and Bollobás (2006) (hypercube). From there the study moved naturally to the bootstrap percolation model on regular random graphs, e.g., Balogh and Pittel (2007). In the context of neuroscience Tlusty and Eckmann (2009) studied bootstrap percolation on more sophisticated metric networks, using it to model activity propagation in the 2D neuronal cultures. Tlusty and Eckmann (2009) also addressed the important neuro-modeling issue of the critical size of the network.

Consider a random graph on n vertices, where each undirected edge is present independently with probability p. The standard notation for this graph is $G_{n,p}$. Vallier (2007) studied bootstrap percolation on $G_{n,p}$. This is the simplest version of a

feed-forward spread of excitation in the neural network described by Turova and Villa (2007). Janson et al. (2010) recently extended the results of Vallier (2007), and provided a complete analysis of bootstrap percolation on $G_{n,p}$. In particular, the phase transition established by Janson et al. (2010) is important in the present paper.

The parameters of the bootstrap percolation model on $G_{n,p}$ are r (fixed as described above), the number of neurons n (tending to ∞), the size $A(0)$ of the initially active set, and the probability $p=p(n)$ that an edge in the graph will be present. (The edges are also called the connections in a neural network). Janson et al. (2010) proved that the model exhibits a sharp phase transition. Meaning that for any $r\geq 2$ and depending on the other parameters of the model, the final size of activation with a high probability is either n–$o(n)$, or $o(n)$. (As usual, $o(n)$ denotes a function much smaller than n, such that $o(n)/n\to 0$ as $n\to\infty$.) This says that if a network is composed only of excitatory neurons, and if any neuron with at least 2 simultaneously incoming impulses becomes active, then with a high probability any amount of initial excitation leads eventually to either an almost entirely firing network ("too much" excitation) or to a tiny fraction of firing neurons ("too little" excitation). Hence, any sensible pruning of the connections is impossible, since either nearly all of the connections are active or nearly all of them are inactive most of the time.

This sharp phase transition between two marginal states ("almost all" or "almost none"), may provide an explanation for the difficulties observed in the simulations by Iglesias et al. (2005) and Iglesias and Villa (2007, 2008). However, keep in mind that in these simulations inhibitory neurons were present as well.

From a theoretical point of view, the results of Janson et al. (2010) confirm the necessity of the inhibitory connections to obtain self-sustained excitation in a random network such as $G_{n,p}$. Note, in addition, that the stable growth rate for activation in a purely excitatory network recently reported by Eckmann et al. (2010) is achieved on a random network, very different than $G_{n,p}$. The modeling of Eckmann et al. (2010) takes advantage of the structure of their network by choosing suitable degree distributions.

Notice also, that the sharp phase transition revealed by Janson et al. (2010) takes places only when $r\geq 2$. The situation when $r=1$ is different. Here only one incoming impulse could be sufficient to excite another neuron. In the latter case, when $p=c/n$, $c>1$, the activation can go through only a positive (strictly less than 1) fraction of the network due to the existence of components in $G_{n,p}$ (Bollobás, 1985).

Next we shall introduce and study a model which incorporates the inhibitory neurons as well.

2.2. *The model of propagation of activation*

Assume that we have n neurons, among which $n_+ := q_0 n$ are excitatory and $n_- := (1-q_0)n$ are inhibitory. Here $0<q_0<1$ is a parameter. Typically it is assumed that $n_+/n_- = 4$, hence typically $q_0 = 4/5$.

Let the probability of the connections between any two different neurons be $p(n)$, and set the strength of any connection be $w>0$. Introduce a random matrix $M(t)=(m_{ij}(t))_{1\leq i,\, j\leq n}$, where each (i, j)-th entry represents the connection from neuron j to neuron i at time t:

$$m_{ij}(t) = \omega\varepsilon_{ij}(t). \tag{1}$$

Assume, that $\varepsilon_{ij}(0)$ are independent (Bernoulli) random variables, taking values 1 or 0. If $\varepsilon_{ij}(0)=1$, we say that there is a link from neuron j to i. We set a probability of connection to be

$$\mathbf{P}\big(\varepsilon_{ij}(0)=1\big) = \frac{c(n)}{n} \geq \frac{1}{n}. \tag{2}$$

The matrix $(\varepsilon_{ij}(0))_{1\leq i,\, j\leq n}$ when $\varepsilon_{ij}(0)=\varepsilon_{ji}(0)$ for all $1\leq i, j\leq n$, corresponds exactly the connections in $G_{n,p}$ graph. We take $1/n$ in Eq. (2) as the lowest value, since it is known from random graph theory (Bollobás, 1985) that otherwise a network will have only small connected components, of order $\log n$.

