Journal of Statistical Physics, Vol. 66, Nos. 5/6, 1992

A Lower Bound for the Memory Capacity in the Potts–Hopfield Model

Pablo A. Ferrari,[1] Servet Martínez,[2] and Pierre Picco[3]

Received October 3, 1991

We consider Potts–Hopfield networks of size N. We prove the result: $\exists \alpha_c > 0$ such that for all $0 < \alpha < \alpha_c$ we can find δ, $\varepsilon > 0$ in such a way that, when $N \to \infty$, we can store αN patterns, all of them being sorrounded by ε-energy barriers at distance δ.

KEY WORDS: Neural network; Hopfield model; Potts model.

INTRODUCTION

Potts–Hopfield networks correspond to q-state neural networks. They have been recently used to model some smoothing techniques and coloring rules (see, for instance, ref. 6). We consider the problem of the memory capacity. We remark that we are storring colored patterns with q colors instead of black and white patterns as would be the case in the Hopfield model.

In our main theorem we consider the relation between local minima of the energy function and the stored patterns. We prove an analog of the result shown by Newman for the Hopfield model.[(1)] To describe it, let us consider a network of q-state neurons, of size N. We prove that there exists $\alpha_c > 0$ such that for any $0 < \alpha < \alpha_c$ we can find $\varepsilon > 0$ and $0 < \delta < 1/2$ such that, in the limit $N \to \infty$, we can store αN patterns, all of them being sorrounded by ε-energy barriers at distance δ.

[1] Instituto de Matemática e Estadística, Universidad de São Paulo, Cx. Postal 20570, 01498 São Paulo, Brasil.

[2] Departamento de Ingeniería Matemática, Universidad de Chile, Casilla 170, Correo 3, Santiago 3, Chile.

[3] Laboratoire Propre, LP7061, Centre de Physique Théorique CNRS, Luminy, Case 907F-13288, Marseille, Cedex 9, France.

Our proof follows that of Newman. But some interesting technical points must be solved.

In Section 1 we describe the model and we show that it reduces to the Hopfield networks when $q=2$. We also code the Hamiltonian as well as the dynamics with a 2-state neural network and at the end of this section we state the main result. The proofs are in Section 2.

A complete discussion of Potts model can be found in ref. 7.

1. DESCRIPTION OF THE MODEL AND MAIN RESULT

The model has N neurons whose configurations are described by an N-vector $\sigma = (\sigma_1,..., \sigma_N)$, where for any $i \in \{1,..., N\}$, $\sigma_i \in \{0, 1,..., q-1\} = Q$, and q is a positive integer $\geqslant 2$.

There are $m \in \mathbb{N}$ patterns $(\xi^\mu)_{\mu=1}^m$ to be stored, each of which is an N-vector $\xi^\mu = (\xi_1^\mu,..., \xi_N^\mu)$, where $\forall \mu \in \{1,..., m\}$, $\forall i \in \{1,..., N\}$, $\xi_i^\mu \in Q$. We consider the random case, which corresponds to choosing (ξ_i^μ) i.i.d. with $\mathbb{P}(\xi_i^\mu = s) = 1/q \ \forall s \in Q$.

Given the patterns ξ^μ, we define the energy

$$H(\sigma) = -\sum_{\mu=1}^m \left[\sum_{i=1}^N \left\{\delta(\xi_i^\mu, \sigma_i) - \frac{1}{q}\right\}\right]^2 \qquad \text{for} \quad \sigma \in Q^N \tag{1.1}$$

where $\delta(\cdot, \cdot)$ is the Kronecker symbol.

Let us associate a possible sequential dynamics compatible with the Hamiltonian. Assume we have updated the ith neuron of the configuration $\sigma = (\sigma_1,..., \sigma_N)$ and call $\sigma' = (\sigma'_1,..., \sigma'_N)$ the new configuration. So

$$\sigma'_j = \sigma_j \qquad \text{for any} \quad j \neq i$$

It can be easily shown that

$$\begin{aligned} H(\sigma') - H(\sigma) = &-2 \sum_{\mu=1}^m \delta(\xi_i^\mu, \sigma'_i) \left\{\left[\sum_{\substack{j=1\\ j\neq i}}^N \left(\delta(\xi_j^\mu, \sigma_j) - \frac{1}{q}\right)\right] + \frac{1}{2} - \frac{1}{q}\right\} \\ &+2 \sum_{\mu=1}^m \delta(\xi_i^\mu, \sigma_i) \left\{\sum_{\substack{j=1\\ j\neq i}}^N \left(\delta(\xi_j^\mu, \sigma_j) - \frac{1}{q}\right)\right] - \frac{1}{q}\right\} \end{aligned}$$

