

Enhancing neural-network performance via assortativity

Sebastiano de Franciscis, Samuel Johnson, and Joaquín J. Torres
*Departamento de Electromagnetismo y Física de la Materia, and
Institute Carlos I for Theoretical and Computational Physics,
Facultad de Ciencias, University of Granada, 18071 Granada, Spain.*

The performance of attractor neural networks has been shown to depend crucially on the heterogeneity of the underlying topology. We take this analysis a step further by examining the effect of degree-degree correlations – or assortativity – on neural-network behavior. We make use of a method recently put forward for studying correlated networks and dynamics thereon, both analytically and computationally, which is independent of how the topology may have evolved. We show how the robustness to noise is greatly enhanced in assortative (positively correlated) neural networks, especially if it is the hub neurons that store the information.

PACS numbers: 64.60.aq, 84.35.+i, 89.75.Fb, 87.85.dm

I. BACKGROUND

For a dozen years or so now, the study of complex systems has been heavily influenced by results from network science – which one might regard as the fusion of graph theory with statistical physics [1, 2]. Phenomena as diverse as epidemics [3], cellular function [4], power-grid failures [5] or internet routing [6], among many others [7], depend crucially on the structure of the underlying network of interactions. One of the earliest systems to have been described as a network was the brain, which is made up of a great many neurons connected to each other by synapses [8–11]. Mathematically, the first neural networks combined the Ising model [12] with the Hebb learning rule [13] to reproduce, very successfully, the storage and retrieval of information [14–16]. Neurons were simplified to binary variables (like Ising spins) representing firing or non-firing cells. By considering the trivial fully-connected topology, exact solutions could be reached, which at the time seemed more important than attempting to introduce biological realism. Subsequent work has tended to focus on considering richer dynamics for the individual cells rather than on the way in which these are interconnected [11, 17, 18]. However, the topology of the brain – whether at the level of neurons and synapses, cortical areas or functional connections – is obviously far from trivial [19–24].

The number of neighbors a given node in a network has is called its degree, and much attention is paid to degree distributions since they tend to be highly heterogeneous for most real networks. In fact, they are often approximately scale-free (i.e., described by power laws) [1, 2, 25, 26]. By including this topological feature in a Hopfield-like neural-network model, Torres *et al.* [27] found that degree heterogeneity increases the system's performance at high levels of noise, since the hubs (high degree nodes) are able to retain information at levels well above the usual critical noise. To prove this analytically, the authors considered the *configurational ensemble* of networks (the set of random networks with a given degree distribution but no degree-degree correlations) and

showed that Monte Carlo (MC) simulations were in good agreement with mean-field analysis, despite the approximation inherent to the latter technique when the network is not fully connected. A similar approach can also be used to show how heterogeneity may be advantageous for the performance of certain tasks in models with a richer dynamics [28]. It is worth mentioning that this influence of the degree distribution on dynamical behavior is found in many other settings, such as the more general situation of systems of coupled oscillators [29].

Another property of empirical networks that is quite ubiquitous is the existence of correlations between the degrees of nodes and those of their neighbors [30, 31]. If the average degree-degree correlation is positive the network is said to be *assortative*, while it is called *disassortative* if negatively correlated. Most heterogeneous networks are disassortative [1], which seems to be because this is in some sense their equilibrium (maximum entropy) state given the constraints imposed by the degree distribution [32]. However, there are probably often mechanisms at work which drive systems from equilibrium by inducing different correlations, as appears to be the case for most social networks, in which nodes (people) of a kind tend to group together. This feature, known as *assortativity* or *mixing by degree*, is also relevant for processes taking place on networks. For instance, assortative networks have lower percolation thresholds and are more robust to targeted attack [31], while disassortative ones make for more stable ecosystems and are – at least according to the usual definition – more synchronizable [33].

The approach usually taken when studying correlated networks computationally is to generate a network from the configuration ensemble and then introduce correlations (positive or negative) by some stochastic rewiring process [34]. A drawback of this method, however, is that results may well then depend on the details of this mechanism: there is no guarantee that one is correctly sampling the phase space of networks with given correlations. For analytical work, some kind of hidden variables from which the correlations originate are often considered [35–38] – an assumption which can also be used to

generate correlated networks computationally [37]. This can be a very powerful method for solving specific network models. However, it may not be appropriate if one wishes to consider all possible networks with given degree-degree correlations, independently of how these may have arisen. Here we get round this problem by making use of a method recently suggested by Johnson *et al.* [32] whereby the ensemble of all networks with given correlations can be considered theoretically without recurring to hidden variables. Furthermore, we show how this approach can be used computationally to generate random networks that are representative of the ensemble of interest (i.e., they are model-independent). In this way, we study the effect of correlations on a simple neural network model and find that assortativity increases performance in the face of noise – particularly if it is the hubs that are mainly responsible for storing information (and it is worth mentioning that there is experimental evidence suggestive of a main functional role played by hub neurons in the brain [39, 40]). The good agreement between the mean-field analysis and our MC simulations bears witness both to the robustness of the results as regards neural systems, and to the viability of using this method for studying dynamics on correlated networks.