We shall construct a discrete time dynamics of pruning of connections. We assume that our time-unit here, denote it τ, is longer than the refractory period, so we assume that the neurons in the model described below may fire consecutively in our sequence of discrete times. We use the rescaled time sequence $t=0, 1, 2, \ldots$, which corresponds to the real times $t=0, \tau, 2\tau, \ldots$

Let $A_+(0)=A_+^0$ and $A_-(0)=A_-^0$ denote the subsets of the excitatory and correspondingly the inhibitory neurons which are activated at time $t=0$. For each set A we shall denote here $|A|$ the number of the elements in A. (Notice that due to the symmetry in the model only values $|A_\pm^0|$ matter.) Assume, that

$$\begin{aligned} |A_+(0)| &= q_+ n_+ = q_+ q_0 n, \\ |A_-(0)| &= q_- n_- = q_-(1-q_0)n, \end{aligned} \tag{3}$$

with some functions $0\leq q_+(n),\ q_-(n)\leq 1$, which are other parameters of the model.

Denote $A(t)=A_+(t)\cup A_-(t)$, where $A_+(t)$ and $A_-(t)$ are the firing subsets of excitatory and inhibitory neurons at time $t=0, 1, \ldots$, correspondingly. We define these sets according to the following dynamics.

Assume that each neuron v has a potential $0\leq X_v(t)\leq 1$. When there are no interactions we assume that $X_v(t)$ are independent for different t and v random variables uniformly distributed on $[0, 1]$.

Given at time t sets $A_+(t)$, $A_-(t)$, and independent potentials $X_v(t)$ for $v\notin A(t)$, we define the state of the network at $t+1$ as follows:

$$A(t+1) = \left\{ 1\leq i\leq n : \frac{X_i(t) + \sum_{j\in A_+(t)} m_{ij}(0)}{1 + \sum_{j\in A_-(t)} m_{ij}(0)} \geq 1 \right\}. \tag{4}$$

In other words, neuron i fires at time $t+1$ if

$$\frac{X_i(t) + \sum_{j\in A_+(t)} m_{ij}(0)}{1 + \sum_{j\in A_-(t)} m_{ij}(0)} \geq 1, \tag{5}$$

which is a typical integrate-and-fire dynamics. Making use of the definition (1) we can rewrite this as follows

$$X_i(t) + w\big(N_i^+(t) - N_i^-(t)\big) \geq 1, \tag{6}$$

where $N_i^+(t)$ and $N_i^-(t)$ denote the numbers of excitatory and inhibitory connections, correspondingly, to neuron i from the firing neurons at time t.

Notice that the assumption of independence of $X_v(t)$ for different times means that we do not assume an accumulation of impulses from different times. This of course simplifies the analysis. However, simultaneous impulses from different neurons are accumulated. The dynamics of our network is described therefore by a sequence

$$\left\{\left(\frac{|A_+(t)|}{n_+},\frac{|A_-(t)|}{n_-}\right), t=0,1,2,\ldots\right\} \tag{7}$$

which forms a Markov chain on $[0,1]\times[0,1]$.

Let us compare now the introduced model with the bootstrap percolation discussed above. If we do not have a population of inhibitory neurons ($n_-=0$) and if $w=1/r$, the condition (6) will be fulfilled, at least when $N_i^+(t)\geq r$. This is the same condition of propagation as in the former bootstrap model. The presence of the inhibitory connections makes, however, the crucial difference: the spread of activation is no longer monotone along the parameter of connectivity of the graph, p. Also observe that we consider not only the dynamics of the feed forward excitation (as in the former bootstrap percolation model) but also we allow excitation "backwards" to the previously fired neurons.

The introduced model combines features of synfire chains and *competing percolation* (the name reflects the counter-effects of excitatory and inhibitory connections). It has also similarities with the model previously studied by Hertz and Prügel-Bennett (1996) and Eckmann et al. (2010).

2.3. *Balanced propagation of impulses*

We consider first the propagation of impulses in the network from the first single stimulus, i.e., the initially activated set $A(0)$. For the entire first training period $[0, T]$ (T will be specified later on) we assume that the connections $\varepsilon_{ij}(t)=\varepsilon_{ij}(0)$, i.e., we maintain the same realization of the random variables which represent the existence of the connections in the network.

We shall find the values of the parameters $c(n)$, q_+, q_-, and w such that the expected values of the numbers of firing excitatory and inhibitory neurons remain approximately the same as at time $t=0$, i.e.,

$$\begin{aligned}&\lim_{n\to\infty}\mathbf{E}|A_+(t)|/n_+ = q_+,\\ &\lim_{n\to\infty}\mathbf{E}|A_-(t)|/n_- = q_-.\end{aligned} \tag{8}$$

This will ensure a certain balance in the propagation of impulses, and will help to find parameters such that

$$\begin{aligned}|A_+(t)| &\approx q_+n_+,\\ |A_-(t)| &\approx q_-n_-,\end{aligned} \tag{9}$$

for all $t\leq T$.