Now define $\sigma'_i = s$, where s is any one of the elements s' of Q, which is a maximizer in the following formula:

$$\max_{s' \in Q} \sum_{\mu=1}^m \delta(\xi_i^\mu, s') \left\{\left[\sum_{\substack{j=1\\ j\neq i}}^N \left(\delta(\xi_j^\mu, \sigma_j) - \frac{1}{q}\right)\right] + \frac{1}{2} - \frac{1}{q}\right\} \tag{1.2}$$

Hence H is a Lyapunov function of this dynamics, that is, it decreases with the evolution. In particular, starting with some initial configuration $\sigma(0)$, the trajectory will follow the decreasing part of the graph of the function $\sigma \to H(\sigma)$, and will eventually reach a local minimum, where it will stop.

To analyze the expressions (1.1) and (1.2) in the case $q=2$, transform the $\{0, 1\}$ variables σ_i, ξ_i^μ into the $\{-1, 1\}$ variables $\tilde{\sigma}_i = 2\sigma_i - 1$, $\tilde{\xi}_i^\mu = 2\xi_i^\mu - 1$. It is easily shown that $H(\sigma)$ given by (1.1) becomes

$$H(\tilde{\sigma}) = -\frac{1}{4} \sum_{\mu=i}^{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \tilde{\xi}_i^\mu \tilde{\xi}_j^\mu \tilde{\sigma}_i \tilde{\sigma}_j$$

which is the usual Lyapunov functional.[1] Now the dynamics (1.2) is written into the new variables as

$$\tilde{\sigma}_i = \operatorname{sign}\left\{ \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \tilde{\xi}_i^\mu \tilde{\xi}_j^\mu \tilde{\sigma}_j \right\}$$

which is the sequential dynamics of the Hopfield model.[1]

Furthermore, the general Potts–Hopfield model given by Hamiltonian (1.1) can be coded like a Hopfield network of size qN, the states of each neuron being $\{0, 1\}$. In this purpose, associate to ξ^μ and σ the following qN-vectors:

$$\hat{\xi}_{i,s}^\mu = \delta(\xi_i^\mu, s), \qquad \hat{\sigma}_{i,s} = \delta(\sigma_i, s)$$

Hence the Hamiltonian (1.1) becomes

$$H(\hat{\sigma}) = -\sum_{\mu=1}^{m} \sum_{i=1}^{N} \sum_{s=1}^{q} \sum_{j=1}^{N} \sum_{u=1}^{q} \hat{\xi}_{i,s}^\mu \hat{\xi}_{j,u}^\mu \hat{\sigma}_{i,s} \hat{\sigma}_{j,u} + 2N \sum_{\mu=1}^{m} \sum_{i=1}^{N} \sum_{s=1}^{q} \hat{\xi}_{i,s}^\mu \hat{\sigma}_{i,s} - mN^2$$

This corresponds to a Hamiltonian associated to the following sequential dynamics,[6] which we write for the updating of the (i, s) neuron:

$$(\hat{\sigma}_{is})' = \mathbb{1}\left(\sum_{\mu=1}^{m} \sum_{j=1}^{N} \sum_{u=1}^{q} \xi_{i,s}^\mu \xi_{j,u}^\mu \hat{\sigma}_{j,u} - N\xi_{i,s}^\mu \right)$$

where $\mathbb{1}(v)$ is the threshold function, $\mathbb{1}(v) = 1$ if $v \geqslant 0$, $=0$ if $v < 0$. This is similar to the iteration of a $\{0, 1\}$ Hopfield network with thresholds $N\xi_{i,s}^\mu$. The difference from the Hopfield model is that the set of variables $\{\hat{\xi}_{i,s}^\mu : \mu = 1,..., m;\ i = 1,..., N;\ s = 1,..., q\}$ are not independent; in fact, the relations $\sum_{s=1}^{q} \xi_{i,s}^{\mu}=1$ must be verified for each pair (i, μ). Then the statistical properties of the Potts–Hopfield model are not simple extensions of the Hopfield case.

