Capacity of an associative memory model on random graph architectures

MATTHIAS LÖWE[1] and FRANCK VERMET[2]

[1]*Fachbereich Mathematik und Informatik, Universität Münster, Einsteinstraße 62, 48149 Münster, Germany. E-mail: maloewe@math.uni-muenster.de*
[2]*Laboratoire de Mathématiques, UMR CNRS 6205, Université de Bretagne Occidentale, 6 avenue Victor Le Gorgeu, CS 93837, F-29238 Brest Cedex 3, France. E-mail: franck.vermet@univ-brest.fr*

We analyze the storage capacity of the Hopfield models on classes of random graphs. While such a setup has been analyzed for the case that the underlying random graph model is an Erdös–Renyi graph, other architectures, including those investigated in the recent neuroscience literature, have not been studied yet. We develop a notion of storage capacity that highlights the influence of the graph topology and give results on the storage capacity for not too irregular random graph models. The class of models investigated includes the popular power law graphs for some parameter values.

Keywords: associative memory; Hopfield model; powerlaw graphs; random graphs; random matrix; spectral theory; statistical mechanics

1. Introduction

Thirty years ago, in 1982, Hopfield introduced a toy model for a brain that renewed the interest in neural networks and has nowadays become popular under the name Hopfield model [18]. This model in its easiest version assumes that the neurons are fully connected and have Ising-type activities, that is, they take the values $+1$, if a neuron is firing and -1, if it is not, and is based on the principles of statistical mechanics. Since Hopfield's ground-breaking work, it has stimulated a large number of researchers from the areas of computer science, theoretical physics and mathematics.

In the latter field, the Hopfield model is particularly challenging, since it also can be considered as a spin glass model and spin glasses are notoriously difficult to study. A survey over the mathematical results in this area can be found in either [4] or [37]. It is worth mentioning that even in the parameter region where no spin glass phase is expected, the Hopfield model still has to offer surprising phenomena such as in [17].

When being considered as a neural network, one of the aspects that have been discussed most intensively is its so-called storage capacity. Here, one tries to store information, so-called patterns in the model, and the question is, how many patterns can be successfully retrieved by the network dynamics, that is, how much information can be stored in a model of N neurons. One of the early mathematical results states that if the patterns are independent and identically distributed (i.i.d. for short) and consist of i.i.d. spins and if their number M is bounded by $\frac{1}{2}N/\log N$, the patterns can be recalled (see [31]) with probability converging to one as $N \to \infty$ and that the constant $\frac{1}{2}$ is optimal (see [3]). Similar results hold true, if one starts with a corrupted input – if more than fifty

percent of the input spins are correct, one still is able to restore the originally "learned" patterns. However, if one also allows for small errors in the retrieval of the patterns one obtains a storage capacity of $M = \alpha N$ for some value of α smaller than 0.14 (see [24,32,36]). This latter result is in agreement with both, computer simulations as well as the predictions of the non-rigorous replica method from statistical physics (see [1]).

The setup of the Hopfield model has been generalized in various aspects, for example, the condition of the independence has been relaxed (see [25,27]), patterns with more than two spins values have been considered (see [15,26,27]), and Hopfield models on Erdös–Renyi graphs were studied [5,6,29,36]. The present paper starts with the observation that even though being more general than the complete graph, also Erdös–Renyi graphs do not seem to be the favorite architectures for a brain for scientists working in neurobiology. There, the standard paradigm currently is rather to model the brain as a small world graph (see [2,35]). We will focus on the question, how many patterns can be stored in a Hopfield model on a random graph, if this graph is no longer necessarily an Erdös–Renyi graph. The classical notion of storage capacity requires that the patterns are fixed points of the retrieval dynamics, that is, local minima of the energy landscape of the Hopfield model (or, in [24,32,36], not too far apart from such minima). It turns out that this notion is already sensitive to the architecture of the network [29]. So it is conceivable that there is a major influence of the underlying graph structure on the model's capability to retrieve corrupted information. Associativity of a network can be described as the potential to repair corrupted information. We will therefore work with a notion of storage capacity that takes this ability into account. Moreover, the relationship between network connectivity and the performance of associative memory models has already been investigated in computer simulations (see, e.g., [8]). Therefore, the goal of the present note is to establish rigorous bounds on the storage capacity of the Hopfield model on a wide class of random graph models, where we interpret "storage" as the ability to retrieve corrupted information. Similar questions have been addressed for the complete graph by Burshtein [7].

We organize the paper in the following way: Section 2 introduces the basic model we will be working with in the present paper. It also addresses the question, what exactly we mean when talking about the storage of patterns. Section 3 contains the main result of this paper. The number of patterns one is able to store in the sense, that a number of errors that is proportional to N can be repaired by $\mathcal{O}(\log N))$ steps of the retrieval dynamics is of order const.$(\lambda_1)^2/(m \log N)$, where λ_1 is the largest eigenvalue of the adjacency matrix of the graph and m its maximal degree. A main ingredient of the proof is thus to analyze the spectrum of the adjacency matrix of the graph that serves as a model of the network architecture. This analysis is provided in Section 4. Eventually, Section 5 contains the proof of the main result. An Appendix will contain estimates on the minimum and maximum degree of an Erdös–Rényi graph. These are needed to apply our main result to the setting of such random graphs and may also be of independent interest.

2. The model

The Hopfield model is a spin model on $N \in \mathbb{N}$ spins. $\sigma \in \Sigma_N := \{-1, +1\}^N$ describes the neural activities of N neurons. The information to be stored in the model are patterns $\xi^1, \ldots, \xi^M \in \{-1, +1\}^N$. As usual, we will assume that these patterns are i.i.d. and consist of i.i.d. spins (ξ_i^μ)

with

$$\mathbb{P}\big(\xi_i^\mu = \pm 1\big) = \tfrac{1}{2}.$$

Note that M may and in the interesting cases will be a function of N. The architecture of the Hopfield model is an undirected graph $G = (V, E)$, where $V = 1, \ldots, N$. With the help of the patterns and the graph, one defines the sequential dynamics $S = T_N \circ T_{N-1} \circ \cdots \circ T_1$ and the parallel dynamics $T = (T_i)$ on Σ_N. By definition T_i only changes the ith coordinate of a configuration σ and

$$S(\sigma) = T_N \circ T_{N-1} \circ \cdots \circ T_1(\sigma), \qquad T(\sigma) = \big(T_1(\sigma), \ldots, T_N(\sigma)\big),$$

$$\text{with } T_i(\sigma) = \operatorname{sgn}\left(\sum_{j=1}^N \sigma_j a_{ij} \sum_{\mu=1}^M \xi_i^\mu \xi_j^\mu\right)$$

(with the convention that $\operatorname{sgn}(0) = 1$, e.g.). Here, $a_{ij} = a_{ji} = 1$ if the edge between i and j is in E and $a_{ij} = a_{ji} = 0$ otherwise. The dynamics can be thought of as governing the evolution of the system from an input toward the nearest learned pattern. ξ^μ being a fixed point of S (or T) can thus be interpreted as recognizing a learned pattern. However, this is not really what one would call an associative memory. An important feature of the standard Hopfield model (the one where $G = K_N$, the complete graph on N vertices) is thus also that under certain restrictions on M (and the number of corrupted neurons), with high probability, a corrupted version of ξ^μ, say $\tilde{\xi}^\mu$ converges to ξ^μ when being evolved under the dynamics. This observation is also crucial for the present paper.

We can associate Hamiltonians (or energy functions) to these dynamics by

$$H_N^S(\sigma) = -\operatorname{Const.}(N) \sum_{i,j=1}^N \sigma_i \sigma_j a_{ij} \sum_{\mu=1}^M \xi_i^\mu \xi_j^\mu$$

and

$$H_N^T(\sigma) = -\operatorname{Const.}(N) \sum_{i=1}^N \left| \sum_{j=1}^N \sigma_j a_{ij} \sum_{\mu=1}^M \xi_i^\mu \xi_j^\mu \right|,$$

such that the energy will decrease along each trajectory of the dynamics:

$$H_N^S\big(S(\sigma)\big) \le H_N^S\big((\sigma)\big) \quad \text{and} \quad H_N^T\big(T(\sigma)\big) \le H_N^T\big((\sigma)\big).$$

The constant is chosen in such a way that the mean free energy of the model is finite and not constantly equal to zero.

One can easily prove that the sequential dynamics will converge to a fixed point of S and that every fixed point of S is a local minimum of H_N^S. In the parallel case, the dynamics T will converge to a fixed point or a 2-cycle of T.

The idea of this setup is that the patterns (as well as their negatives $-(\xi^\mu), \mu = 1, \ldots, M$) are hopefully possible limits of the dynamics. For instance, this is easily checked, if $M \equiv 1$ and G is

the complete graph, that ξ^1 is a local minimum of H_N^S, since then

$$H_N^S(\sigma) = -\mathrm{Const.}(N)\left(\sum_{i=1}^N \sigma_i \xi_i^1\right)^2 + \mathrm{Const._1}(N)$$

and hoped to be inherited by the more general model, as long as M is small enough. Indeed, for $M = 1$, the stored pattern ξ^1 is still a local minimum of H_N^S, if G is only connected. In this case, one obtains that

$$H_N^S(\sigma) = -\mathrm{Const.}(N)\sum_{i,j=1}^N \sigma_i \sigma_j a_{ij} \xi_i^1 \xi_j^1 = -\mathrm{Const.}(N) X^t A X$$

with $X = (\sigma_i \xi_i^1)$ and $A = (a_{ij})$. From here, the assertion is immediate (we are grateful to an anonymous referee for this remark).