II. PRELIMINARY CONSIDERATIONS

A. Model neurons on networks

The attractor neural network model put forward by Hopfield [15] consists of N binary neurons, each with an activity given by the dynamic variable $s_i = \pm 1$. Every time step (MCS), each neuron is updated according to the stochastic transition probability $P(s_i \rightarrow \pm 1) = \frac{1}{2}[1 \pm \tanh(h_i/T)]$ (parallel dynamics), where the field h_i is the combined effect on i of all its neighbors, $h_i = \sum_j \hat{w}_{ij} s_j$, and T is a noise parameter we shall call *temperature*, but which represents any kind of random fluctuations in the environment. This is the same as the Ising model for magnetic systems, and the transition rule can be derived from a simple interaction energy such that aligned variables s (spins) contribute less energy than if they were to take opposite values. However, this system can store P given configurations (*memory patterns*) $\xi_i^\nu = \pm 1$ by having the interaction strengths (*synaptic weights*) set according to the Hebb rule [13]: $\hat{w}_{ij} \propto \sum_{\nu=1}^P \xi_i^\nu \xi_j^\nu$. In this way, each pattern becomes an attractor of the dynamics, and the system will evolve towards whichever one is closest to the initial state it is placed in. This mechanism is called *associative memory*, and is nowadays used routinely for tasks such as image identification. What is more, it has been established that something similar to the Hebb rule is implemented in nature via the processes of long-term potentiation and depression at the synapses [41], and this phenomenon is indeed required for learning [42].

To take into account the topology of the network, we

shall consider the weights to be of the form $\hat{w}_{ij} = \hat{\omega}_{ij} \hat{a}_{ij}$, where the element \hat{a}_{ij} of the adjacency matrix represents the number of directed edges (usually interpreted as synapses in a neural network) from node j to node i , while $\hat{\omega}$ stores the patterns, as before:

$$\hat{\omega}_{ij} = \frac{1}{\langle k \rangle} \sum_{\nu=1}^P \xi_i^\nu \xi_j^\nu.$$

For the sake of coherence with previous work, we shall assume \hat{a} to be symmetric (i.e., the network is undirected), so each node is characterized by a single degree $k_i = \sum_j \hat{a}_{ij}$. However, all results are easily extended to directed networks – in which nodes have both an *in* degree, $k_i^{\text{in}} = \sum_j \hat{a}_{ij}$, and an *out* degree, $k_i^{\text{out}} = \sum_j \hat{a}_{ji}$ – by bearing in mind it is only a neuron’s pre-synaptic neighbors that influence its behavior. The mean degree of the network is $\langle k \rangle$, where the angles stand for an average over nodes: $\langle \cdot \rangle \equiv N^{-1} \sum_i (\cdot)$ [43].

B. Network ensembles

When one wishes to consider a set of networks which are randomly wired while respecting certain constraints – that is, an *ensemble* – it is usually useful to define the expected value of the adjacency matrix, $E(\hat{a}) \equiv \hat{\epsilon}$ [44]. The element $\hat{\epsilon}_{ij}$ of this matrix is the mean value of \hat{a}_{ij} obtained by averaging over the ensemble. For instance, in the Erdős-Rényi (ER) ensemble all elements (outside the diagonal) take the value $\hat{\epsilon}_{ij}^{ER} = \langle k \rangle / N$, which is the probability that a given pair of nodes be connected by an edge. For studying networks with a given degree sequence, (k_1, \dots, k_N) , it is common to assume the *configuration ensemble*, defined as

$$\hat{\epsilon}_{ij}^{\text{conf}} = \frac{k_i k_j}{\langle k \rangle N}$$

This expression can usually be applied also when the constraint is a given degree distribution, $p(k)$, by integrating over $p(k_i)$ and $p(k_j)$ where appropriate. One way of deriving $\hat{\epsilon}^{\text{conf}}$ is to assume one has k_i dangling half-edges at each node i ; we then randomly choose pairs of half-edges and join them together until the network is wired up. Each time we do this, the probability that we join i to j is $k_i k_j / (\langle k \rangle N)^2$, and we must perform the operation $\langle k \rangle N$ times. Bianconi showed that this is also the solution for Barabási-Albert evolved networks [46]. However, we should bear in mind that this result is only strictly valid for networks constructed in certain particular ways, such as in these examples. It is often implicitly assumed that were we to average over all random networks with a given degree distribution, the mean adjacency matrix obtained would be $\hat{\epsilon}^{\text{conf}}$. As we shall see, however, this is not necessarily the case [32].