We want to study the relation between the local minima of the energy $H(\sigma)$ and the stored patterns $\{\xi^{\mu}: \mu=1,..., m\}$ choosen at random with uniform distribution.

In the Hopfield case $(q=2)$, if $m=N/(c \log N)$ for some real number c, it is proved that the original patterns are local minima.[4,5] The case $m=\alpha N$ was studied by Amit *et al.*[3] using a nonrigorous argument that strongly suggests the presence of a threshold α_c such that if $0<\alpha<\alpha_c(\alpha_c \sim 0.15)$, the Hamiltonian has minima that are near the patterns, while if $\alpha>\alpha_c$, this is not the case.

An important rigorous result in this direction was made by Newman,[2] who proved that $\exists \alpha>0$ such that if $\alpha<\alpha$, then for almost all realizations of random patterns there exists, around each pattern, an energy barrier. Let us be more precise.

Call $d(\sigma, \sigma')$ the Hamming distance between two configurations:

$$d(\sigma, \sigma')=\sum_{i=1}^{N} [1-\delta(\sigma_i, \sigma_i')] \tag{1.3}$$

that is, the number of sites where the two configurations disagree. For any configuration σ and real number $\delta \in [0, 1]$, denote by

$$S(\sigma, \delta)=\{\sigma': d(\sigma, \sigma')=[\delta N]\}$$

the sphere of radius $[\delta N]$.

Call $\underline{h}(\sigma, \delta)=\min\{H(\sigma'): \sigma' \in S(\sigma, \delta)\}$. We shall say that there exists an energy barrier at distance δ around a pattern ξ^{μ} if for some $\varepsilon>0$

$$H(\xi^{\mu})+\varepsilon N^2<\underline{h}(\xi^{\mu}, \delta) \tag{1.4}$$

We remark that the presence of an energy barrier at distance δ around ξ^{μ} does not imply that ξ^{μ} is a local minimum, but merely that at least one minimum is inside a ball of radius $[\delta N]$ centered at ξ^{μ}. If we accept the idea of storing patterns that could be retrievied with errors, the notion of energy barriers is a good one.

The main result of this paper is the extension to q-state neural networks of the Newman result for $q=2$.

Theorem. Given $q \geqslant 2$, there exists an $\underline{\alpha}_c=\underline{\alpha}_c(q)>0$ such that for any $\alpha \leqslant \underline{\alpha}_c$, we can find $\delta \in (0, 1/2)$ and an $\varepsilon>0$ such that, in the limit

$N \to \infty$, for almost all configurations of patterns, all patterns are surrounded by energy barriers at distance δ.

Precisely,

$$\mathbb{P}\left(\bigcup_{K=1}^{\infty} \bigcap_{N \geqslant K} \{\forall \mu \in \{1,..., \alpha N\}; \forall \sigma \in S(\xi^{\mu}, \delta); H(\sigma) > H(\xi^{\mu}) + \varepsilon N^2\} \right) = 1 \tag{1.5}$$

2. PROOF OF THE THEOREM

Let us first remark that, given $\mu \in \{1,..., \alpha N\}$ and a pattern ξ^{μ}, the set of configurations σ such that $d(\xi^{\mu}, \sigma) = [\delta N]$ can be obtained by choosing first the subset $J \subset \{1,..., N\}$, where σ and ξ^{μ} disagree, and then $\sigma_j \neq \xi_j^{\mu}$ $\forall j \in J$. In particular,

$$|S(\xi^{\mu}, \delta)| = \binom{N}{[\delta N]} (q-1)^{[\delta N]} \tag{2.1}$$

From (2.1) we get

$$\begin{aligned} &\mathbb{P}(\exists \mu \subset \{1,..., \alpha N\}, \exists \sigma \in S(\xi^{\mu}, \delta); H(\sigma) \leqslant H(\xi^{\mu}) + \varepsilon N^2) \\ &\quad \leqslant \alpha N \binom{N}{[\delta N]} (q-1)^{[\delta N]} \, \mathbb{P}(H(\xi_J^{\mu}) - H(\xi^{\mu}) \leqslant \varepsilon N^2) \end{aligned} \tag{2.2}$$

where ξ_J^{μ} is any pattern in $S(\xi^{\mu}, \delta)$; one can choose $J = \{1, 2,..., [\delta N]\}$ and μ is any index $\in \{1,..., \alpha N\}$.