When considering the stability of a random pattern ξ^μ under S or T in the above setting, we need to check whether $T_i(\xi^\mu) = \xi_i^\mu$ holds for any i. Now

$$T_i\left(\xi^\mu\right) = \mathrm{sgn}\left(\sum_{j=1}^N a_{ij} \sum_{\nu=1}^M \xi_i^\nu \xi_j^\nu \xi_j^\mu\right) = \mathrm{sgn}\left(\sum_{j=1}^N a_{ij} \xi_i^\mu + \sum_{j=1}^N a_{ij} \sum_{\nu \neq \mu} \xi_i^\nu \xi_j^\nu \xi_j^\mu\right).$$

That is, we have a signal term of strength $d(i)$, the degree of vertex i (given by the first summand on the right-hand side of the above equation) and a random noise term. The first observation is that the network topology enters via the degrees of the nodes. Indeed in such a simple setup – the stability of stored information – the minimum degree of the vertices is clearly decisive to compute the models's storage capacity: in the case where a vertex i has a small degree, the noise term will exceed the signal term, except for a very small number of stored patterns. However, it seems to be obvious that also global aspects, for example, whether or not the graph is connected, must play a role. This is confirmed if we are setting up a Hopfield model on graph G consisting of a complete graph K_m (on the vertices $1, \ldots, m$) and the graph K_{N-m} on the vertices $m+1, \ldots, N$ with $\log N \ll m \ll N$ and if we assume that these two subgraphs are disconnected or just connected by one arc. Each of the vertices thus has at least degree m and it can be computed along the lines of [31] or [33] that at least $\frac{m}{2\log N}$ patterns can be stored as fixed points of the dynamics. However, if we try to store one pattern, for example, ξ^1 with $\xi_i^1 = 1$ for all $i = 1, \ldots, N$, and start with a corrupted input $\tilde{\xi}^1$ with

$$\tilde{\xi}_i^1 = \begin{cases} -1, & i \leq m, \\ 1, & m+1 \leq i \leq N, \end{cases}$$

we see that

$$T_i\left(\tilde{\xi}^1\right) = \tilde{\xi}_i^1.$$

Hence, $\tilde{\xi}^1$ is a fixed point implying that the retrieval dynamics is not able to correct $m \ll N$ errors, even if we just want to store one pattern. So, if we insist that a neural network should

also exhibit some associative abilities (and this has always been a central argument for the use of neural networks), we have to take the graph topology into account.

This topology is encoded in the so called adjacency matrix A of G. Here, $A = (a_{ij})$ and $a_{ij} = 1$, if $e_{i,j} \in E$ and $a_{ij} = 0$ otherwise. If G is sufficiently regular, the connectivity of G (which played an important role in the above counterexample) can be characterized in terms of the spectral gap. To define it, let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$ be the (necessarily real) eigenvalues of A in decreasing order. Define κ to be the second largest modulus of the eigenvalues, that is,

$$\kappa := \max_{i \geq 2} |\lambda_i| = \max\{\lambda_2, |\lambda_N|\}.$$

Then the spectral gap is the difference between the largest eigenvalue and κ, that is, $\lambda_1 - \kappa$. However, also the degrees of the vertices are important. Hence, let $d_i = \sum_j a_{ij}$ be the degree of vertex i. We will denote by

$$\delta := \min_i d_i \quad \text{and} \quad m := \max_i d_i$$

the minimum and maximum degree of G, respectively.

In this paper, we will concentrate on the parallel dynamics, which is easier to handle when we iterate the dynamics.

3. Results

We will now state the main result of the present paper.

In order to formulate it, let us define the usual Hamming distance on the space of configurations Σ_N,

$$d_H(\sigma, \sigma') = \tfrac{1}{2}[N - (\sigma, \sigma')],$$

where (σ, σ') is the standard inner product of σ and σ'. In other words, d_H counts the number of indices where σ and σ' disagree. For any $\sigma \in \Sigma_N$ and $\varrho \in [0, 1]$, let $\mathcal{S}(\sigma, \varrho N)$ the sphere of radius ϱN centered at σ, that is,

$$\mathcal{S}(\sigma, \varrho N) = \{\sigma' : d_H(\sigma, \sigma') = [\varrho N]\},$$

where $[\varrho N]$ denotes the integer part of ϱN.

For the rest of the paper, we will suppose that the following hypothesis is true:

(H1) There exists $c_1 \in]0, 1[$, such that $\delta > c_1 \lambda_1$ (recall that δ is the minimum degree of the graph G, and λ_1 is the largest eigenvalue of its adjacency matrix).

Remark 3.1. Condition (H1) seems to be new. To understand it, recall that for a regular graph with degree d the largest eigenvalue of A equals d and so does its minimum degree δ. Condition (H1) can thus be interpreted as the requirement that G is sufficiently regular. Indeed, it turns out that, for example, for a Erdös–Rényi graph $G(N, p)$ is fulfilled, if and only if $p \gg \frac{\log N}{N}$, that is, when the graph is fully connected. Hence, for Erdös–Rényi graphs condition (H1) rules out

the sparse case, when the graph is not only disconnected asymptotically almost surely, but also very irregular, in the sense that the degree distribution is a Poisson distribution and the relative fluctuations of the degrees are large. Moreover, it will turn out that also certain power law graphs satisfy condition (H1).

We will need a second condition that keeps track on how well the graph is connected.

(H2) We say that a graph satisfies (H2), if the following relation holds between the largest eigenvalue λ_1 of the adjacency matrix and the modulus of its second largest eigenvalue κ:

$$\lambda_1 \geq c \log(N)\kappa \tag{3.1}$$

for some $c > 0$ large enough.

Remark 3.2. Roughly speaking, condition (3.1) reveals connectivity properties of the underlying graph. Clearly, it holds for the complete graph K_N, where $\lambda_1 = N - 1$ and all the other eigenvalues are equal to -1. Also, as pointed out below, condition (3.1) is fulfilled for an Erdös–Rényi random graph, if p is large enough, since the spectral gap, that is, the difference between the largest and the second largest modulus of an eigenvalue is of order $Np(1 - 1/\sqrt{Np})$.

To understand, that indeed (3.1) can be interpreted as a measure for the connectivity of the graph, assume for a moment that the graph were d-regular. Then $\lambda_1 = d$. If the graph is disconnected, there is (at least) one more eigenvalue equal to one, and hence (3.1) cannot hold. More generally, for a regular graph, the spectrum of the adjacency matrix can be computed from the spectrum of the Laplacian. On the other hand, the spectral gap of the Laplacian can be estimated by Poincaré or Cheeger type inequalities (see [12]), which roughly state that the spectral gap of the Laplacian is small, if there are vertex sets of large volume, but small surface, or if the graph has small bottlenecks. Both quantities are a measure for how well the graph is connected.

Under the above conditions, we will prove, that we can store a number M of patterns depending on λ_1 and the spectral gap of A – even in the sense that the dynamics T repairs a corrupted input. Mathematically speaking, we show the following.

Theorem 3.3. *With the notation introduced in Section* 2, *if* (H1) *and* (H2) *are satisfied, then there exists* $\alpha_c > 0$ *and* $\varrho_1 \in]0, 1/2[$ *such that if*

$$M = \alpha \frac{\lambda_1^2}{m \log N} - \frac{\kappa \lambda_1}{m},$$

for some $\alpha < \alpha_c$, *then that for all* $\varrho \in]0, \varrho_1]$ *we obtain*

$$P\left[\forall \mu = 1, \ldots, M, \forall x \text{ s.t. } d_H\left(x, \xi^\mu\right) \leq \varrho N \colon T^k(x) = \xi^\mu\right] \to 1 \qquad \text{as } N \to \infty,$$

for any $k \geq C(\max\{\log \log N, \frac{\log(N)}{\log(\lambda_1/(\kappa \log(N)))}\})$ *for a sufficiently large constant* C.

Here, T^k *is defined as the* kth *iterate of the map* T.

In other words, Theorem 3.3 states that we are able to store the given number of patterns in such a way that a number of errors that is proportional to N can be repaired by a modest (at most $\mathcal{O}(\log N)$) number of iterations of the retrieval dynamics. The number of patterns depends on the largest eigenvalue and the spectral gap of the adjacency matrix and is larger for large spectral gaps.

Before advancing to the proof, we will apply this result to some classical models of random and non-random graphs.

Corollary 3.4. *If $G = K_N$, that is, in the case of the classical Hopfield model, the storage capacity in the sense of Theorem 3.3 is $M = \alpha \frac{N}{\log N}$ for some constant α. The number of steps needed to repair a corrupted input is of order $\mathcal{O}(\log \log N)$.*

Proof. The complete graph is regular, hence condition (H1) is satisfied. From Theorem 3.3, we obtain the numerical values for M and the number of steps by observing that in the case of the complete graph the eigenvalues of A are $N - 1$ and -1 (the latter being an $N - 1$-fold eigenvalue). $\qquad\square$

Remark 3.5. It should be remarked that similar results were obtained by Komlos and Paturi [21]. In [22], even the case of regular graphs is treated. The results of these two authors were probably inspired by the results in [31], where the maximum number of patterns that are (with high probability) fixed points of the retrieval dynamics is determined. A similar result to [21] for the Hopfield model on the complete graph is due to Burshtein [7], who shows that the capacity of the Hopfield model obtained in [31] does not change, if one starts with corrupted patterns and allows for several reconstruction steps. Also, a bound for the number of necessary steps is given. These results are closely related to our result, and actually Burshtein is able to determine our α in the case of the Hopfield model on the complete graph. However, while he is working only with a random corrupted input, we consider a worst case scenario since we require that all vectors at distance $[\varrho N]$ from the originally stored pattern are attracted to this pattern by the retrieval dynamics. A similar result for a Hopfield model with $q > 2$ different states was proven in [28].

These results are to be contrasted to the findings in [24,32] or [36]. There, one is satisfied with a corrupted input being attracted to some point "close" to the stored pattern. Naturally, the resulting capacities are larger. Also, a bound on the number of iterations until this point is reached is not given.