Consider the limiting (as $n\to\infty$) mean drifts of the Markov chain (7), which are defined conditionally on $|A_+(t)|$ and $|A_-(t)|$ as follows

$$D_+(q_+,q_-) := \lim_{n\to\infty}\frac{1}{n_+}\mathbf{E}\{|A_+(t+1)|-|A_+(t)|\,||A_+(t)| = q_+n_+, \ |A_-(t)| = q_-n_-\}, \tag{10}$$

$$D_-(q_+,q_-) := \lim_{n\to\infty}\frac{1}{n_-}\mathbf{E}\{|A_-(t+1)|-|A_-(t)|\,||A_+(t)| = q_+n_+, \ |A_-(t)| = q_-n_-\}. \tag{11}$$

Our first task is to find for which q_+, q_- (if for any)

$$D_+(q_+,q_-)=0, \quad D_-(q_+,q_-)=0. \tag{12}$$

This condition roughly means that the numbers of firing excitatory and inhibitory neurons in the network remain constant within a single trial.

Denote the conditional probability (see Eq. (5))

$$p_n(q_+,q_-)=\mathbf{P}\{X_i+w(N_i^+(t)-N_i^-(t))\geq 1\,||A_+(t)|=q_+n_+,|A_-(t)|=q_-n_-\}.$$

Then conditionally on $|A_+(t)|=q_+n_+$ and $|A_-(t)|=q_-n_-$, the distribution of $|A_+(t+1)|$ is binomial $Bin(n_+, p_n(q_+, q_-))$. Hence,

$$\begin{aligned}D_+(q_+,q_-) &= \lim_{n\to\infty}\left(\frac{n_+p_n(q_+,q_-)}{n_+}-\frac{q_+n_+}{n_+}\right)\\ &= \lim_{n\to\infty}p_n(q_+,q_-)-q_+ = p(q_+,q_-)-q_+,\end{aligned} \tag{13}$$

where

$$p(q_+,q_-) = \lim_{n\to\infty}p_n(q_+,q_-),$$

and similarly

$$D_-(q_+,q_-) = p(q_+,q_-)-q_-. \tag{14}$$

Then Eq. (12) becomes

$$p(q_+,q_-)-q_+=0, p(q_+,q_-)-q_-=0. \tag{15}$$

This immediately implies that we have to assume

$$0<q_-=q_+=:q<1 \tag{16}$$

in order to have any solution to Eq. (12).

Observe that we get here condition $q_-=q_+$ solely due to the symmetry in the model, namely that both excitatory and inhibitory neurons receive the same impulses with the same strength. One can change this assumption in order to have more realistic model. However, qualitatively the result will be similar.

Denote further

$$p(q) = \lim_{n\to\infty}p_n(q,q),$$

and consider system (12), which is now one equation

$$p(q)=q. \tag{17}$$

Notice that conditionally on $|A_+(t)|=qn_+$, $|A_-(t)|=qn_-$, the random variable $N_i^+(t)$ has a binomial distribution $Bin\left(qn_+,\frac{c(n)}{n}\right)$, when $i\notin A_+(t)$, and $Bin\left(qn_+-1,\frac{c(n)}{n}\right)$, otherwise, where $qn_+=qq_0n$.

Assume from now on that function $c(n)$ in definition (2) is constant, i.e., we set

$$c(n)=c. \tag{18}$$

This assumption implies that as $n\to\infty$ the distribution of $N_i^+(t)$ converges to a Poisson distribution with parameter qq_0c. Similarly, the distribution of $N_i^-(t)$ converges to a Poisson distribution with parameter $q(1-q_0)c$. Let further $Y(a)$, $Y'(a)$ for any $a>0$ denote independent random variables with Poisson distributions with parameter a. Taking into account that the random variables $N_i^+(t)$ and $N_i^-(t)$ are independent for each t and i, we derive

$$p(q) = \lim_{n\to\infty}p_n(q,q) = \mathbf{P}\{wY(qq_0c)-wY'(q(1-q_0)c)\geq 1-X\}, \tag{19}$$

where X has a uniform distribution on [0, 1] independent of Y and Y'.

Assume that $1/w>1$ is a natural positive number. (Typically $1/w\sim 7$–100, since 7–100 simultaneously incoming impulses are considered to be sufficient for firing of a neuron.) Then, since $1-X$ has the same distribution as X, we have

$$\begin{aligned} p(q) &= \mathbf{P}\left\{Y(qcq_0) - Y'(q(1-q_0)c) \geq \frac{X}{w}\right\} \\ &= \sum_{n=0}^{\frac{1}{w}-1} \int_{nw}^{(n+1)w} \mathbf{P}\{Y(qcq_0) - Y'(q(1-q_0)c) \geq n+1\}\,dx \\ &= \sum_{n=0}^{\frac{1}{w}-1} w\mathbf{P}\{Y(qcq_0) - Y'(q(1-q_0)c) \geq n+1\} \\ &= \sum_{n=0}^{\frac{1}{w}-1} w \sum_{k=0}^{\infty} \sum_{l=k+n+1}^{\infty} e^{-cq} \frac{(qcq_0)^l}{l!} \frac{(q(1-q_0)c)^k}{k!}. \end{aligned} \tag{20}$$

Clearly,

$$p(0) = 0.$$

It is also straightforward to compute that

$$p'(q) = cq_0 \sum_{n=0}^{\frac{1}{w}-1} w \sum_{k=0}^{\infty} e^{-cq} \frac{(qcq_0)^{k+n}}{(k+n)!} \frac{(q(1-q_0)c)^k}{k!}, \tag{21}$$

which is obviously positive for all $q>0$.