In order to estimate the probability in the right-hand side of (2.2), we introduce some notations.

Let

$$V_1 = 2 \frac{\delta}{q}$$

$$V_2 = 4(1-\delta)\left(\frac{1}{q}\right)\left(1 - \frac{1}{q}\right) + 2\delta \left(\frac{1}{q}\right)\left(1 - \frac{2}{q}\right)$$

$$\gamma = \delta \left[(1-\delta) + \left(1 - \frac{2}{q}\right) \right]$$

Define the variables

$$Z_{1,J}^{\mu'} = \sum_{i \in J} \frac{\delta(\xi_i^{\mu'}, (\xi_J^{\mu})_i) - \delta(\xi_i^{\mu'}, \xi_i^{\mu})}{(N V_1)^{1/2}} \tag{2.3}$$

$$Z^{\mu'}_{2,J} = \sum_{i \in J} \frac{\delta(\xi^{\mu'}_i, (\xi^{\mu}_J)_i) + \delta(\xi^{\mu'}_i, \xi^{\mu}_i) - 2/q}{(NV_2)^{1/2}} \tag{2.4}$$

$$Z^{\mu'}_{2,J^c} = \sum_{i \in J^c} \frac{2(\delta(\xi^{\mu'}_i, \xi^{\mu}_i) - 1/q)}{(NV_2)^{1/2}} \tag{2.5}$$

Using the identity $x^2 - y^2 = (x-y)(x+y)$, it is not difficult to check that

$$H(\xi^{\mu}_J) - H(\xi^{\mu}) = -(V_1 V_2)^{1/2} N \sum_{\substack{\mu'=1 \\ \mu' \neq \mu}}^{\alpha N} Z^{\mu'}_{1,J}(Z^{\mu'}_{2,J} + Z^{\mu'}_{2,J^c}) + N^2 \gamma \tag{2.6}$$

In particular, the event which appears in the right-hand side of (2.2) is

$$B_N \equiv \left\{ u \leqslant \frac{1}{\alpha N} \sum_{\substack{\mu'=1 \\ \mu' \neq \mu}}^{\alpha N} Z^{\mu'}_{1,J}(Z^{\mu'}_{2,J} + Z^{\mu'}_{2,J^c}) \right\} \tag{2.7}$$

where $u = (\gamma - \varepsilon)/\alpha(V_1 V_2)^{1/2}$.

Now given μ and a fixed realization of ξ^{μ}, it is easy to check that the family

$$\{Z^{\mu'}_{1,J}(Z^{\mu'}_{2,J} + Z^{\mu'}_{2,J^c})\}^{\alpha N}_{\substack{\mu'=1 \\ \mu' \neq \mu}}$$

is a family of i.i.d. random variables.

Remark. We emphasize that, given ξ^{μ}, the two random variables $(Z^{\mu'}_{1,J}, Z^{\mu'}_{2,J})$ are not independent, but $Z^{\mu'}_{1,J}$ and $Z^{\mu'}_{2,J}$ are independent of $Z^{\mu'}_{2,J^c}$. This phenomenon occurs only if $q > 2$; if $q = 2$, it can be verified that $Z^{\mu'}_{2,J} = 0$ $\forall \mu' \in \{1,..., \alpha N\}$. This is the main difference between the two cases.

It is easy to show that $\mathbb{E}(Z^{\mu'}_{1,J} Z^{\mu'}_{2,J}) = 0$ [in fact, it follows from expression (2.11) below]; then B_N is the event of a large deviation from the mean; it is natural to use the exponential Markov inequality to get

$$\mathbb{P}(B_N) \leqslant [\inf_{t \geqslant 0} e^{-tu} \mathbb{E}(e^{tZ_{1,J}(Z_{2,J} + Z_{2,J^c})})]^{\alpha N} \tag{2.8}$$

where $(Z_{1,J}; Z_{2,J}; Z_{2,J^c})$ have the same distribution as any given $(Z^{\mu'}_{1,J}; Z^{\mu'}_{2,J}; Z^{\mu'}_{2,J^c})$ for some $\mu' \neq \mu$.