We mainly want to apply our results to some random architectures, that is, G will be the realization of some random graph. The most popular model of a random graph is the Erdös–Renyi graph $G(N, p)$. Here, all the possible $\binom{N}{2}$ edges occur with equal probability $p = p(N)$ independently of each other. Hopfield models on $G(N, p)$ have already been discussed in [5,36] or [29].

Here, we obtain the following corollary.

Corollary 3.6. *If G is chosen randomly according to the model $G(N, p)$, then if $p \geq c_0 \frac{(\log N)^2}{N}$ for some $c_0 > 0$, for a set of realizations of G the probability of which converges to one as $N \to$*

∞, *the capacity* (*in the above sense*) *of the Hopfield model is* $cpN/\log(N)$ *for some constant* $c > 0$.

Proof. For the eigenvalues of an Erdös–Rényi graph, it is well known that with probability converging to 1, as $N \to \infty$ (such a statement in random graph theory is said to hold asymptotically almost surely), $\lambda_1 = (1 + o(1))Np$ and $\kappa \leq c\sqrt{Np}$ (see, e.g., [14,16,23] and these facts were also used in [29]). Moreover, we can control the minimum and maximum degree in $G(N, p)$. Indeed, for our values of p we have $m = (1 + o(1))Np$ and $\delta = (1 + o(1))Np$ asymptotically almost surely. Surprisingly, we could not find this result in the literature, and thus proved it in the Appendix.

Hence, (H1) is satisfied. $\square$

Remark 3.7. As mentioned above, the Hopfield model on an Erdös–Rényi graph has already been discussed in [5,36] or [29]. The first two of these papers treat the case of rather dense graphs, more precisely the regime of $p \geq \text{const.}\sqrt{\frac{\log N}{N}}$. This regime seems to be a bit artificial, since a realization of $G(N, p)$ is already connected, once p is larger than $\frac{\log N}{N}$. The regime of $\text{const.}_1 \frac{\log N}{N} \leq p \leq \text{const.}_2 \sqrt{\frac{\log N}{N}}$ was analyzed in [29]. However, in all of these papers the notion of storage capacity is the one, where we just require stored patterns to be close to minima of the energy function, that is, fixed points of the retrieval dynamics. As motivated above, this notion is unable to reflect the different reconstruction abilities for various network architectures. Corollary 3.6 deals with the notion of storage capacity introduced in Section 2; one might naturally wonder, whether the restriction $p \geq c_0 \frac{(\log N)^2}{N}$ could be weakened or whether this is the optimal condition, when we consider this notion of storage capacity. However, by now we do not have an answer to this question, especially since the reverse bound on the storage capacity is usually much harder to obtain.

The next example is one of the central results of the present paper: We analyze the Hopfield model on an architecture that comes closer to the models used in neuroscience, the so-called power law graphs. To introduce it, let us give a general construction of random graph models, which is standard in graph theory (see, e.g., [10] or [9]) and nowadays referred to as the Inhomogeneous Random Graph (see, e.g., the very recommendable lecture notes [38]). To this end, let i_0 and N positive integers and $L = \{i_0, i_0 + 1, i_0 + N - 1\}$. For a sequence $w = (w_i)_{i \in L}$, we consider random graphs $G(w)$ in which edges are assigned independently to each pair of vertices (i, j) with probability

$$p_{ij} = \varrho w_i w_j,$$

where $\varrho = 1/\sum_{k \in L} w_k$. We assume that

$$\max_i w_i^2 < \sum_{k \in L} w_k$$

so that $p_{ij} \leq 1$ for all i and j. It is easy to see that the expected degree of i is w_i. This allows for a very general construction of random graphs. Note in particular that for $w_i = pN$ for all $i = 1, \ldots, N$, one recovers the Erdös–Rényi graph.

For notational convenience, let

$$d = \sum_{i \in L} w_i / N$$

be the expected average degree, $\overline{m}$ the expected maximum degree and

$$\tilde{d} = \sum_{i \in L} w_i^2 \Big/ \sum_{i \in L} w_i$$

be the so-called second-order average degree of the graph $G(w)$. From these definitions, the advantage of this kind of construction of a random graph becomes transparent: We are able to construct random graphs, with expected degrees that are up to our own choice.

We now turn to a subclass of random graphs that have recently become very popular, power law graphs [13]. Power law random graphs are random graphs in which the number of vertices of degree k is proportional to $1/k^\beta$ for some fixed exponent β. It has been realized that this "power law"-behavior is prevalent in realistic graphs arising in various areas. Graphs with power law degree distribution are ubiquitously encountered, for example, in the internet, the telecommunications graphs, the neural networks and many biological applications [20,34,39]. The common feature of such networks is that they are large, have small diameter, but have small average degree. This behavior can be achieved by hubs, a few vertices with a much larger degree than others. A possible choice would be a power law graph, where the degrees obey a power law distribution. Keeping in mind that the $G(w)$ model allows to build a graph model with a given expected degree sequence, it is plausible that this model can be used to model the networks of the given examples. Indeed, using the $G(w)$ model, we can build random power law graphs in the following way. Given a power law exponent β, a maximum expected degree $\overline{m}$, and an average degree d, we take $w_i = c i^{-1/(\beta-1)}$ for each $i \in \{i_0, \ldots, i_0 + 1, i_0 + N - 1\}$, with

$$c = \frac{\beta - 2}{\beta - 1} d N^{1/(\beta-1)}$$

and

$$i_0 = N \left(\frac{d(\beta - 2)}{\overline{m}(\beta - 1)} \right)^{\beta-1}.$$

For such power law graphs, we obtain the following.

Corollary 3.8. *If G is chosen randomly according to a power law graph with $\beta > 3$, then if*

$$\overline{m} \gg d > c \sqrt{\overline{m}} \big(\log(N) \big)^{3/2}$$

or

$$\overline{m} \gg d > c \sqrt{\overline{m}} \big(\log(N) \big) \quad \text{and} \quad \overline{m} \gg (\log N)^4,$$

for some constant $c > 0$ for a set of realizations of G the probability of which converges to one as $N \to \infty$, the capacity (in the above sense) of the Hopfield model is $C(\beta)\frac{d^2}{\overline{m}\log(N)}$ for a constant C that only depends on β.

Remark 3.9.

- One might indeed wonder, whether the restriction of $\beta > 3$ is an artefact of our proof below or whether there is some intrinsic reason, why storing become much more difficult for $\beta < 3$. A recent paper by Jacob and Mörters [19] may shed some light on this question. There a spatial preferential attachment graph is constructed (for details, see the construction in [19]). It turns out that the graphs have powerlaw behavior for the degree distribution. The parameter β depends on the parameters of the model. For their model, the authors are able to show that for $\beta > 3$ the models exhibits clustering, that is, many triangles occur, while for $\beta < 3$ there is no clustering. On the other hand, the storage of patterns in a Hopfield is basically a collective phenomenon for which a strong interaction of the neurons is necessary. Clustering is a measure for such a strong interaction.
- The second condition in Corollary 3.8 basically states that we assume that there are so-called hubs, that is, vertices with a much larger degree than the average one, but that the graph may not be too irregular, for example, for a star graph (one vertex connected to all other vertices that are not connected otherwise), this condition would be violated, and indeed we would not be able to repair corrupted patterns on such a graph.

Proof of Corollary 3.8. By definition, if $d \ll \overline{m}$, then the minimum expected degree $w_{\min} = c(i_0 + N - 1)^{-1/(\beta - 1)}$ satisfies $w_{\min} = \frac{\beta - 2}{\beta - 1}d(1 + o(1))$.

From [9], we learn about the second-order average degree that

$$\tilde{d} = (1 + o(1))\frac{(\beta - 2)^2}{(\beta - 1)(\beta - 3)}d,$$

if $\beta > 3$.

On the other hand, Chung and Radcliffe prove in [11] the following: if the maximum expected degree $\overline{m}$ satisfies $\overline{m} > \frac{8}{9}\log(\sqrt{2}N)$, then with probability at least $1 - \frac{1}{N}$, we have

$$\lambda_1(G(w)) = (1 + o(1))\tilde{d} \quad \text{and} \quad \kappa(G(w)) \leq \sqrt{8\overline{m}\log(\sqrt{2}N)}.$$

We will now use the following exponential bound due to Chung and Lu. As shown by these authors in [10], we have the following estimate, using Chernoff inequalities: for all $c > 0$, there exist two constants $c_0, c_1 > 0$ such that

$$P[\exists i \in L\colon |d_i - w_i| > cw_i] \leq \sum_{i \in L}\exp(-c_0 w_i) \leq \exp(-c_1 d + \log N), \qquad (3.2)$$

since $w_{\min} = \mathcal{O}(d)$. Applying this with, for example, $c = 1/2$, we see (applying the Borel–Cantelli lemma) that for almost all realizations of the random graphs, we have that for all $i \in L$,

$$d_i > \frac{1}{2}w_i \geq \frac{1}{2}w_{\min} = \frac{1}{2}\frac{\beta - 2}{\beta - 1}d(1 + o(1)) = \frac{1}{2}\frac{\beta - 3}{\beta - 2}\lambda_1(1 + o(1)),$$

and thus

$$\delta > \frac{1}{2}\frac{\beta - 3}{\beta - 2}(1 + o(1))\lambda_1,$$

which is (H1).

To apply Theorem 3.3, we also need to compare $\overline{m}$ to the maximum degree m of a graph G, chosen randomly according to a power law graph with $\beta > 3$. We again use (3.2).