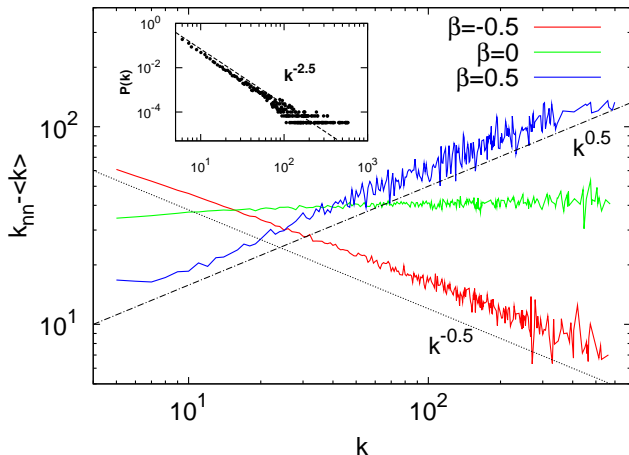


FIG. 1: Mean-nearest-neighbor functions $\bar{k}_{nn}(k)$ for scale-free networks with $\beta = -0.5$ (disassortative), 0.0 (neutral), and 0.5 assortative, generated according to the algorithm described in Sec. III B. Inset: degree distribution (the same in all three cases). Other parameters are $\gamma = 2.5$, $\langle k \rangle = 12.5$, $N = 10^4$.

C. Correlated networks

In the configuration ensemble, the expected value of the mean degree of the neighbors of a given node is $\bar{k}_{nn,i} = k_i^{-1} \sum_j \hat{\epsilon}_{ij}^{conf} k_j = \langle k^2 \rangle / \langle k \rangle$, which is independent of k_i . However, as mentioned above, real networks often display degree-degree correlations, with the result that $\bar{k}_{nn,i} = \bar{k}_{nn}(k_i)$. If $\bar{k}_{nn}(k)$ increases with k , the network is said to be assortative – whereas it is disassortative if it decreases with k (see Fig. 1). This is from the more general nomenclature (borrowed from sociology) in which sets are assortative if elements of a kind group together, or assort. In the case of degree-degree correlated networks, positive assortativity means that edges are more than randomly likely to occur between nodes of a similar degree. A popular measure of this phenomenon is Pearson’s coefficient applied to the edges [1, 2, 31]: $r = ([k_l k_l'] - [k_l]^2) / ([k_l']^2 - [k_l]^2)$, where k_l and k_l' are the degrees of each of the two nodes belonging to edge l , and $[\cdot] \equiv (\langle k \rangle N)^{-1} \sum_l (\cdot)$ is an average over edges.

The ensemble of all networks with a given degree sequence (k_1, \dots, k_N) contains a subset for all members of which $\bar{k}_{nn}(k)$ is constant (the configuration ensemble), but also subsets displaying other functions $\bar{k}_{nn}(k)$. We can identify each one of these subsets (regions of phase space) with an expected adjacency matrix $\hat{\epsilon}$ which simultaneously satisfies the following conditions: **i**) $\sum_j k_j \hat{\epsilon}_{ij} = k_i \bar{k}_{nn}(k_i)$, $\forall i$ (by definition of $\bar{k}_{nn}(k)$), and **ii**) $\sum_j \hat{\epsilon}_{ij} = k_i$, $\forall i$ (for consistency). An ansatz which fulfills these

requirements is any matrix of the form

$$\hat{\epsilon}_{ij} = \frac{k_i k_j}{\langle k \rangle N} + \int d\nu \frac{f(\nu)}{N} \left[\frac{(k_i k_j)^\nu}{\langle k^\nu \rangle} - k_i^\nu - k_j^\nu + \langle k^\nu \rangle \right], \quad (1)$$

where $\nu \in \mathbb{R}$ and the function $f(\nu)$ is in general arbitrary [32]. (If the network were directed, then $k_i = k_i^{\text{in}}$ and $k_j = k_j^{\text{out}}$ in this expression.) This ansatz yields

$$\bar{k}_{nn}(k) = \frac{\langle k^2 \rangle}{\langle k \rangle} + \int d\nu f(\nu) \sigma_{\nu+1} \left[\frac{k^{\nu-1}}{\langle k^\nu \rangle} - \frac{1}{k} \right] \quad (2)$$