Calling $\mathbb{E}_A$ for $A = J$ or $A = J^c$ the expectation with respect to the random variables $(\xi^{\mu'}_i)(\xi^{\mu}_i)$ for $i \in A$, and taking into account the previous remark, we get

$$\mathbb{E}(e^{tZ_{1,J}(Z_{2,J} + Z_{2,J^c})}) = \mathbb{E}_J(e^{tZ_{1,J}Z_{2,J}} \mathbb{E}_{J^c}(e^{tZ_{1,J}Z_{2,J^c}})) \tag{2.9}$$

Now looking at (2.5), it is easy to check that

$$\mathbb{E}_{J^c}(e^{tZ_{1,J}Z_{2,J^c}}) = \sum_{n=0}^{|J^c|} \binom{|J^c|}{n}\left(\frac{1}{q}\right)^n \left(1-\frac{1}{q}\right)^{|J^c|-n} \times \exp Z_{1,J} \frac{2t}{(NV_2)^{1/2}}\left(n-\frac{|J^c|}{q}\right) \tag{2.10}$$

Now in order to perform the expectation with respect to $\mathbb{E}_J$ we need the explicit form of the joint distribution of $(Z_{1,J}, Z_{2,J})$. Assume for the moment that for any function $f\colon \mathbb{R}^2 \to \mathbb{R}$,

$$\begin{aligned}&\mathbb{E}_J(f(Z_{2,J}, Z_{1,J}))\\ &= \sum_{K=0}^{|J|}\sum_{l=-K}^{+K}\binom{J}{K}\binom{K}{(K+l)/2}\left(1-\frac{2}{q}\right)^{|J|-K}\left(\frac{2}{q}\right)^K\left(\frac{1}{2}\right)^K\\ &\quad\times f\left(\frac{K-2\,|J|/q}{(NV_2)^{1/2}}, \frac{l}{(NV_1)^{1/2}}\right)\end{aligned} \tag{2.11}$$

Therefore we get

$$\begin{aligned}&\mathbb{E}(e^{tZ_{1,J}(Z_{2,J}+Z_{2,J^c})})\\ &= \sum_{K=0}^{|J|}\sum_{l=-K}^{+K}\sum_{n=0}^{|J^c|}\binom{|J|}{K}\binom{K}{(K+l)/2}\binom{|J^c|}{n}\left(1-\frac{2}{q}\right)^{|J|-K}\left(\frac{2}{q}\right)^K\left(\frac{1}{2}\right)^K\\ &\quad\times\left(1-\frac{1}{q}\right)^{|J^c|-n}\left(\frac{1}{q}\right)^n \exp\frac{tl}{(NV_1)^{1/2}}\left(\frac{(K-2J/q)+2(n-|J^c|/q)}{(V_2N)^{1/2}}\right)\end{aligned} \tag{2.12}$$

Let us now check (2.11).

For each $K \in 0,..., |J|$, we can choose A, $|A| = |J| - K$, such that, for $i \in A$,

$$\delta(\xi_i^{\mu'}, (\xi_J^{\mu})_i) + \delta(\xi_i^{\mu'}, \xi_i^{\mu}) = 0$$

and for $i \in J\backslash A$

$$\delta(\xi_i^{\mu'}, (\xi_J^{\mu})_i) + \delta(\xi_i^{\mu'}, \xi_i^{\mu}) = 1$$

This choice can be done in $\binom{|J|}{K}$ different ways and the probability of any choice is $(1-2/q)^{|J|-K}(2/q)^K$.