Under our assumption that $d \gg \log N$, we deduce from this estimate that $m = C\overline{m}(1 + o(1))$, for some $C > 0$, and we finally obtain that the capacity of the Hopfield model on power law graphs (for a sequence of sets of graphs with probability converging to one) is at least

$$\text{const.} \frac{\lambda_1^2}{m \log(N)} - \frac{\kappa\lambda_1}{m} = C(\beta)\frac{d^2}{\overline{m}\log(N)},$$

if, $\beta > 3$, and $\kappa < c_2\frac{\lambda_1}{\log(N)}$ for some $c_2 > 0$ small enough. This is true in particular, if

$$\sqrt{8\overline{m}\log(\sqrt{2}N)} < c_3\frac{d}{\log(N)}$$

for some c_3 small enough, that is,

$$d > c\sqrt{\overline{m}}\big(\log(N)\big)^{3/2}.$$

In fact, this condition on d can be slightly weakened, if we consider the slightly stronger condition on the maximum expected degree: $\overline{m} \gg (\log N)^4$. Indeed, in a recent paper [30], Lu and Peng prove that under this condition on $\overline{m}$, we have

$$\lambda_1\big(G(w)\big) = \big(1 + o(1)\big)\tilde{d} \quad \text{and} \quad \kappa\big(G(w)\big) \le 2\sqrt{\overline{m}}\big(1 + o(1)\big),$$

a.s., if $\tilde{d} \gg \sqrt{\overline{m}}$. Finally, we get as previously a capacity of order

$$C(\beta)\frac{d^2}{\overline{m}\log(N)},$$

if $\overline{m} \gg d > c\sqrt{\overline{m}}(\log(N))$ and $\overline{m} \gg (\log N)^4$. □

4. Technical preparations on random graphs

We first present the results we will use in the proof of our theorem. Let G be a simple graph with N vertices and l edges. Recall that for such a graph

$$\lambda_1 \ge \cdots \ge \lambda_N$$

are the (real) eigenvalues of its adjacency matrix and $\kappa = \max\{\lambda_2, |\lambda_N|\}$.

We begin with an estimate of the moment generating function of a sum of i.i.d. random variables, related to G. We assign i.i.d. random variables X_i to the vertices of G, taking values ± 1 with equal probability. Let us define the "quadratic form" over G

$$S = \sum_{\{i,j\} \in E} X_i X_j.$$

The following theorem due to Komlos and Paturi [22] gives an upper bound on the moment generating function of S, which appears naturally when we use an exponential Markov inequality for an upper bound.

Theorem 4.1 ([22]). *The moment generating function of S can be bounded as*

$$E\big[e^{-tS}\big] \leq E\big[e^{tS}\big] \leq \exp\Big(\frac{lt^2}{2(1-\lambda_1 t)}\Big),$$

for $0 \leq t < 1/\lambda_1$.

Remark 4.2. The attentive reader may wonder, whether the above theorem is really difficult to prove, as the random variables $X_i X_j$ are Bernoulli random variables. However, note that they are not independent, which is the basic difficulty in this estimate.

Not unexpectedly, a bound on the moment generating function implies a concentration of measure result.

Corollary 4.3. *For any $y > 0$, we have*

$$P[S > y] \leq \exp\Big(-\frac{y^2}{2(l+\lambda_1 y)}\Big).$$

Proof. Apply the exponential Markov inequality together with Theorem 4.1 to see that

$$P[S > y] \leq e^{-ty} E\big[e^{tS}\big] \leq \exp\Big(-ty + \frac{lt^2}{2(1-\lambda_1 t)}\Big),$$

for $0 \leq t < 1/\lambda_1$. The desired estimate is obtained by the choice of $t = \frac{y}{l+\lambda_1 y}$ which is smaller than $1/\lambda_1$. □

As we will apply this result for subgraphs in the proof of our main result, we need also an estimate of the largest eigenvalue $\lambda_1(H)$ of particular subgraphs H of G. To this end, we will quote another result by Komlos and Paturi [22].

Lemma 4.4 ([22]). *Let G be a simple graph with N vertices. If I and J are two subsets of the vertex set of G with $|I| = \varrho N$ and $|J| = \varrho' N$, where $\varrho, \varrho' \in (0, 1)$, the number of edges $e(J; I)$ going from J to I is at most*

$$e(J, I) \leq \big[\varrho\varrho'\lambda_1(G) + \sqrt{\varrho\varrho'}\kappa(G)\big]N.$$

Moreover, the largest eigenvalue (of the adjacency matrix) of the graph H determined by the edges from I to J is bounded as

$$\lambda_1(H) \leq 2\left[\sqrt{\varrho\varrho'}\lambda_1(G) + \left(1 - \sqrt{\varrho\varrho'}\right)\kappa(G)\right].$$

The proof of this lemma basically involves estimating quadratic forms by their eigenvalues together with Cauchy's interlacing theorem for eigenvalues of matrices. However, it is not trivial (see the proof in [22]).

5. Proof of the main result

We are now ready to begin with the proof of Theorem 3.3. We first present an important lemma that determines the behavior of the system for one step of the synchronous dynamics, more precisely it controls, how many errors are corrected by one step of the dynamics.

Lemma 5.1. *Recall that m denotes the maximum degree of the random graph G in question and let*

$$\varrho_0 = \exp\left(-c_2 \frac{\lambda_1}{\kappa + Mm/\lambda_1}\right),$$

for some constant $c_2 > 0$. If $M \leq c\lambda_1$ for some constant $c > 0$, there exists $\varrho_1 \in (0, \frac{1}{2})$ and a constant $c_1 > 0$, such that for all $\varrho \in [\varrho_0, \varrho_1]$ we have

$$P\left[\forall \mu \in \{1, \dots, M\}, \forall x \in \mathcal{S}\left(\xi^\mu, \varrho N\right): d_H\left(T(x), \xi^\mu\right) \leq f(\varrho)N\right] \geq 1 - \varepsilon_N,$$

where

$$f(\varrho) = \max\left\{c_1\varrho\left(\frac{\kappa}{\lambda_1}\right)^2, c_1\varrho h(\varrho), c_1\frac{\kappa}{\lambda_1}h(\varrho), c_1\varrho\left(\frac{M\kappa}{(\lambda_1)^2}\log\left(\frac{1}{\varrho}\right)\right)^{2/3}, \varrho_0\right\} \leq \varrho,$$

$\varepsilon_N \geq 0$, $\varepsilon_N \to 0$ as $N \to +\infty$ and

$$h(\varrho) = -\varrho \log\varrho - (1 - \varrho)\log(1 - \varrho)$$

is the entropy function.

Proof. This lemma is of central importance for our main result. However, its proof is rather technical. Let us therefore first describe its basic idea.

To this end, recall that it suffices to prove that

$$\sum_{\mu=1}^{M} P\left[\exists x \in \mathcal{S}\left(\xi^\mu, \varrho N\right): d_H\left(T(x), \xi^\mu\right) > f(\varrho)N\right] \leq \varepsilon_N. \tag{5.1}$$

To simplify notation, we can assume that the fundamental memory in question is ξ^1.

Now assume that we start with a corrupted input (i.e., a corrupted pattern) $x \in \{-1, 1\}^N$ such that $d_H(\xi^1, x) = \varrho N$. Let I be the set of coordinates in which x and ξ^1 differ. Let $T(x)$ be the vector resulting after one step of the parallel dynamics, and J be the set of coordinates in which $T(x)$ and ξ^1 differ. Now define the weight matrix W as $W = (w_{ij})$ and

$$w_{ij} = a_{ij} \sum_{\nu=1}^{M} \xi_i^\nu \xi_j^\nu.$$

Then, since ξ^1 is not properly reconstructed for the coordinates $j \in J$, for all $j \in J$, we have $\xi_j^1 (Wx)_j \leq 0$, which implies $\sum_{j \in J} \xi_j^1 (Wx)_j \leq 0$.

The idea is now to analyze the contributions to $\sum_{j \in J} \xi_j^1 (Wx)_j$. Similar to what we said in the analysis of the dynamics T_i in Section 2, there is a "signal term" stemming from the closeness of x to ξ^1 and there are noise terms from the influence of the other patterns. We will first show that the signal term grows at least linearly in $|J|$. On the other hand, we are also able to give an upper bound on the influence of the random noise terms that are also controlled by the size of I and J. While all these computations are relatively straight forward in the Hopfield model on the complete graph, the estimates become much more involved on a general graph. The key observation is that we are able to control the probability to find sets I and J with the above properties with the help of the spectrum of the adjacency matrix (using the results of the previous section). Technically to this end, we have to split up the noise terms according to where the vertices i in $\sum_{j \in J} \sum_I \xi_j^1 a_{ij} \xi_i^\mu \xi_j^\mu$ come from. The bottom line is, that if $|J|$ is too large, the probability to find sets I, with $|I| = \varrho N$ and J (such that ξ^1 is not reconstructed correctly on J when starting with an x differing from ξ^1 in the coordinates I) converges to 0 – even when being multiplied by the number of patterns M, if M is of the given size (cf. equation (5.3) below).

Let us now carry out this idea.

For later use, set

$$S^\mu(J, I) = \sum_{j \in J} \sum_{k=1}^{N} a_{jk} \xi_j^1 \xi_j^\mu \xi_k^\mu x_k$$

and

$$S(J, I) = \sum_{\mu=1}^{M} S^\mu(J, I) =: \sum_{j \in J} \xi_j^1 (Wx)_j.$$

Observe that, if the patterns are chosen i.i.d. with i.i.d. coordinates their typical distance is $N/2 \pm \text{const.} \sqrt{N}$. This in turn implies that, if $\varrho < 1/2$ and $d_H(x, \xi^1) = \varrho N$, then x tends to be closer to ξ^1 than to any other pattern, and $S^1(J, I)$ will be the dominating term in $S(J, I)$. We will first give a lower bound for $S^1(J, I)$. We can rewrite $S^1(J, I)$ as

$$S^1(J, I) = \sum_{j \in J} \sum_{k=1}^{N} a_{jk} \xi_k^1 x_k = \sum_{j \in J} \left(e(j, \bar{I}) - e(j, I) \right) = e(J, V) - 2e(J, I),$$

where again we use the notation $e(J, I)$ and $e(j, I)$, to denote the number edges going from the set J to the set I, or, respectively, from the vertex j to the set I. Moreover, $\bar{I}$ denotes the complement of the set I in V.