(the first term being the result for the configuration ensemble), where $\sigma_{b+1} \equiv \langle k^{b+1} \rangle - \langle k \rangle \langle k^b \rangle$. To prove the uniqueness of a matrix $\hat{\epsilon}$ obtained in this way (i.e., that it is the only one compatible with a given $\bar{k}_{nn}(k)$) assume that there exists another valid matrix $\hat{\epsilon}' \neq \hat{\epsilon}$. Writing $\hat{\epsilon}'_{ij} - \hat{\epsilon}_{ij} \equiv h(k_i, k_j) = h_{ij}$, then Condition **i**) implies that $\sum_j k_j h_{ij} = 0$, $\forall i$, while Condition **ii**) means that $\sum_j h_{ij} = 0$, $\forall i$. It follows that $h_{ij} = 0$, $\forall i, j$. This means that $\hat{\epsilon}$ is not just one possible way of obtaining correlations according to $\bar{k}_{nn}(k)$; rather, there is a two-way mapping between $\hat{\epsilon}$ and $\bar{k}_{nn}(k)$: every network with this particular function $\bar{k}_{nn}(k)$ and no other ones are contained in the ensemble defined by $\hat{\epsilon}$. Thanks to this, if we are able to consider random networks drawn according to this matrix (whether we do this analytically or computationally; see Section III B), we can be confident that we are correctly taking account of the whole ensemble of interest. In other words, whatever the reasons behind the existence of degree-degree correlations in a given network, we can study the effects of these with only information on $p(k)$ and $\bar{k}_{nn}(k)$ by obtaining the associated matrix $\hat{\epsilon}$. This is not to say, of course, that all topological properties are captured in this way: a particular network may have other features – such as higher order correlations, modularity, etc. – the consideration of which would require concentrating on a sub-partition of those with the same $p(k)$ and $\bar{k}_{nn}(k)$. But this is not our purpose here.

In many empirical networks, $\bar{k}_{nn}(k)$ has the form $\bar{k}_{nn}(k) = A + Bk^\beta$, with $A, B > 0$ [2, 30] – the mixing being assortative if β is positive, and disassortative when negative. Such a case is fitted by Eq. (2) if

$$f(\nu) = C \left[\frac{\sigma_2}{\sigma_{\beta+2}} \delta(\nu - \beta - 1) - \delta(\nu - 1) \right], \quad (3)$$

with C a positive constant, since this choice yields

$$\bar{k}_{nn}(k) = \frac{\langle k^2 \rangle}{\langle k \rangle} + C \sigma_2 \left[\frac{k^\beta}{\langle k^{\beta+1} \rangle} - \frac{1}{\langle k \rangle} \right]. \quad (4)$$

Johnson *et al.* [32] obtained the entropy of ensembles of networks with scale-free degree distributions ($p(k) \sim k^{-\gamma}$) and correlations given by Eq. (4), and found that

the most likely configurations (those maximizing the entropy) generally correspond to correlated networks. In particular, the expected mixing, all other things being equal, is usually a certain degree of disassortativity – which explains the predominance of these networks in the real world. They also showed that the maximum entropy is usually obtained for values of C close to one. Here, we shall use this result to justify concentrating on correlated networks with $C = 1$, so that the only parameter we need to take into account is β . It is worth mentioning that Pastor-Satorras *et al.* originally suggested using this exponent as a way of quantifying correlations [30], since this seems to be the most relevant magnitude. Because β does not depend directly on $p(k)$ (as r does), and can be defined for networks of any size (whereas r , in very heterogeneous networks, always goes to zero for large N due to its normalization [47]), we shall henceforth use β as our assortativity parameter.

So, after plugging Eq. (3) into Eq. (1), we find that the ensemble of networks exhibiting correlations given by Eq. (4) (and $C = 1$) is defined by the mean adjacency matrix

$$\hat{\epsilon}_{ij} = \frac{1}{N}[k_i + k_j - \langle k \rangle] + \frac{\sigma_2}{\sigma_{\beta+2}} \frac{1}{N} \left[\frac{(k_i k_j)^{\beta+1}}{\langle k^{\beta+1} \rangle} - k_i^{\beta+1} - k_j^{\beta+1} + \langle k^{\beta+1} \rangle \right] \quad (5)$$

III. ANALYSIS AND RESULTS

A. Mean field

Let us consider the single-pattern case ($P = 1$, $\xi_i = \xi_i^1$). Substituting the adjacency matrix \hat{a} for its expected value $\hat{\epsilon}$ (as given by Eq. (5)) in the expression for the local field at i – which amounts to a mean-field approximation – we have

$$h_i = \frac{1}{\langle k \rangle} \xi_i \left\{ \left[(k_i - \langle k \rangle) + \frac{\sigma_2}{\sigma_{\beta+2}} (\langle k^{\beta+1} \rangle - k_i^{\beta+1}) \right] \mu_0 + \langle k \rangle \mu_1 + \frac{\sigma_2}{\sigma_{\beta+2}} (k_i^\beta - \langle k^{\beta+1} \rangle) \mu_{\beta+1} \right\},$$

where we have defined

$$\mu_\alpha \equiv \frac{\langle k_i^\alpha \xi_i s_i \rangle}{\langle k_i^\alpha \rangle}$$

for $\alpha = 0, 1, \beta + 1$. These order parameters measure the extent to which the system is able to recall information in spite of noise [28]. For the first order we have $\mu_0 = m \equiv \langle \xi_i s_i \rangle$, the standard overlap measure in neural networks (analogous to magnetization in magnetic systems), which takes account of memory performance. However, μ_1 , for instance, weighs the sum with the degree of each node, with the result that it measures information per synapse instead of per neuron. Although the overlap m is often assumed to represent, in some sense, the *mean*

firing rate of neurological experiments, it is possible that μ_1 is more closely related to the empirical measure, since the total electric potential in an area of tissue is likely to depend on the number of synapses transmitting action potentials. In any case, a comparison between the two order parameters is a good way of assessing to what extent the performance of neurons depends on their degree – larger-degree model neurons can in general store information at higher temperatures than ones with smaller degree can [27].