We proved therefore that $p(q)$ is a monotonically increasing function on [0, 1], strictly less than 1 (as a probability of an event which does not cover the entire space of all possible outcomes), and $p(0)=0$.

Let us also consider $p(q)$ as a function of c and w, and therefore we write $p(q)=p_{c,\,w}(q)$. It follows from the last line in Eq. (20) that $p_{c,\,w}(q)=p_{1,\,w}(cq)$, and thus $p_{c,\,w}(q)$ monotonically increases in c as well, since we proved that it increases in q for all $c>0$, $w>0$. Also it follows straight from the first line in Eq. (20) that $p_{c,\,w}(q)$ is increasing in w as well.

Since function $p(q)$ for any fixed c and w is monotone increasing from 0 at $q=0$ to 1 at $q=1$, and infinitely differentiable, we conclude that equation $p(q)=q$ (see Eq. (17)) has at most a finite number of the solutions.

In particular, formula (21) implies

$$p'(0) = q_0 c\omega.$$

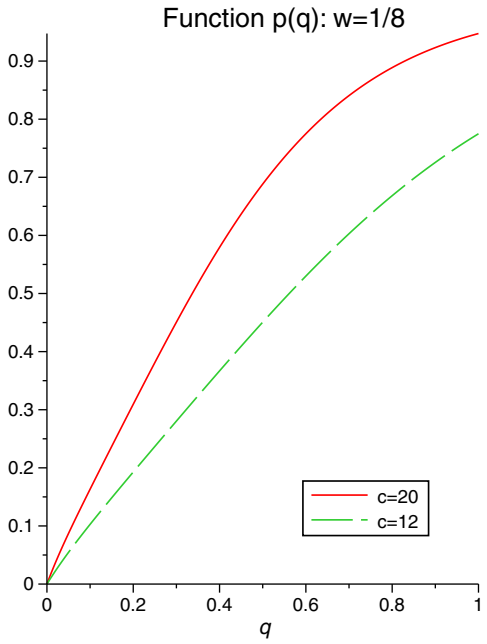
Hence, a sufficient condition for the existence of at least one non-zero solution to Eq. (17) is

$$q_0 c\omega > 1. \tag{22}$$

Recall that here $q_0=n_+/n=4/5$ is a proportion of the excitatory neurons in the entire network, $1/\omega$ is a typical number of excitatory inputs sufficient to fire a neuron, and constant c is the parameter in our assumption that the probability of any single connection is c/n. We can rewrite the sufficient conditions (22) for the existence of at least one solution to Eq. (17) as

$$c > \frac{n_+}{n}\frac{1}{\omega} = \frac{5}{4}\frac{1}{\omega}. \tag{23}$$

Numerical analysis shows that function $p(q)$ defined in Eq. (20) for some reasonable parameters (as e.g., $q_0=4/5$, $7<1/\omega<100$, and $5/(4\omega)\leq c<100$) is convex ($q(p)<0$). See, e.g., enclosed plots for the function $p(q)$ with $w=1/8$ for 2 choices of c: $c=20$, $c=12$.



Hence, for the appropriate range of parameters condition (22) is not only sufficient but also necessary for the existence of a strictly positive solution to Eq. (17). When this is the case, condition (23) defines a *threshold* value for the connection strength c in the following sense: if $c<5/(4\omega)$, then for any input the numbers of neurons in the consecutive layers of firing neurons will decrease, and eventually activity will not propagate any further.

Under condition $q^*(p)<0$ (which holds at least for the range of parameters that we are interested in) for every value of a which satisfies Eq. (22), the strictly positive solution to Eq. (17) is unique; let us denote it q^*, so that $q^*=p(q^*)$ and $q^*>0$. Recall that function $p(q)=p_{c,\,w}(q)$ is increasing both in c and in w for any fixed q. This implies that q^* is also increasing in c and in w.

The strictly positive solution q^* is a *stable* fixed point of the dynamical system

$$q_{t+1} = p(q_t), \tag{24}$$

with initial state $q_0\in(0, 1)$. This stability allows us to conclude that the Markov chain (7) (for large n) will spend a long time within a small neighborhood of the fixed point (q^*, q^*).