Under the assumption of hypothesis (H1) and with the help of Lemma 4.4, we have for all I and J,

$$S^1(J, I) \geq c_1 \lambda_1 |J| - 2\left(|I||J|\frac{\lambda_1}{N} + \sqrt{|I||J|}\kappa \right)$$

$$= \lambda_1 |J|\left(c_1 - 2\varrho - 2\sqrt{\frac{\varrho}{\varrho'}}\frac{\kappa}{\lambda_1} \right),$$

where $\varrho' = \frac{|J|}{N}$. If we assume that $\varrho' \geq c_2 \varrho (\frac{\kappa}{\lambda_1})^2$ for some $c_2 > 0$ large enough, and $\varrho < \varrho_1$ for some $\varrho_1 \in (0, 1/2)$ small enough, we get

$$S^1(J, I) \geq C_1 \lambda_1 |J|, \tag{5.2}$$

for some constant $C_1 \in (0, 1)$.

For $\mu \geq 2$, we compute

$$S^\mu(J, I) = \sum_{(j,k)\in E(J,\bar{I})} u_j^\mu u_k^\mu - \sum_{(j,k)\in E(J,I)} u_j^\mu u_k^\mu$$

$$= \sum_{(j,k)\in E(J,V)} u_j^\mu u_k^\mu - 2 \sum_{(j,k)\in E(J,I)} u_j^\mu u_k^\mu,$$

where $u_i^\mu = \xi_i^1 \xi_i^\mu$, for all $i = 1, \ldots, N$ and $\mu = 1, \ldots, M$. To apply the results for the moment generating function of quadratic forms introduced in Theorem 4.1 and Corollary 4.3, we need to rewrite these sums over ordered pairs of vertices as sums over unordered pairs. We have

$$E(J, V) = E(J, J) + E(J, \bar{J}) = 2E\{J, J\} + E\{J, \bar{J}\} = E\{J, V\} + E\{J, J\},$$

where for $K, L \subset V$ $E(K, L)$ is the edges set of the directed graph between the sets K and L induced by our original graph. Likewise, $E\{K, L\}$ denotes the corresponding set of undirected edges. In the same way, we obtain

$$E(J, I) = E(J \cap \bar{I}, J \cap I) + E(J \cap I, J \cap I) + E(J, I \cap \bar{J})$$

$$= E\{J \cap \bar{I}, J \cap I\} + 2E\{J \cap I, J \cap I\} + E\{J, I \cap \bar{J}\}$$

$$= E\{J, I\} + E\{J \cap I, J \cap I\}.$$

Eventually,

$$E(J, V) - 2E(J, I) = E\{J, V\} - 2E\{J, I\} + E\{J, J\} - 2E\{J \cap I, J \cap I\}.$$

We want to prove that for ϱ' larger than $f(\varrho)$ we have that

$$MP\left[\exists I, |I| = \varrho N, \exists J, |J| = \varrho' N, S(J, I) < 0\right] \longrightarrow 0, \tag{5.3}$$

as $N \to +\infty$.

To this end, set

$$S_1^{\mu}(J) = \sum_{(j,k) \in E\{J,V\}} u_j^{\mu} u_k^{\mu}, \qquad S_2^{\mu}(J, I) = \sum_{(j,k) \in E\{J,I\}} u_j^{\mu} u_k^{\mu},$$

$$S_3^{\mu}(J) = \sum_{(j,k) \in E\{J,J\}} u_j^{\mu} u_k^{\mu} \quad \text{and} \quad S_4^{\mu}(J, I) = \sum_{(j,k) \in E\{J \cap I, J \cap I\}} u_j^{\mu} u_k^{\mu}.$$

Then

$$S(J, I) = S^1(J, I) + \sum_{\mu=2}^{M} S_1^{\mu}(J) - 2 \sum_{\mu=2}^{M} S_2^{\mu}(J, I) + \sum_{\mu=2}^{M} S_3^{\mu}(J) - \sum_{\mu=2}^{M} S_4^{\mu}(J, I).$$

Let $\gamma_1, \gamma_2, \gamma_3, \gamma_4 \geq 0$, such that $\gamma_1 + 2\gamma_2 + \gamma_3 + \gamma_4 = 1$.

We will consider the four sums separately. First, using (5.2), we have

$$P\left[\exists I, |I| = \varrho N, \exists J, |J| = \varrho' N, \sum_{\mu=2}^{M} S_1^{\mu}(J) < -\gamma_1 S^1(J, I)\right]$$

$$\leq \sum_{J: |J| = \varrho' N} P\left[\sum_{\mu=2}^{M} S_1^{\mu}(J) < -\gamma_1 C_1 \lambda_1 |J|\right].$$

Given the vector $\xi^1 = (\xi_i^1)_{i=1,\ldots,N}$, the random variables $(u_i^{\mu})_{i=1,\ldots,N}^{\mu=2,\ldots,M}$ are conditionally independent and uniformly distributed on $\{-1, +1\}$. As the estimates we will get for the conditional probabilities and the moment generating function will not depend on the choice of ξ^1, they will be true also for the unconditional probabilities.

Given the vector ξ^1, the random variables $S_1^{\mu}(J), \mu = 2, \ldots, M$, are independent. Similar to the estimate of Corollary 4.3, we obtain

$$P\left[\sum_{\mu=2}^{M} S_1^{\mu}(J) < -\gamma_1 C_1 \lambda_1 |J|\right] \leq \exp\left(-\frac{1}{2} \frac{\gamma_1 C_1 \lambda_1 |J|}{\lambda_J + Me\{J, V\}/(\gamma_1 C_1 \lambda_1 |J|)}\right),$$

where $\lambda_J = \lambda_1(E\{J, V\})$ is the largest eigenvalue of the graph determined by the undirected edges in $E\{J, V\}$. Using Lemma 4.4, we have

$$\lambda_J \leq 2\left[\sqrt{\varrho'} \lambda_1 + (1 - \sqrt{\varrho'})\kappa\right],$$

and moreover, $e\{J, V\} \leq e(J, V) \leq m|J|$ is trivially true. We deduce that

$$P\left[\sum_{\mu=2}^{M} S_1^{\mu}(J) < -\gamma_1 C_1 \lambda_1 |J|\right] \leq \exp\left(-\frac{\gamma_1 C_1}{2} \frac{\varrho' N}{2\sqrt{\varrho'} + 2\kappa/\lambda_1 + Mm/(\gamma_1 C_1 (\lambda_1)^2)}\right).$$

Now there are $\binom{N}{|J|}$ ways to choose the set J, and by Stirling's formula

$$\binom{N}{|J|} \le \exp\bigl(h(\varrho')N\bigr),$$

where

$$h(x) = -x \log x - (1-x)\log(1-x)$$

is the entropy function introduced above.

Using $h(\varrho') \le -2\varrho' \log(\varrho')$, we obtain that

$$\sum_{J:\,|J|=\varrho'N} P\left[\sum_{\mu=2}^{M} S_1^{\mu}(J) < -\gamma_1 C_1 \lambda_1 |J|\right]$$

$$\le \exp\left(-2\varrho'N\left(\frac{\gamma_1 C_1}{4}\frac{1}{2\sqrt{\varrho'}+2\kappa/\lambda_1+Mm/(\gamma_1 C_1(\lambda_1)^2)}+\log(\varrho')\right)\right).$$

The exponent is negative, if

$$\frac{\gamma_1 C_1}{4}\frac{1}{2\sqrt{\varrho'}+2\kappa/\lambda_1+Mm/(\gamma_1 C_1(\lambda_1)^2)}+\log(\varrho') > 0,$$

which is true if

$$\frac{\gamma_1 C_1}{8}\frac{1}{2\sqrt{\varrho'}}+\log(\varrho') > 0, \tag{5.4}$$

as well as

$$\frac{\gamma_1 C_1}{8}\frac{\lambda_1}{2\kappa+Mm/(\gamma_1 C_1\lambda_1)}+\log(\varrho') > 0. \tag{5.5}$$

This gives a first bound on $f(\varrho)$ in the sense, that if ϱ' is so large, then we will have small probabilities to find the corresponding sets I and J.

Now, there exists a $\varrho_1 \in (0, 0.1)$, such that the first condition (5.4) is true if $\varrho' < \varrho_1$. The second condition (5.5) is true if

$$\varrho' > \exp\left(-c\frac{\lambda_1}{2\kappa+Mm/(\gamma_1 C_1\lambda_1)}\right),$$

where $c = \frac{\gamma_1 C_1}{8}$. This implies that, if there exists a constant $c_2 > 0$ such that

$$\varrho' \ge \varrho_0 := \exp\left(-c_2\frac{\lambda_1}{\kappa+Mm/\lambda_1}\right),$$

then (5.5) is true.