Substituting s_i for its expected value according to the transition probability, $s_i \rightarrow \tanh(h_i/T)$, we have, for any α ,

$$\langle k_i^\alpha \xi_i s_i \rangle = \langle k_i^\alpha \xi_i \tanh(h_i/T) \rangle;$$

or, equivalently, the following 3-D map of closed coupled equations for the macroscopic overlap observables μ_0 , μ_1 and $\mu_{\beta+1}$ – which describes, in this mean-field approximation, the dynamics of the system:

$$\begin{aligned} \mu_0(t+1) &= \int p(k) \tanh[F(t)/(\langle k \rangle T)] dk \\ \mu_1(t+1) &= \frac{1}{\langle k \rangle} \int p(k) k \tanh[F(t)/(\langle k \rangle T)] dk \\ \mu_{\beta+1}(t+1) &= \frac{1}{\langle k^{\beta+1} \rangle} \int p(k) k^{\beta+1} \tanh[F(t)/(\langle k \rangle T)] dk, \end{aligned} \quad (6)$$

with

$$\begin{aligned} F(t) &\equiv (k \mu_0(t) + \langle k \rangle \mu_1(t) - \langle k \rangle \mu_0(t)) \\ &+ \frac{\sigma_2}{\sigma_{\beta+2}} [k^{\beta+1} (\mu_{\beta+1}(t) - \mu_0(t)) \\ &+ \langle k^{\beta+1} \rangle (\mu_0(t) - \mu_{\beta+1}(t))]. \end{aligned}$$

This can be easily computed for any degree distribution $p(k)$. Note that taking $\beta = 0$ (the uncorrelated case) the system collapses to the 2-D map obtained in Ref. [27], while it becomes the typical 1-D case for a homogeneous $p(k)$ – say a fully-connected network [15]. It is in principle possible to do similar mean-field analysis for any number P of patterns, but the map would then be 3P-dimensional, making the problem substantially more complex.

At a critical temperature T_c , the system will undergo the characteristic second order phase transition from a phase in which it exhibits memory (akin to ferromagnetism) to one in which it does not (paramagnetism). To obtain this critical temperature, we can expand the hyperbolic tangent in Eqs. (6) around the trivial solution $(\mu_0, \mu_1, \mu_{\beta+1}) \simeq (0, 0, 0)$ and, keeping only linear terms, write

$$\begin{aligned}
\mu_0 &= \mu_1/T_c, \\
\mu_1 &= \frac{1}{\langle k \rangle^2 T_c} [\langle k \rangle^2 \mu_1 + \sigma_2 \mu_{\beta+1}], \\
\mu_{\beta+1} &= \frac{1}{T_c \langle k \rangle \langle k^{\beta+1} \rangle} \left[\sigma_{\beta+2} \mu_0 \right. \\
&\quad + \frac{\sigma_2}{\sigma_{\beta+2}} \left(\langle k^{\beta+1} \rangle^2 - \langle k^{2(\beta+1)} \rangle \right) \mu_0 \\
&\quad \left. + \langle k \rangle \langle k^{\beta+1} \rangle \mu_1 - \frac{\sigma_2}{\sigma_{\beta+2}} \left(\langle k^{\beta+1} \rangle^2 - \langle k^{2(\beta+1)} \rangle \right) \mu_{\beta+1} \right].
\end{aligned}$$

Defining

$$\begin{aligned}
A &\equiv \frac{\sigma_2}{\langle k \rangle^2}, \\
B &\equiv \frac{\sigma_2}{\sigma_{\beta+2}} \frac{\langle k^{2(\beta+1)} \rangle - \langle k^{\beta+1} \rangle^2}{\langle k \rangle \langle k^{\beta+1} \rangle}, \\
D &\equiv \frac{\sigma_{\beta+2}}{\langle k \rangle \langle k^{\beta+1} \rangle},
\end{aligned}$$

T_c will be the solution to the third order polynomial equation:

$$T_c^3 - (B+1)T_c^2 + (B-A)T_c + A(B-D) = 0. \quad (7)$$

Note that for neutral (i.e., uncorrelated) networks, $\beta = 0$, and so $A = B = D$. We then have $T_c = \langle k^2 \rangle / \langle k \rangle^2$, as expected [28].