Now for fixed K and A, we choose on $J\backslash A$ a subset B of cardinal $(K+l)/2$, where for any $j \in B$,

$$\delta(\xi_i^{\mu'}, (\xi_J^{\mu})_i) - \delta(\xi_i^{\mu'}, \xi_i^{\mu}) = 1$$

and for any $j \in (J \backslash A) \backslash B$,

$$\delta(\xi_i^{\mu'}, (\xi_J^{\mu})_i) - \delta(\xi_i^{\mu'}, \xi_i^{\mu}) = -1$$

We have $\binom{K}{(K+l)/2}$ choices of such subset B and each choice has probability $(1/2)^K$. Note that

$$\sum_{i \in J \backslash A} \delta(\xi_i^{\mu'}, (\xi_J^{\mu})_i) - \delta(\xi_i^{\mu'}, \xi_i^{\mu}) = l$$

In order to estimate (2.12), we will use the two following lemmas.

Lemma 2.1. For any real number λ and any integer $K \geqslant 0$

$$\sum_{l=-K}^{+K} \binom{K}{(K+l)/2} \left(\frac{1}{2}\right)^k e^{\lambda l} \leqslant e^{\lambda^2 K/2} \tag{2.13}$$

Proof. This follows from

$$(\cosh \lambda)^K \leqslant \exp \frac{\lambda^2}{2} K$$

Lemma 2.2. Let $S_n = \sum_{i=1}^n X_i$, where $X_i = 0$ with probability p and $X_i = 1$ with probability $1 - p = r$. Then for any real number λ

$$\mathbb{E}(e^{\lambda(S_n - \mathbb{E}(S_n))}) \leqslant e^{(\lambda^2/2)n} \tag{2.14}$$

Proof. Using the Jensen inequality, we get, for any $i \in \{1,..., n\}$,

$$\mathbb{E}(e^{\lambda(X_i - \mathbb{E}(X_i))}) \leqslant \mathbb{E}(e^{\lambda(X_i - X_i')}) \tag{2.15}$$

when X_i' have the same distribution as X_i and is independent of X_i. Remark that $X_i - X_i' = \pm 1$ with probability rp and $X_i - X_i' = 0$ with probability $p^2 + r^2$. Therefore

$$\mathbb{E}(e^{\lambda(X_i - X_i')}) = p^2 + r^2 + 2pr \left(\frac{e^{\lambda} + e^{-\lambda}}{2}\right) \leqslant \cosh \lambda \leqslant \exp \frac{\lambda^2}{2} \quad \blacksquare$$

Now using Lemma 2.1, we get, for any $K \geqslant 0$,

$$\begin{aligned}
&\sum_{l=-K}^{+K} \binom{K}{(K+l)/2} \left(\frac{1}{2}\right)^K \exp \frac{tl}{(NV_1)^{1/2}} \left(\frac{(K - 2\,|J|/q) + 2(n - |J^c|/q)}{(V_2 N)^{1/2}}\right) \\
&\leqslant \exp \frac{t^2}{2} \frac{K}{NV_1} \left(\frac{(K - 2\,|J|/q) + 2(n - |J^c|/q)}{(V_2 N)^{1/2}}\right)^2 \\
&\leqslant \mathbb{E}\left[\exp\left(\frac{t\sqrt{\delta}}{\sqrt{V_1}} Z\right)\right] \frac{(K - 2\,|J|/q) + 2(n - |J^c|/q}{(V_2 N)^{1/2}}\Bigg)
\end{aligned} \tag{2.16}$$

where Z is a normalized Gaussian random variable and we have used $K/N \leqslant \delta$.

Now inserting (2.16) in (2.12) and using Lemma 2.2 with $p_1 = (1-2/q)$ and $p_2 = (1-1/q)$, we get

$$\mathbb{E}(e^{tZ_{1,J}(Z_{2,J}+Z_{2,J^c})}) \leqslant \mathbb{E}\left(\exp \frac{t^2}{2} \frac{\delta Z^2}{V_1 V_2} \left(\frac{|J| + 4\,|J^c|}{N}\right)\right)$$

$$= \left(1 - t^2 \frac{\delta(\delta + 4(1-\delta))}{V_1 V_2}\right)^{-1/2} \tag{2.17}$$

Calling $v = \delta(\delta + 4(1-\delta))/V_1 V_2$, we have

$$\inf_{t \geqslant 0} e^{-tu} \frac{1}{(1-t^2 v)^{1/2}} = e^{\Gamma(u,v)} \tag{2.18}$$

where

$$\Gamma(u, v) = \frac{1}{2}\left\{\left(1 + \frac{4u^2}{v}\right)^{1/2} - 1 + \log \frac{v}{u^2}\left[\left(1 + \frac{4u^2}{v}\right)^{1/2} - 1\right]\right\} \tag{2.19}$$

Therefore, using (2.19), (2.18), (2.8), and (2.2), if we prove that

$$-\delta \log \delta - (1-\delta)\log(1-\delta) + \delta \log(q-1) - \alpha\Gamma(u, v) < 0 \tag{2.20}$$

then using the Borel–Cantelli lemma, we can finish the proof of the theorem.