For the second term, we have

$$P\left[\exists I, |I| = \varrho N, \exists J, |J| = \varrho' N, \sum_{\mu=2}^{M} S_2^{\mu}(J, I) > \gamma_2 S^1(J, I)\right]$$

$$\leq \sum_{I: |I|=\varrho N} \sum_{J: |J|=\varrho' N} P\left[\sum_{\mu=2}^{M} S_2^{\mu}(J, I) > \gamma_2 C_1 \lambda_1 |J|\right]$$

$$\leq \sum_{I: |I|=\varrho N} \sum_{J: |J|=\varrho' N} \exp\left(-\frac{1}{2} \frac{\gamma_2 C_1 \lambda_1 |J|}{\lambda_{\{J,I\}} + M e\{J, I\}/(\gamma_2 C_1 \lambda_1 |J|)}\right),$$

where $\lambda_{\{J,I\}} = \lambda_1(E\{J, I\})$ is the largest eigenvalue of the graph determined by the undirected edges in $E\{J, I\}$. Using Lemma 4.4, we get

$$\lambda_{\{J,I\}} \leq 2\left[\sqrt{\varrho\varrho'}\lambda_1 + \kappa\right]$$

and

$$e\{J, I\} \leq \left(\varrho\varrho'\lambda_1 + \sqrt{\varrho\varrho'}\kappa\right)N,$$

which implies

$$P\left[\sum_{\mu=2}^{M} S_2^{\mu}(J) > \gamma_2 C_1 \lambda_1 |J|\right]$$

$$\leq \exp\left(-\frac{\gamma_2 C_1}{2} \frac{\varrho' N}{2\sqrt{\varrho\varrho'} + 2\kappa/\lambda_1 + M\varrho/(\gamma_2 C_1 \lambda_1) + (M\kappa/(\gamma_2 C_1(\lambda_1)^2))\sqrt{\varrho/\varrho'}}\right).$$

There are $\binom{N}{|I|}\binom{N}{|J|}$ ways to choose the sets I and J and

$$\binom{N}{|I|}\binom{N}{|J|} \leq \exp((h(\varrho) + h(\varrho')N) \leq \exp(2h(\varrho)n),$$

as we assume that $\varrho' \leq \varrho \leq 1/2$. These considerations yield that

$$P\left[\exists I, |I| = \varrho N, \exists J, |J| = \varrho' N, \sum_{\mu=2}^{M} S_2^{\mu}(J, I) > \gamma_2 S^1(J, I)\right]$$

becomes small, once the condition

$$\frac{\gamma_2 C_1}{2} \frac{\varrho'}{2\sqrt{\varrho\varrho'} + 2\kappa/\lambda_1 + M\varrho/(\gamma_2 C_1 \lambda_1) + (M\kappa/(\gamma_2 C_1(\lambda_1)^2))\sqrt{\varrho/\varrho'}} > 2h(\varrho),$$

is satisfied. This is true if

$$\gamma_2 C_1 \frac{\varrho'}{4\sqrt{\varrho\varrho'}} > 8h(\varrho), \qquad \gamma_2 C_1 \frac{\varrho'}{4\kappa/\lambda_1} > 8h(\varrho), \qquad \frac{\gamma_2 C_1}{2} \frac{\varrho'}{M\varrho/(\gamma_2 C_1 \lambda_1)} > 8h(\varrho)$$

and

$$\frac{\gamma_2 C_1}{2} \frac{\varrho'}{(M\kappa/(\gamma_2 C_1 (\lambda_1)^2))\sqrt{\varrho/\varrho'}} > 16\varrho \log\left(\frac{1}{\varrho}\right) \geq 8h(\varrho).$$

From here, we obtain the four conditions

$$\varrho' > C^2 \varrho h(\varrho)^2, \qquad \varrho' > C\frac{\kappa}{\lambda_1}h(\varrho), \qquad \varrho' > C'\frac{M}{\lambda_1}\varrho h(\varrho)$$

and

$$\varrho' \geq \varrho\left(\frac{2}{C'} \frac{M\kappa}{(\lambda_1)^2} \log\left(\frac{1}{\varrho}\right)\right)^{2/3},$$

where $C = \frac{32}{\gamma_2 C_1}$ and $C' = \frac{16}{(\gamma_2 C_1)^2}$.

For the third term, we have

$$P\left[\exists I, |I| = \varrho N, \exists J, |J| = \varrho' N, \sum_{\mu=2}^{M} S_3^\mu(J, J) < -\gamma_3 S^1(J, I)\right]$$

$$\leq \sum_{J:\, |J|=\varrho' N} P\left[\sum_{\mu=2}^{M} S_3^\mu(J, J) < -\gamma_3 C_1 \lambda_1 |J|\right]$$

$$\leq \sum_{J:\, |J|=\varrho' N} \exp\left(-\frac{1}{2} \frac{\gamma_3 C_1 \lambda_1 |J|}{\lambda_{\{J,J\}} + Me\{J, J\}/(\gamma_3 C_1 \lambda_1 |J|)}\right),$$

where $\lambda_{\{J,J\}} = \lambda_1(E\{J, J\})$ is the largest eigenvalue of the graph determined by the undirected edges in $E\{J, J\}$. Using Lemma 4.4, we have

$$\lambda_{\{J,J\}} \leq 2\varrho' \lambda_1 + 2\kappa$$

and $e\{J, J\} \leq (\varrho'\lambda_1 + \kappa)\varrho' N$, and we obtain as for the previous terms

$$\exp\left(-\frac{1}{2} \frac{\gamma_3 C_1 \lambda_1 |J|}{\lambda_{\{J,J\}} + Me\{J, J\}/(\gamma_3 C_1 \lambda_1 |J|)}\right)$$

$$\leq \exp\left(-\frac{\gamma_3 C_1}{2} \frac{\varrho' N}{(2 + M/(\gamma_3 C_1 \lambda_1))\varrho' + (\kappa/\lambda_1)(2 + M/(\gamma_3 C_1 \lambda_1))}\right).$$

There are $\binom{N}{|J|}$ ways to choose the set J. From this, we see that

$$P\left[\exists I, |I| = \varrho N, \exists J, |J| = \varrho' N, \sum_{\mu=2}^{M} S_3^{\mu}(J, J) < -\gamma_3 S^1(J, I)\right]$$

becomes small, if the condition

$$\frac{\gamma_3 C_1}{2(2 + M/(\gamma_3 C_1 \lambda_1))} \frac{1}{1 + \kappa/(\lambda_1 \varrho')} > h(\varrho'),$$

is fulfilled, which is true if

$$h(\varrho') < C \quad \text{and} \quad h(\varrho') < C\varrho'\frac{\lambda_1}{\kappa} \qquad \text{where } C = \frac{\gamma_3 C_1}{4(2 + M/(\gamma_3 C_1 \lambda_1))}. \qquad (5.6)$$

As we assume that $M \le c\lambda_1$, there exists a $\varrho_2(\gamma_3, C_1) \in (0, 0.1)$, such that the first inequality in (5.6) is true if $\varrho' < \varrho_2$.

Using the bound $h(\varrho') \le -2\varrho' \log(\varrho')$ again, we get that there exists $c > 0$ such that the second condition in (5.6) is true if

$$\varrho' > \exp\left(-c\frac{\lambda_1}{\kappa}\right).$$

For the fourth term, we have

$$P\left[\exists I, |I| = \varrho N, \exists J, |J| = \varrho' N, \sum_{\mu=2}^{M} S_4^{\mu}(J, I) > \gamma_4 S^1(J, I)\right]$$

$$\le \sum_{I:\, |I|=\varrho N} \sum_{J:\, |J|=\varrho' N} P\left[\sum_{\mu=2}^{M} S_4^{\mu}(J, I) > \gamma_4 C_1 \lambda_1 |J|\right]$$

$$\le \sum_{I:\, |I|=\varrho N} \sum_{J:\, |J|=\varrho' N} \exp\left(-\frac{1}{2}\frac{\gamma_4 C_1 \lambda_1 |J|}{\lambda_{J \cap I} + Me\{J \cap I, J \cap I\}/(\gamma_4 C_1 \lambda_1 |J|)}\right),$$

where $\lambda_{J \cap I} = \lambda_1(E\{J \cap I, J \cap I\})$ is the largest eigenvalue of the graph determined by the undirected edges in $E\{J \cap I, J \cap I\}$.

Using Lemma 4.4 and assuming that $\varrho' \le \varrho$, we have $\lambda_{J \cap I} \le 2\varrho' \lambda_1 + 2\kappa$ and $e\{J \cap I, J \cap I\} \le (\varrho'\lambda_1 + \kappa)\varrho' N$, which are the same bounds as for the third term. There are $\binom{N}{|I|}\binom{N}{|J|}$ ways to choose the sets I and J and using again

$$\binom{N}{|I|}\binom{N}{|J|} \le \exp\left((h(\varrho) + h(\varrho')N\right) \le \exp\left(2h(\varrho)N\right),$$

we finally arrive at the same conditions as for the third term, with possibly a different constant C.

Finally, the various conditions can be summarized as

$$\varrho' \geq c_2\varrho\left(\frac{\kappa}{\lambda_1}\right)^2, \qquad \varrho_1 \geq \varrho \geq \varrho' \geq \varrho_0, \qquad \varrho' > C^2\varrho h(\varrho)^2,$$

$$\varrho' > C\frac{\kappa}{\lambda_1}h(\varrho), \qquad \varrho' > C'\frac{M}{\lambda_1}\varrho h(\varrho) \quad \text{and} \quad \varrho' \geq \varrho\left(\frac{2}{C'}\frac{M\kappa}{(\lambda_1)^2}\log\left(\frac{1}{\varrho}\right)\right)^{2/3}.$$

Finally, taking into account all the conditions, we get that (5.1) is true if we choose

$$f(\varrho) = \max\left\{c_1\varrho\left(\frac{\kappa}{\lambda_1}\right)^2, c_1\varrho h(\varrho), c_1\frac{\kappa}{\lambda_1}h(\varrho), c_1\varrho\left(\frac{M\kappa}{(\lambda_1)^2}\log\left(\frac{1}{\varrho}\right)\right)^{2/3}, \varrho_0\right\}$$

for some $c_1 > 0$ large enough and we see that $f(\varrho) \leq \varrho$ if $\varrho \in (\varrho_0, \varrho_1)$ with ϱ_1 small enough. □

In order to prove the Theorem 3.3, we will apply Lemma 5.1 repeatedly until the system attains an original pattern. Using

$$\varrho_0 = \exp\left(-c_2\frac{\lambda_1}{\kappa + Mm/\lambda_1}\right),$$

we get that the system can attain an original pattern, that is, $\varrho_0 N < 1$, only if

$$\kappa + \frac{Mm}{\lambda_1} < c_2\lambda_1/\log(N)$$

(which follows from the choice of M made in Theorem 3.3).

To determine the maximal number of steps the synchronous dynamics needs to converge, we analyze the following sequences.