We conclude that for any given parameters ω and n_+/n (which we assume to be $n_+/n=4/5$) and c, which satisfies Eq. (23), there exists a unique value $0<q^*<1$, such that the activity of the system when (on the average) q^*n_+ out of n_+ excitatory, and q^*n_- out of n_- inhibitory neurons fire, is stable. In other words, given an input of q^*n_+ excitatory and q^*n_- inhibitory neurons, the system will preserve approximately the same numbers of firing excitatory and inhibitory neurons. The neurons, which fire simultaneously at the same time, can be pictured as layers in the network. We may call q^* the self-sustaining activity load, such that the mean number of

excitatory (inhibitory) neurons in every layer remains to be q^*n_+ (q^*n_-). Furthermore, this value q^* is stable, which means that if we start with lesser or greater amount of excitation it will stabilize around the value q^*n.

Observe, however, that the theoretical prediction $0<q^*<1$ one should take with a caution when dealing with numerical computations. For example, when $w=1/8$ condition (22) tells us that for $c>10$ we should find a value $0<q^*<1$. It turns out (calculations in Maple) that, as predicted, $q^*=0$ for $c\leq 10$. Then for $c=12$ we have $q^*=0.13$, and further increase of c quickly moves q^* towards value 1, so that we have $q^*=0.99$ already for $c=30$. The positive solution q^* is increasing in c as we argued above, therefore $q^*\geq 0.99$ for all $c\geq 30$. But $q^*=1$ is the saturation point when the entire network is firing. When $w=20$ the condition $c>25$ predicts existence $0<q^*<1$. Numerical computations in this case give $q^*=0.1$ for $c=30$ and $q^*=0.99$ already for $c=50$.

Hence, even when condition (23) is satisfied, in practice it is only a small range of values of the remaining parameter w which gives a stable solution q^* substantially less than 1 and greater than 0. This well may be the explanation for the dichotomy observed by Iglesias and Villa (2007) and Iglesias and Villa (2008).

2.4. Numbers of in- and outcoming firing impulses per neuron

Let $N_n^{in,\upsilon}(q)$ ($N_n^{out,\upsilon}(q)$) denote a random number of incoming (outcoming) excitatory impulses for a firing at time $t+1$ of neuron υ in a network with the self-sustaining activity load of excitation in a network, i.e., assuming that $|A_+(t)|=qn_+, |A_-(t)|=qn_-$ with positive q satisfying $q=p(q)$. Similarly to Eq. (20) we derive for any firing neuron υ

$$\nu^{in}(q) := \lim_{n\to\infty} \mathbf{E}N_n^{in,\upsilon}(q) = \mathbf{E}\left\{ Y(qcq_0) \,|\, Y(qcq_0) - Y'(q(1-q_o)c) \geq \frac{X}{w} \right\} = \frac{qcq_0}{p(q)}\left(p(q) + \frac{p'(q)}{cq_0} \right). \quad (25)$$

Hence, under assumption of convexity of $p(q)$, since $q=p(q)$ and $p'(q)<1$ (because $p'(q)>0$ and q is a positive solution to $q=p(q)$) we get from here that

$$qcq_0 \leq \nu^{in}(q) = cq_0 q + p'(q) < qcq_0 + 1. \quad (26)$$

Clearly, we have the following balance equation, which simply says that the number of outcoming connections at time t equals the number of incoming connections at time $t+1$:

$$\sum_{\upsilon\in A_+(t)} N_n^{out,\upsilon}(q) = \sum_{\upsilon\in A_+(t+1)} N_n^{in,\upsilon}(q).$$

This together with the assumption of stability, i.e., that the fraction of firing neurons remains constant $|A_+(t)|=|A_+(t+1)|=qn_+$, yields for $\nu^{out}(q) = lim_{n\to\infty}\mathbf{E}N_n^{out,\upsilon}(q)$:

$$\nu^{out}(q) = \nu^{in}(q).$$

Hence, the average out-degree in the subgraph on the excitatory neurons only, will approximately be equal to the in-degree, which by Eq. (26) is about qq_0c.

2.5. Training sets and pruning

Assume now that for a long time T the Markov chain (7) evolved in the neighborhood of the stable point (q, q). Notice, that the probability that some neuron does not fire within time T is approximately $(1-p(q))^T$. Therefore it is appropriate to choose T so that $(1-p(q))^T$ is small.

Then we introduce a new stimulus $A^1=A_+^1\cup A_-^1$, choosing set A_+^1 (A_-^1) uniformly among n_+ excitatory (n_- inhibitory) neurons with $|A_+^1|=qn_+$ ($|A_-^1|=qn_-$). We also reset the connections $\varepsilon_{ij}(T)$ to be again a new independent realization of the binary random variables, and let the excitation pass through the network for another T time units.

We repeat these training trials for αn number of times. After αn trials we modify the parameters $m_{ij}(T\alpha n+1)$ of the network, setting $m_{ij}(T\alpha n+1)=0$ if the connection from i to j has never been realized or activated during the entire history of training.