B. Generating correlated networks

Given a degree distribution $p(k)$, the ensemble of networks compatible with this constraint and with degree-degree correlations according to Eq. (4) (with some exponent β) is defined by the mean adjacency matrix $\hat{\epsilon}$ of Eq. (5) – as described in Section II C and in Ref. [32]. Therefore, although there will generally be an enormous number of possible networks in this volume of phase space, we can sample them correctly simply by generating them according to $\hat{\epsilon}$. To do this, first we have to assign to each node a degree drawn from $p(k)$. If the elements of $\hat{\epsilon}$ were probabilities, it would suffice then to connect each pair of nodes (i, j) with probability $\hat{\epsilon}_{ij}$ to generate a valid network. Strictly speaking, $\hat{\epsilon}$ is an expected value, which in certain cases can be greater than one. To get round this, we write a probability matrix $\hat{p} = \hat{\epsilon}/a$ with a some value such that all elements of \hat{p} are smaller than one. If we then take random pairs of nodes (i, j) and, with probability \hat{p}_{ij} , place an edge between them, repeating the operation until $\frac{1}{2}\langle k \rangle N$ edges have been placed, the expected value of edges joining i and j will be $\hat{\epsilon}_{ij}$. This method is like the *hidden variable* technique [37] in that edges are placed with a predefined probability (which is why the resulting ensemble is canonical). The difference lies in the fact that in the method here described correlations only depend on the degrees of nodes.

We are interested here in neural networks, in which a given pair of nodes can be joined by several synapses, so we shall not impose the restriction of so-called simple networks of allowing only one edge at most per pair. We shall, however, consider networks with a *structural cutoff*: $k_i < \sqrt{\langle k \rangle N}$, $\forall i$ [48]. This ensures that, at least for $\beta \leq 0$, all elements of $\hat{\epsilon}$ are indeed smaller than one.

Because we can expect effects due to degree-degree correlations to be largest when $p(k)$ is very broad, and since most networks in nature and technology seem to exhibit approximately power-law degree distributions [1, 7, 25, 26], we shall here test our general theoretical results against simulations of scale-free networks: $p(k) \sim k^{-\gamma}$. This means that a network (or the region of phase space to which it belongs) is characterized by the set of parameters $\{\langle k \rangle, N, \gamma, \beta\}$.

C. Assortativity and dynamics

In Fig. 2 we plot the stationary value of μ_1 against the temperature T , as obtained from simulations and Eqs. (6), for disassortative, neutral and assortative networks. The three curves are similar at low temperatures, but as T increases their behavior becomes quite different. The disassortative network is the least robust to noise. However, the assortative one is capable of retaining some information at temperatures considerably higher than the critical value, $T_c = \langle k^2 \rangle / \langle k \rangle$, of neutral networks. A comparison between μ_1 and μ_0 (see Fig. 3) shows that it is the high degree nodes that are mainly responsible for this difference in performance. This can be seen more clearly in Fig. 4, which displays the difference $\mu_1 - \mu_0$ against T for the same networks. It seems that, because in an assortative network a sub-graph of hubs will have more edges than in a disassortative one, it has a higher effective critical temperature. Therefore, even when most of the nodes are acting randomly, the set of nodes of sufficiently high degree nevertheless displays associative memory.

The phase diagram in Fig. 5 shows the critical temperature, T_c , as obtained from Eq. (7). In addition to the effect reported in Ref. [27] whereby the T_c of scale-free networks grows with degree heterogeneity (decreasing γ), it also increases very significantly with positive degree-degree correlations (increasing β).

At large values of N , the critical temperature scales as $T_c \sim N^b$, with $b \geq 0$ a constant. However, because the moments of k appearing in the coefficients of Eq. (7) can have different asymptotic behavior depending on the values of γ and β , the scaling exponent b differs from one region to another in the space of these parameters. These are the seven regions shown in Fig. 6, along with the scaling behavior exhibited by each one. This can be seen explicitly in Fig. 7, where T_c , as obtained from MC simulations, is plotted against N for cases in each of the regions with $\gamma < 3$. In each case, the scaling is as given by Eq. (7) and shown in Fig. 6. For the four regions with $\gamma < 3$, from lowest to highest assortativity

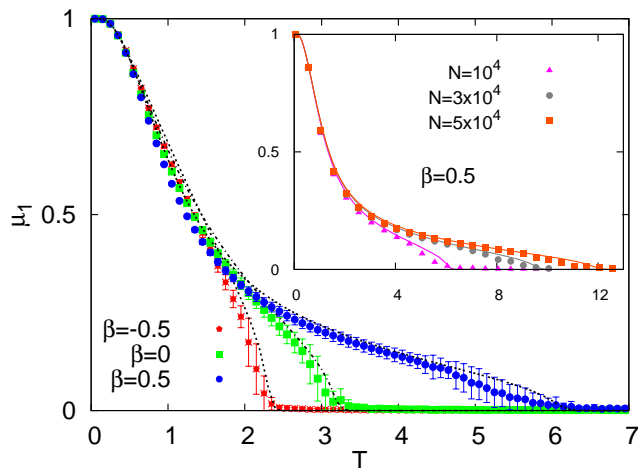


FIG. 2: Stable stationary value of the weighted overlap μ_1 against temperature T for scale-free networks with correlations according to $\bar{k}_{nn} \sim k^\beta$, for $\beta = -0.5$ (disassortative), 0.0 (neutral), and 0.5 (assortative). Symbols from MC simulations, with errorbars representing standard deviations, and lines from Eqs. (6). Other network parameters as in Fig. 1. Inset: μ_1 against T for the assortative case ($\beta = 0.5$) and different system sizes: $N = 10^4$, $3 \cdot 10^4$ and $5 \cdot 10^4$.