Lemma 5.2. *Let $(w_n)_{n\in\mathbb{N}}, (x_n)_{n\in\mathbb{N}}, (y_n)_{n\in\mathbb{N}}$ and $(z_n)_{n\in\mathbb{N}}$ such that*

$$w_0 = x_0 = y_0 = z_0 = \varrho \in \left[\exp\left(-\frac{1}{2c}\frac{\lambda_1}{\kappa}\right), 1/e\right]$$

and

$$w_{n+1} = cw_n\left(\frac{\kappa}{\lambda_1}\right)^2, \qquad x_{n+1} = cx_n h(x_n),$$

$$y_{n+1} = c\frac{\kappa}{\lambda_1}h(y_n) \quad \text{and}$$

$$z_{n+1} = cz_n\left(\frac{M\kappa}{(\lambda_1)^2}\log\left(\frac{1}{z_n}\right)\right)^{2/3},$$

for $n \in \mathbb{N}$ and $c > 0$. Let us assume that $\frac{\lambda_1}{\kappa} > C_1 \log N$ for some $C_1 > 1$ large enough and that $M \leq C_2\lambda_1$ for some $C_2 > 0$. Then the sequences $(w_n), (x_n), (y_n)$ and (z_n) are decreasing and

there exists $C_3 > 0$ and

$$n_0 \geq C_3 \max\left\{\log\log N, \frac{\log(N)}{\log(\lambda_1/(\kappa\log N))}\right\}$$

such that $\max\{w_{n_0}, x_{n_0}, y_{n_0}, z_{n_0}\} < 1/N$.

Proof. Let us first consider the sequence (w_n). Iterating $w_{n+1} = aw_n$, with $a = c(\frac{\kappa}{\lambda_1})^2$, we get trivially $w_n = a^n w_0$ from which we deduce that $w_n < \frac{1}{N}$ as soon as $n > c_1 \frac{\log(N)}{\log(\lambda_1/\kappa)}$ for some $c_1 > 0$.

For the sequence (x_n), using $h(x) \leq -2x\log(x) \leq 2\sqrt{x}$ for $x \in [0, 1/2]$, we have $x_{n+1} \leq (Cx_n)^{3/2}$ for some constant $C > 0$. Iterating, we get $x_n \leq (C^3 x_0)^{(3/2)^n}$, from which we deduce that $x_n < \frac{1}{N}$ if $n \geq c_2 \log\log N$ for some $c_2 > 0$, if x_0 is small enough.

For the sequence (y_n), using again $h(x) \leq -2x\log(x)$, we have to iterate the relation $y_{n+1} = ay_n \log(\frac{1}{y_n})$, with $a = 2c\frac{\kappa}{\lambda_1}$. If we consider $y_0 \in [\exp(-1/a), \exp(-1)]$, the inductively defined sequence $y_{n+1} = g(y_n)$ is decreasing and converges to $\exp(-1/a)$ since the function $g(x) = -ax\log(x)$ is increasing on the interval $[\exp(-1/a), \exp(-1)]$, $y_1 \leq y_0$ and $\exp(-1/a)$ is the single fixed point of g. Moreover, we have

$$y_{n+2} = a^2 y_n \log\left(\frac{1}{y_n}\right)\left(\log\left(\frac{1}{y_n}\right) + \log\left(\frac{1}{a}\right) + \log\left(\frac{1}{\log(1/y_n)}\right)\right)$$
$$\leq a^2 y_n \log\left(\frac{1}{y_n}\right)\left(\log\left(\frac{1}{y_n}\right) + \log\left(\frac{1}{a}\right)\right),$$

if $y_n \leq 1/e$. By iteration, if we set $b = \log(\frac{1}{\min\{\varrho,a\}})$, we get similarly for all $n \in \mathbb{N}$,

$$y_n \leq a^n y_0 \prod_{i=0}^{n-1}\left[\log\left(\frac{1}{y_0}\right) + i\log\left(\frac{1}{a}\right)\right]$$
$$\leq (ab)^n y_0 n!$$
$$\leq c_3\left(ab\frac{n}{e}\right)^n \sqrt{n}$$
$$= c_3 \exp\left(n\big(\log(a) + \log(b) + \log(n) - 1\big) + \frac{1}{2}\log(n)\right)$$
$$\leq c_3 \exp\left(-c_4 \log N\big(1 + o(1)\big)\right),$$

for some $c_3 > 0$, if $n = c_4 \log(N)/(-\log a - \log\log N)$. In particular, this justifies the hypothesis $\frac{\lambda_1}{\kappa} > c_5 \log N$ for some $c_5 > 1$ large enough. We therefore see that there exists some $c_6 > 0$ such that $e^{-1/a} \leq y_n < \frac{1}{N}$ for $n = c_6 \frac{\log(N)}{\log(\lambda_1/(\kappa\log N))}$.

The third sequence can be rewritten as $z_{n+1} = a z_n (\log \frac{1}{z_n})^{2/3}$, with $a = c(\frac{M}{\lambda_1} \frac{\kappa}{\lambda_1})^{2/3}$. With the same technique as for y_n, we get that the sequence (z_n) converges to $\exp(-1/a^{3/2})$ and $z_n < 1/N$ if

$$n \geq c_7 \log(N) \bigg/ \log\left(\frac{(\lambda_1)^2}{M\kappa \log(N)}\right).$$

This proves the lemma. □

The combination of the previous considerations and Lemma 5.2 then yields the Theorem 3.3.

Appendix: On the degrees of the Erdös–Renyi graph

To prove the Corollary 3.6, we need to estimate the minimum and the maximum degrees of a typical Erdös–Renyi graph $G(N, p)$. The following result could not be found in the literature. We prove in this appendix the following.

Lemma A.1. *If G be is chosen randomly according to the model $G(N, p)$, then if $p \gg \frac{\log N}{N}$, for a set of realizations of G the probability of which converges to one as $N \to \infty$, we have $m = (1 + o(1))Np$ and $\delta = (1 + o(1))Np$.*

Proof. Let G chosen randomly according to the model $G(N, p)$. The law of the degree d_i of an arbitrary vertex i of G is the binomial distribution $B(N, p)$. Hence, using the exponential Markov inequality, we arrive at the following bound: for $p < a < 1$ and $N \geq 1$,

$$P[d_i \geq aN] \leq \exp(-NH(a, p)),$$

where H is the relative entropy or Kullback–Leibler information

$$H(a, p) = a \log\left(\frac{a}{p}\right) + (1 - a) \log\left(\frac{1 - a}{1 - p}\right).$$

If we now set $m = \max_i d_i$ as above, we obtain

$$P[m \geq aN] \leq \sum_{i=1}^{N} P[d_i \geq aN] \leq N \exp(-NH(a, p)).$$

If we choose $a = (1 + \varepsilon)p$, for some $\varepsilon > 0$ such that $a < 1$, we therefore get

$$P\big[m \geq (1+\varepsilon)pN\big] \leq N \exp\left(-N\left(p(1+\varepsilon)\log(1+\varepsilon) + \big(1 - (1+\varepsilon)p\big)\log\left(\frac{1 - (1+\varepsilon)p}{1 - p}\right)\right)\right).$$

Moreover, we have $(1 - (1 + \varepsilon)p) \log(\frac{1-(1+\varepsilon)p}{1-p}) \geq -p\varepsilon$.

Indeed, if we set $q = 1 - p$ and $u = 1 - (1 + \varepsilon)p = q - \varepsilon p$, the last inequality is equivalent to $\log(q/u) \le q/u - 1$ which is true since $q/u > 1$. Thus,

$$P\big[m \ge (1 + \varepsilon)pN\big] \le N \exp\big(-Np\big((1 + \varepsilon)\log(1 + \varepsilon) - \varepsilon\big)\big)$$

$$\le N \exp\left(-Np\frac{\varepsilon^2}{2}(1 + o(1))\right),$$

if we suppose that $\varepsilon = o(1)$ as $N \to \infty$. Choosing $\varepsilon = 2\sqrt{\frac{\log N}{pN}}$, we have $\varepsilon = o(1)$ for $p \gg \frac{\log N}{N}$, and

$$P\big[m \ge (1 + \varepsilon)pN\big] \to 0 \qquad \text{as } N \to \infty.$$

Moreover, we have $m \ge \lambda_1$ and $\lambda_1 = (1 + o(1))pN$ (with probability converging to 1 as $N \to \infty$), which gives the result for m.

Now, if we set $\delta := \min_i d_i$, we want to prove that $P[\delta \ge (1 + \varepsilon')pN] \to 1$, as $N \to \infty$, for some $\varepsilon' = o(1)$. We consider the complementary graph, that is, the random graph $\overline{G}$, such that exactly those edges are missing in a realization of $\overline{G}$, that occur in the corresponding realization of the original random graph G. Now the maximum degree $\overline{m}$ of $\overline{G}$ and the minimum degree δ of G are linked via the relation $\delta = N - 1 - \overline{m}$.