Note that the probability that one particular connection will not be chosen in any of the trials is $\left(1-\frac{c}{n}\right)^{\alpha n} \approx e^{-c\alpha}$. Hence, if we choose α so that $e^{-c\alpha}=0.1$, we achieve the average number of connections per neuron $0.1n$, which agrees with data on the synaptic connections in the brain (see, e.g., Rolls, 2008), and the resulting network will fulfill the conditions of a diluted network (in terms of Rolls, 2008; Treves, 1991).

One can modify the probabilities of connections in a different way as well, depending on the further purpose of the model. But in any case, the important feature of the system will be the underlying architecture of the connections, which will enable a steady flow of impulses (information) through the system.

2.6. Relation to the Hopfield model on random graphs

One of the most successful and versatile model of neural network is the Hopfield model (introduced originally as a simple model of a spin glass by Pastur and Figotin (1978)) considered by Hopfield (1982) as a model for auto-associative memory. After the seminal work of Amit et al. (1987), who provided statistical mechanics formalism for a more complicated version of the Hopfield model, this model was in a focus of attention of theoretical physicists and computational neuroscientists.

In the previous section we studied a certain dynamics of activation on a random graph. Here we shall view the Hopfield model in the same manner: dynamics of activation on a given graph. In particular, we are interested in the properties of this model on a random graph.

For the convenience of a reader let us shortly remind the definition of the Hopfield model. The Hopfield model suggests a mechanism to retrieve a given set of patterns of activity in a network. Thus it models a memory function, one of principle tasks of the brain. Assume, that a network consists of N neurons, fully interconnected. Each neuron can be in one of two states: -1 or 1 which represent the main two levels of activity. Thus the entire state space of this network is $S_n=\{\sigma=(\sigma_1,\ldots,\sigma_n):\sigma_i\in\{-1,1\}\}$. Then the given set of binary patterns $\xi^1,\ldots,\xi^M\in S_n$ belongs (under certain conditions on value M) to the set of the local minima of the Hamiltonian

$$H_n(\sigma) = -\sum_{i,j=1}^{n} J_{ij}\sigma_i\sigma_j, \quad (27)$$

where $J_{ij} = \frac{1}{n}\sum_{\mu=1}^{M}\xi_i^\mu\xi_j^\mu$. The *storage capacity* of the model is the greatest number $M(N)$ such that the patterns ξ^μ are stable (or close to stable in some precise sense). Numerically it was shown already by Hopfield (1982) that there exists critical constant α_c such that for $M \le \alpha_c N$ the patterns are stable, while above this number this is not the case. Newman (1988) rigorously proved that for $M \le 0.055N$ there exist local minima near every original pattern surrounded by high energy barrier.

Numerous versions of Eq. (27) are still being introduced and analyzed. One particular direction is to relax the condition of full and symmetric connectivity of the original Hopfield model. Some of the first results in this area were obtained by physicists already in 80s: Sompolinsky (1986), Derrida et al. (1987) (highly dilute model), Treves and Amit (1988) (weakly dilute model) were among the first.

Let us now write the Hamiltonian for the Hopfield model on a network consisting of n neurons where a synaptic connection between each pair of neurons exists with probability p and is absent with probability $1-p$ as follows

$$H_n(\sigma,\varepsilon) = -\frac{1}{n}\sum_{i,j=1}^{n}\sum_{\mu=1}^{M}\varepsilon_{ij}\sigma_i\sigma_j\xi_i^\mu\xi_j^\mu, \tag{28}$$

where ε_{ij} are $i.i.d.Be(p)$ for different pairs (i, j).

With Hamiltonian (28) we associate a dynamics of activation on a *random* graph $G_{n,p}$, while with Hamiltonian (27) the dynamics on a *fully connected* graph is associated.

The question of how the storage capacity depends on the two parameters p and n, and in particular how small p is allowed to be taken to obtain a functioning memory, has been rigorously analyzed by Bovier and Gayrard (1992) (see also Treves and Rolls (1991) for more biologically justified computations). It has been proved by Bovier and Gayrard (1992) that when $p \ge c[(\ln n)/n]^{1/2}$, the model can store at least $m = \alpha_c pN$ patterns, where $\alpha_c \approx 0.027$ if $c \ge \sim 3$ and decreases proportionally to $1/(-\ln c)$ for c small. This generalizes the results of Newman (1988) for the standard Hopfield model Hopfield (1982). Although the upper bounds for the storage capacity were not given, Bovier and Gayrard (1992) provided also an argument that p cannot go below $1/\sqrt{n}$.