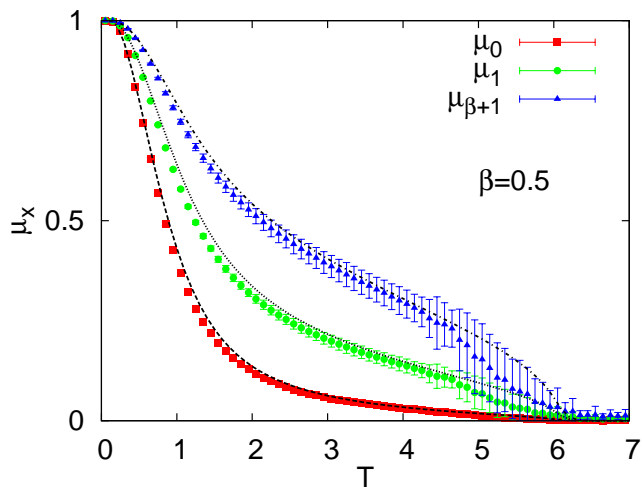


FIG. 3: Stable stationary values of order parameters μ_0 , μ_1 and $\mu_{\beta+1}$ against temperature T , for assortative networks according to $\beta = 0.5$. Symbols from MC simulations, with errorbars representing standard deviations, and lines from Eqs. (6). Other parameters as in Fig. 1.

we have scaling exponents which are dependent on: only γ (region I), only β (region II), both γ and β (region III), and, perhaps most interestingly, neither of the two (region IV) – with T_c scaling, in the latter case, as \sqrt{N} . As for the more homogeneous $\gamma > 3$ part, regions V and VI have a diverging critical temperature despite the fact that the second moment of $p(k)$ is finite, simply as a

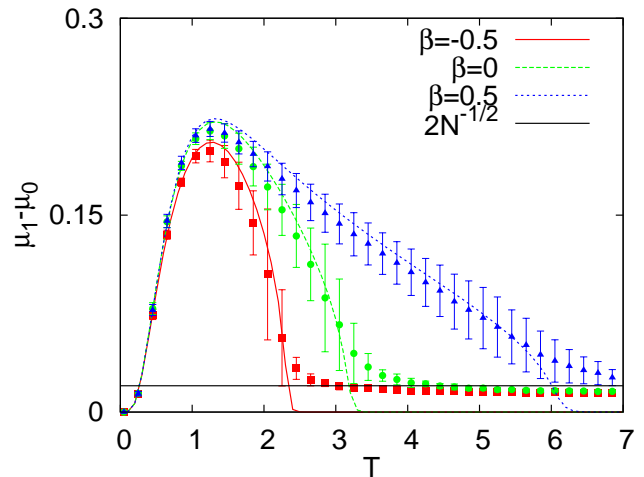


FIG. 4: Difference between the stationary values μ_1 and μ_0 for networks with $\beta = -0.5$ (disassortative), 0.0 (neutral) and 0.5 (assortative), against temperature. Symbols from MC simulations, with errorbars representing standard deviations, and lines from Eqs. (6). Line shows the expected level of fluctuations due to noise, $\sim N^{-\frac{1}{2}}$. Other parameters as in Fig. 1.

result of assortativity.

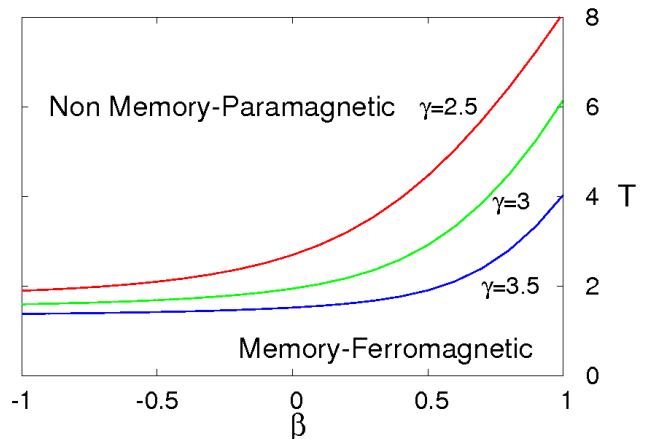


FIG. 5: Phase diagrams for scale-free networks with $\gamma = 2.5$, 3, and 3.5. Lines show the critical temperature T_c marking the second-order transition from a memory (ferromagnetic) phase to a memoryless (paramagnetic) one, against the assortativity β , as given by Eq. (7). Other parameters as in Fig. 1.