As $\overline{G}$ is chosen randomly according to the model $G(N, 1 - p)$, we have

$$P\big[\overline{m} \ge (1 + \varepsilon)(1 - p)N\big] \le N \exp\big(-NH\big((1 + \varepsilon)(1 - p), 1 - p\big)\big),$$

for all $\varepsilon > 0$ such that $(1 + \varepsilon)(1 - p) < 1$. Now

$$H\big((1 + \varepsilon)(1 - p), 1 - p\big) = (1 + \varepsilon)(1 - p)\log(1 + \varepsilon) + (p - \varepsilon + p\varepsilon)\log\left(1 - \frac{\varepsilon(1 - p)}{p}\right).$$

If we suppose that $\varepsilon = o(1)$ and $\varepsilon \ll p$, using the inequality $\log(1 - x) \ge -x - x^2/2 - x^3$ for $x \in (0, 1/2)$ to bound the last term, we obtain the estimate

$$H\big((1 + \varepsilon)(1 - p), 1 - p\big) \ge \frac{\varepsilon^2}{2p}\left((1 - p) - C\left(\varepsilon + \frac{\varepsilon}{p}\right) + \mathcal{O}(p\varepsilon^2)\right),$$

for some $C > 0$ and

$$P\big[\overline{m} \ge (1 + \varepsilon)(1 - p)N\big] \le \exp\left(-N\frac{\varepsilon^2}{2p}\left((1 - p) - C\left(\varepsilon + \frac{\varepsilon}{p}\right) + \mathcal{O}(p\varepsilon^2)\right) + \log(N)\right).$$

There exists some $c > 0$ such that if we choose $\varepsilon = \sqrt{\frac{4p}{c(1-p)}\frac{\log(N)}{N}}$, we get

$$P\big[\overline{m} \ge (1 + \varepsilon)(1 - p)N\big] \le \exp\left(-cN\frac{\varepsilon^2}{2p}(1 - p) + \log(N)\right) \to 0,$$

under the conditions $p \gg \frac{\log N}{N}$ and $1 - p \gg (\frac{\log N}{N})^{1/3}$.

Finally, we get $\delta \geq N - 1 - (1 + \varepsilon)(1 - p)N = (1 + o(1))Np$, which is the result under these two conditions.

Eventually, we will extend this result for all p such that $p \to 1$, as $N \to +\infty$. As previously, using the exponential Markov inequality, we get the following bound: for $0 < b < p < 1$, and $N \geq 1$,

$$P[d_i \leq bN] \leq \exp\big(-NH(b, p)\big).$$

We set $p = 1 - a_N$ and $b = 1 - b_N$, for some strictly positive sequences (a_N) and (b_N) such that $a_N + b_N \to 0$, as $N \to \infty$, $a_N \ll b_N$, and we can restrict to the case $a_N < (c\frac{\log N}{N})^{1/3}$ for some $c > 0$. We get

$$P[\delta \leq bN] \leq N \exp\left(-N\left((1 - b_N)\log\left(\frac{1 - b_N}{1 - a_N}\right) + b_N \log\left(\frac{b_N}{a_N}\right)\right)\right)$$

$$\leq \exp\left(-N\left(b_N \log\left(\frac{b_N}{a_N}\right) - 2b_N\right) + \log(N)\right).$$

So, we need to choose b_N such that

$$b_N \log\left(\frac{b_N}{a_N}\right) > \frac{\log(N)}{N}.$$

We have

$$b_N \log\left(\frac{b_N}{a_N}\right) > b_N \log\left(b_N\left(\frac{N}{c\log(N)}\right)^{1/3}\right) > \frac{\log(N)}{N},$$

if we choose for instance $b_N = (\frac{\log N}{N})^\gamma$ with $\gamma \in (0, 1/3)$.

Finally, we get for all $p \to 1$ that $\delta \geq (1 - b_N)N = (1 + o(1))Np$, with probability converging to 1 as $N \to \infty$. $\qquad\square$

References

[1] Amit, D.J., Gutfreund, H. and Sompolinsky, H. (1985). Spin-glass models of neural networks. *Phys. Rev. A* (3) **32** 1007–1018. MR0797031

[2] Bassett, D. and Bullmore, E. (2006). Small-world brain networks. *Neuroscientist* **12** 512–523.

[3] Bovier, A. (1999). Sharp upper bounds on perfect retrieval in the Hopfield model. *J. Appl. Probab.* **36** 941–950. MR1737066

[4] Bovier, A. (2006). *Statistical Mechanics of Disordered Systems*: *A Mathematical Perspective*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge: Cambridge Univ. Press. MR2252929

[5] Bovier, A. and Gayrard, V. (1992). Rigorous bounds on the storage capacity of the dilute Hopfield model. *J. Stat. Phys.* **69** 597–627. MR1193852

[6] Bovier, A. and Gayrard, V. (1993). Rigorous results on the thermodynamics of the dilute Hopfield model. *J. Stat. Phys.* **72** 79–112. MR1233027

[7] Burshtein, D. (1994). Nondirect convergence radius and number of iterations of the Hopfield associative memory. *IEEE Trans. Inform. Theory* **40** 838–847. MR1295316

[8] Chen, W., Adams, R., Calcraft, L., Steuber, V. and Davey, N. (2008). Connectivity graphs and the performance of sparse associative memory models. In *IEEE World Congress on Computational Intelligence (Conference IJCNN 2008)* 2742–2749.

[9] Chung, F. and Lu, L. (2002). The average distances in random graphs with given expected degrees. *Proc. Natl. Acad. Sci. USA* **99** 15879–15882 (electronic). MR1944974

[10] Chung, F. and Lu, L. (2002). Connected components in random graphs with given expected degree sequences. *Ann. Comb.* **6** 125–145. MR1955514

[11] Chung, F. and Radcliffe, M. (2011). On the spectra of general random graphs. *Electron. J. Combin.* **18** Paper 215, 14. MR2853072

[12] Diaconis, P. and Stroock, D. (1991). Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Probab.* **1** 36–61. MR1097463

[13] Durrett, R. (2010). *Random Graph Dynamics. Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge: Cambridge Univ. Press. MR2656427

[14] Feige, U. and Ofek, E. (2005). Spectral techniques applied to sparse random graphs. *Random Structures Algorithms* **27** 251–275. MR2155709

[15] Ferrari, P.A., Martínez, S. and Picco, P. (1992). A lower bound for the memory capacity in the Potts–Hopfield model. *J. Stat. Phys.* **66** 1643–1652. MR1156420

[16] Füredi, Z. and Komlós, J. (1981). The eigenvalues of random symmetric matrices. *Combinatorica* **1** 233–241. MR0637828

[17] Gentz, B. and Löwe, M. (1999). The fluctuations of the overlap in the Hopfield model with finitely many patterns at the critical temperature. *Probab. Theory Related Fields* **115** 357–381. MR1725405

[18] Hopfield, J.J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA* **79** 2554–2558. MR0652033

[19] Jacob, E. and Mörters, P. (2012). Spatial preferential attachment networks: Power laws and clustering coefficients. Preprint. Available at arXiv:1210.3830v1.

[20] Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N. and Barabasi, A.L. (2000). The large-scale organization of metabolic networks. *Nature* **407** 651–654.

[21] Komlós, J. and Paturi, R. (1988). Convergence results in an associative memory model. *Neural Networks* **1** 239–250.

[22] Komlós, J. and Paturi, R. (1993). Effect of connectivity in an associative memory model. *J. Comput. System Sci.* **47** 350–373. MR1246181

[23] Krivelevich, M. and Sudakov, B. (2003). The largest eigenvalue of sparse random graphs. *Combin. Probab. Comput.* **12** 61–72. MR1967486

[24] Loukianova, D. (1994). Capacité de mémoire dans le modèle de Hopfield. *C. R. Acad. Sci. Paris Sér. I Math.* **318** 157–160. MR1260330

[25] Löwe, M. (1998). On the storage capacity of Hopfield models with correlated patterns. *Ann. Appl. Probab.* **8** 1216–1250. MR1661188

[26] Löwe, M. and Vermet, F. (2005). The storage capacity of the Blume–Emery–Griffiths neural network. *J. Phys. A* **38** 3483–3503. MR2131428

[27] Löwe, M. and Vermet, F. (2005). The storage capacity of the Hopfield model and moderate deviations. *Statist. Probab. Lett.* **75** 237–248. MR2212355

[28] Löwe, M. and Vermet, F. (2007). The capacity of q-state Potts neural networks with parallel retrieval dynamics. *Statist. Probab. Lett.* **77** 1505–1514. MR2395600

[29] Löwe, M. and Vermet, F. (2011). The Hopfield model on a sparse Erdős–Renyi graph. *J. Stat. Phys.* **143** 205–214. MR2787981

[30] Lu, L. and Peng, X. (2013). Spectra of edge-independent random graphs. *Electron. J. Combin.* **20** 1–18. MR3158266

[31] McEliece, R.J., Posner, E.C., Rodemich, E.R. and Venkatesh, S.S. (1987). The capacity of the Hopfield associative memory. *IEEE Trans. Inform. Theory* **33** 461–482. MR0901677

[32] Newman, C. (1988). Memory capacity in neural network models: Rigorous lower bounds. *Neural Networks* **1** 223–238.

[33] Petritis, D. (1996). Thermodynamic formalism of neural computing. In *Dynamics of Complex Interacting Systems* (*Santiago*, 1994). *Nonlinear Phenom. Complex Systems* **2** 81–146. Dordrecht: Kluwer Academic. MR1421794

[34] Piekniewski, F. and Schreiber, T. (2008). Spontaneous scale-free structure of spike flow graphs in recurrent neural networks. *Neural Networks* **21** 1530–1536.

[35] Rubinov, M., McIntosh, A., Valenzuela, M. and Breakspear, M. (2009). Simulation of neuronal death and network recovery in a computational model of distributed cortical activity. *Am. J. Geriatr. Psychiatry* **17** 210–217.

[36] Talagrand, M. (1998). Rigorous results for the Hopfield model with many patterns. *Probab. Theory Related Fields* **110** 177–276. MR1609015

[37] Talagrand, M. (2003). *Spin Glasses*: *A Challenge for Mathematicians. Cavity and Mean Field Models. Ergebnisse der Mathematik und Ihrer Grenzgebiete. 3 Folge. A Series of Modern Surveys in Mathematics* [*Results in Mathematics and Related Areas*. 3*rd Series. A Series of Modern Surveys in Mathematics*] **46**. Berlin: Springer. MR1993891

[38] van der Hofstadt, R. (2013). Random graphs and complex networks. Lecture notes. Available at http://www.win.tue.nl/~rhofstad/NotesRGCN.pdf.

[39] Zhou, H. and Lipowsky, R. (2005). Dynamic pattern evolution on scale-free networks. *Proc. Natl. Acad. Sci. USA* **102** 10052–10057.