One can rephrase this question on p in a different way:

A: *What are the properties of the underlying (random) graph that are needed for a properly functioning memory?*

Bovier and Gayrard (1992) pointed out a rather obvious, but still an important partial answer to this question: it is not just the *connectedness* of the network which one needs here. Indeed, it is known (Bollobás, 1985) that a $G_{n,p}$ graph with $p = \ln n/n$ most likely consists of a single connected component. However, the critical threshold for the memory provided by Bovier and Gayrard (1992) is considerably higher, it is of order $1/\sqrt{n}$.

One may also ask therefore an inverse question:

B. *For which properties of $G_{n,p}$ the order $p \sim 1/\sqrt{n}$ is critical?*

The answers for the posed questions on the structure of the underlying graph might help neuromodeling by defining the most efficient (in some sense) structure.

Here we shall argue (and speculate) as follows: we shall give an answer for the question B, i.e., we describe the property of $G_{n,p}$ for which the order $p \sim 1/\sqrt{n}$ is critical, and then we can speculate that this property is necessary for a properly functioning memory, i.e., we give some answer to the question A.

To give the answer for the question B consider the process of bootstrap percolation (see the definition in Section 1) on random graph $G_{n,p}$ (n vertices and probability of each edge is p). When the final size of the activation is $n-o(n)$ we say that we observe 'almost percolation' in the network.

We shall use the following result on the critical value of p permitting almost percolation in $G_{n,p}$.

Theorem (Janson et al. (2010)). *Consider bootstrap percolation in $G_{n,p}$. Suppose that $r \ge 2$ and that $A(0) = a \to \infty$ as $n \to \infty$ but with $A(0) = a = o(n)$. Then the threshold for p for almost percolation is*

$$p_c = \left(\frac{(r-1)^{r-1}(r-1)!}{r^{r-1}}\right)^{1/r}\left(na^{r-1}\right)^{-1/r} \tag{29}$$

in the sense that if, for some $\delta > 0$, $p \le (1-\delta)p_c$ then the final size of the activation is at most $2A(0) = o(n)$, while if $p \ge (1+\delta)p_c$, then the final size of the activation is $n-o(n)$ with a high probability.

Consider the case when $r = 2$. This is the minimal (and hence, critical in some sense) value which leads to the sharp phase transition as we discussed in Section 1. The meaning of this parameter is the minimal number of the incoming impulses sufficient for a firing of a neuron in our model. (Of course, the real parameter for the neuromodeling should be much higher, e.g., 7–100, see e.g., Rolls (2008).) When $r = 2$ formula (29) tells us that if $p \ll 1/\sqrt{n}$ then the excitation which requires at least 2 incoming impulses does not cover more than $2A(0) = o(n)$ neurons, while for $p \gg 1/\sqrt{n}$ this excitation covers almost the entire network. Hence, the value $p \sim 1/\sqrt{n}$ is critical for the graph $G_{n,p}$ for the bootstrap percolation with the minimal parameter $r = 2$.

Thus, the order $1/\sqrt{n}$ for the critical value of p for properly functioning memory in the Hopfield model [8] coincides with the critical value p in Eq. (29) when $r = 2$, which permits almost percolation.

This suggests the following answer to the question A: the auto-associative memory functions properly, or more efficient on the graphs which permit bootstrap percolation.

Furthermore, if one assumes that the structure of the network which includes inhibitory neurons (as e.g., the model we studied here) is more biologically justified, it is an open interesting question to check whether the auto-associative memory in Hopfield model might be even more efficient on such networks. (This is also very much in lines of the research by Hertz and Prügel-Bennett (1996).) But this will be a subject of a separate study.

2.7. Conclusions

It has been argued here that the architecture of random networks is one of the determining parameters for the storage capacity. The study of the graph structure of the network helps to understand better the functioning of auto-associative memory.

Acknowledgments

The author thanks Alessandro Villa for bringing the author's attention to the problem investigated and helpful discussions

at the early stage of the project. The author is grateful to Edmund Rolls for inspiring and encouraging discussions on the project and related matters, including emphasizing the relation of the architecture to memory function.

The author thanks the referees for many useful suggestions.

REFERENCES

Aizenman, M., Lebowitz, J.L., 1988. Metastability effects in bootstrap percolation. J. Phys. A 21 (19), 3801–3813.

Amit, D.J., Gutfreund, H., Sompolinsky, H., 1987. Statistical mechanics of neural networks near saturation. Ann. Phys. 173, 30–67.

Balogh, J., Bollobás, B., 2006. Bootstrap percolation on the hypercube. Probab. Theory Relat. Fields 134 (4), 624–648.

Balogh, J., Pittel, B.G., 2007. Bootstrap percolation on the random regular graph. Random Struct. Algorithms 30 (1–2), 257–286.

Bienenstock, E., 1995. A model of neocortex. Network: Comput. Neural Syst. 6, 179–224.

Bollobás, B., 1985. Random Graphs. Academic Press.

Bourgeois, J., Rakic, P., 1993. Changes of synaptic density in the primary visual cortex of the macaque monkey from fetal to adult stage. J. Neurosci. 13, 2801–2820.