Recall

$$u = \frac{\delta[(1-\delta) + (1-2/q)] - \varepsilon}{\alpha[(2\delta/q)\, V_2]^{1/2}}$$

and $\lim_{\delta \to 0} V_2 = 4(1/q)(1-1/q)$. Choose for convenience $\varepsilon = \delta(1-2/q)$; then

$$u \sim \frac{\sqrt{\delta}}{\alpha} C_1(q) \xrightarrow[\alpha \to 0]{} \infty$$

for some constant $C_1(q)$.

Now, it can be checked that $\lim_{u \to \infty} [\Gamma(u, v)/u] = 1/\sqrt{v}$; therefore $\alpha\Gamma(u, v) \sim \sqrt{\delta}\, C_2(q)$ for some constant $C_2(q)$.

Now

$$\frac{C_2(q)\sqrt{\delta}}{-\delta \log \delta - (1-\delta)\log(1-\delta) + \delta \log(q-1)}$$

$$\sim \frac{C_2(q)}{-\sqrt{\delta}\log \delta + \sqrt{\delta}\log(q-1)} \xrightarrow[\delta \to 0]{} +\infty$$

This proves (2.20), choosing first α small enough and then δ.

Remark. Making an explicit computation, if we are interested in q very large, it can be checked that as $q \to \infty$, $C_1(q) \sim [(q\delta)^{1/2}/\alpha]\, C_1$, $C_2(q) \sim (q/\delta)^{1/2}$ (C_2 and therefore

$$\frac{(\delta/q)^{1/2}}{-\delta \log \delta - (1-\delta)\log(1-\delta) + \delta \log(q-1)}$$
$$\sim \frac{\sqrt{q}}{-\sqrt{\delta}\log\delta + \sqrt{\delta}\log(q-1)} \to +\infty$$

by choosing an appropriate δ depending on q.

At this point we remark that in the limit $q \to \infty$ the estimate (2.14) is too poor to take seriously when one considers what happens when $q \to \infty$.

ACKNOWLEDGMENTS

The authors are thankful for interesting discussions with E. Goles. S.M. and P.P. are indebted for the very kind hospitality of the Instituto de Matemática e Estatística, Universidade de São Paulo. This work was partially supported by Fundação de Amparo à Pesquisa de Estado de São Paulo. S.M. was partially financed by Fondecyt grant 1208-91, Chile.

REFERENCES

1. J. J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, *Proc. Natl. Acad. Sci. USA* **79**:2554–2558 (1982).
2. C. M. Newman, Memory capacity in neural networks models: Rigorous lower bounds, *Neural Networks* **1**:223–238 (1988).
3. D. J. Amit, H. Gutfreund, and H. Sompolinsky, Spin-glass models of neural networks, *Phys. Rev. A* **32**:1007–1018 (1985); Storing infinite numbers of patterns in a spin-glass model of neural networks, *Phys. Rev. Lett.* **55**:1530–1533 (1985).
4. R. J. McEliece, E. C. Posner, and E. R. Rodemich, and S. S. Venkateshe, The capacity of the Hopfield associative memory, *IEEE Trans. Inf. Theory* **33**:461–482 (1987).
5. G. Weisbuch and F. Fogelman-Soulié, Scaling laws for the attractors of Hopfield networks, *J. Phys. Lett.* (Paris) **46**:L-623–L-630 (1985).
6. E. Goles and S. Martínez, Neural and automata networks, *Mathematics and Its Applications* (Kluwer, 1990).
7. Y. F. Wu, The Potts model, *Rev. Mod. Phys.* **54**(1):235–315 (1982).

Communicated by Anna Demasi