Bovier, A., Gayrard, V., 1992. Rigorous bounds on the storage capacity of the dilute Hopfield model. J. Stat. Phys. 69, 597–627.

Bovier, A., Picco, P. (Eds.), 1998. Mathematical Aspects of Spin Glasses and Neural Networks. : Progress in Probability, 41. Birkhäuser.

Cerf, R., Manzo, F., 2002. The threshold regime of finite volume bootstrap percolation. Stochastic Process. Appl. 101 (1), 69–82.

Chalupa, J., Leath, P.L., Reich, G.R., 1979. Bootstrap percolation on a Bethe lattice. J. Phys. C. 12, L31–L35.

Derrida, B., Gardner, E., Zippelius, A., 1987. An exactly solvable asymmetric neural network model. Europhys. Lett. 4, 167.

Eccles, J.C., 1952. The Neurophysiological Basis of Mind. Oxford University Press.

Eckmann, J.-P., Moses, E., Stetter, O., Tlusty, T., Zbinden, C., 2010. Leaders of neuronal cultures in a quorum percolation model. Front. Comput. Neurosci. 4, 132.

Harris, T.E., 1963. The Theory of Branching Processes. Springer.

Hertz, J., Prügel-Bennett, A., 1996. Learning short synfire chains by self-organization. Network: Comput. Neural Syst. 7 (2), 357–363.

Holroyd, A.E., 2003. Sharp metastability threshold for two-dimensional bootstrap percolation. Probab. Theory Relat. Fields 125 (2), 195–224.

Hopfield, J.J., 1982. Neural networks and physical systems with emergent collective computational abilities. Proc. Natl. Acad. Sci. U. S. A. 79, 2554–2558.

Huttenlocher, P.R., 1979. Synaptic density in human frontal cortex — developmental changes and effects of aging. Brain Res. 163, 195–205.

Huttenlocher, P.R., de Courten, C., Garey, L.J., Van der Loos, H., 1982. Synaptogenesis in human visual cortex — evidence for synapse elimination during normal development. Neurosci. Lett. 33, 247–252.

Iglesias, J., Villa, A.E., 2007. Effect of stimulus-driven pruning on the detection of spatiotemporal patterns of activity in large neural networks. Biosystems 89, 287–293.

Iglesias, J., Villa, A.E., 2008. Emergence of preferred firing sequences in large spiking neural networks during simulated neuronal development. Int. J. Neural Syst. 18, 267–277.

Iglesias, J., Eriksson, J., Grize, F., Tomassini, M., Villa, A.E.P., 2005. Dynamics of pruning in simulated large-scale spiking neural networks. Biosystems 79, 11–20.

Janson, S., Luczak, T., Turova, T., Vallier, T., 2010. Bootstrap Percolation on the Random Graph $G_{n,p}$. arXiv:1012.3535 [math. PR]2010.

Kozma, R., Puljic, M., Balister, P., Bollobás, B., Freeman, W., 2005. Phase transitions in the neuropercolation model of neural populations with mixed local and non-local interactions. Biol. Cybernet. 92 (6), 367–379.

Montgomery, J., Madison, D., 2004. Discrete synaptic states define a major mechanism of synapse plasticity. Trends Neurosci. 27, 744–750.

Newman, Ch.M., 1988. Memory capacity in neural network models: rigorous lower bounds. Neural Netw. 1, 223–238.

Pastur, L.A., Figotin, A.L., 1978. On the theory of disordered spin systems. Theor. Math. Phys. 35, 403–414.

Puljic, M., Kozma, R., 2005. Activation clustering in neural and social networks. Complexity 10 (4), 42–50.

Rolls, E.T., 2008. Memory, Attention, and Decision-Making. A Unifying Computational Neuroscience Approach. Oxford University Press, Oxford.

Sompolinsky, H., 1986. Neural networks with nonlinear synapses and a static noise. Phys. Rev. A 34 (3).

Talagrand, M., 1998. Rigorous results for the Hopfield model with many patterns. Probab. Theory Relat. Fields 110 (2), 177–276.

Tlusty, T., Eckmann, J.-P., 2009. Remarks on bootstrap percolation in metric networks. J. Phys. A: Math. Theor. 42, 205004.

Treves, A., 1991. Dilution and sparse coding in threshold-linear nets. J. Phys. A: Math. Gen. 24 (1), 327335.

Treves, A., Amit, D.J., 1988. Metastable states in asymmetrically diluted Hopfield networks. J. Phys. A: Math. Gen. 21, 31–55.

Treves, A., Rolls, E.T., 1991. What determines the capacity of autoassociative memories in the brain? Network 2, 371397.

Turova, T., Villa, A., 2007. On a phase diagram for random neural networks with embedded spike timing dependent plasticity. Biosystems 89, 280–286.

Vallier, T. (2007) Random graph models and their applications, PhD Thesis in Mathematics, Lund University, Sweden.