IV. DISCUSSION

We have shown that assortative networks of simple model neurons are able to exhibit associative memory

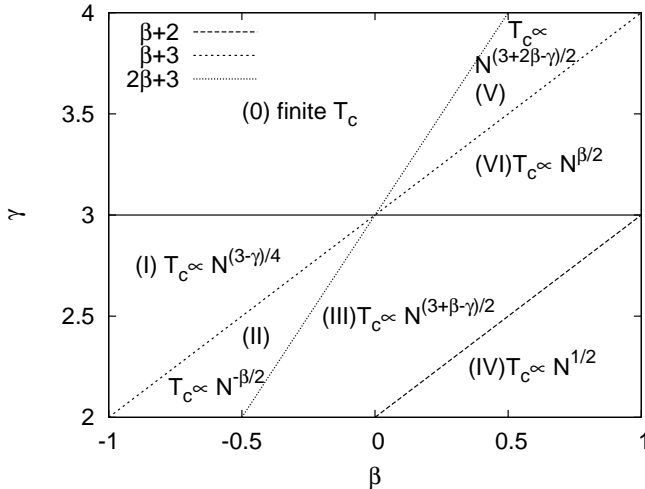


FIG. 6: Parameter space $\beta - \gamma$ partitioned into the regions in which $b(\beta, \gamma)$ has the same functional form – where b is the scaling exponent of the critical temperature: $T_c \sim N^b$. Exponents obtained by taking the large N limit in Eq. (7).

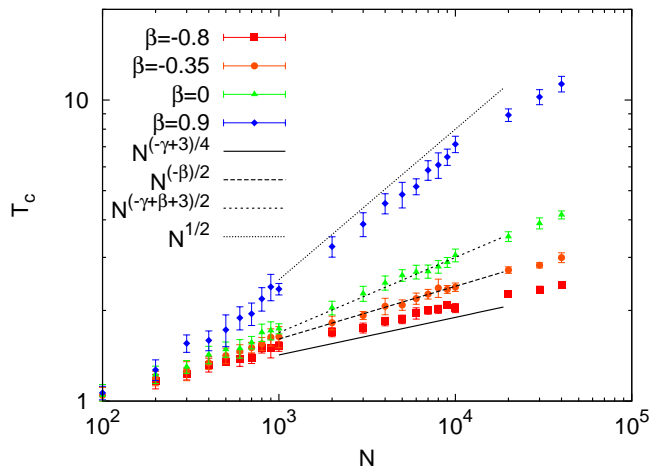


FIG. 7: Examples of how T_c scales with N for networks belonging to regions I, II, III and IV of Fig. 6 ($\beta = -0.8, -0.35, 0.0$ and 0.9 , respectively). Symbols from MC simulations, with errorbars representing standard deviations, and slopes from Eq. (7). All parameters – except for β and N – are as in Fig. 1.

in the presence of levels of noise such that uncorrelated (or disassortative) networks cannot. This may appear to be in contradiction with a recent result obtained using spectral graph analysis – that synchronizability of a set of coupled oscillators is highest for disassortative networks [33]. A synchronous state of model oscillators and a memory phase of model neurons are both sets of many simple dynamical elements coupled via a network in such a way that a macroscopically coherent situation is main-

tained [29]. Obviously both systems require the effective transmission of information among the elements. So why are opposite results as regards the influence of topology reported for each system? The answer is simple: whereas the definition of a synchronous state is that every single element oscillate at the same frequency, it is precisely when most elements are actually behaving randomly that the advantages to assortativity we report become apparent. In fact, it can be seen in Fig. 2 that at low temperatures disassortative networks perform the best, although the effect is small. This is reminiscent of percolation: at high densities of edges the giant component is larger in disassortative networks, but in assortative ones a non-vanishing fraction of nodes remain interconnected even at densities below the usual percolation threshold [31]. Because in the case of targeted attacks it is this threshold which is taken as a measure of resilience, we say that assortative networks perform the best. In general, the optimal network for good conditions (i.e., complete synchronization, high density of edges, low levels of noise) is not necessarily the one which performs the best in bad conditions (partial synchronization, low density of edges, high levels of noise). It seems that optimality – whether in resilience or robustness – should thus be defined for particular conditions.

We have used the technique suggested in Ref. [32] to study the effect of correlations on networks of model neurons, but many other systems of dynamical elements should be susceptible to a similar treatment. In fact, Ising spins [46], Voter Model agents [49], or Boolean nodes [50], for instance, are similar enough to binary neurons that we should expect similar results for these models. If a moral can be drawn, it is that persistence of partial synchrony, or coherence of a subset of highly connected dynamical elements, can sometimes be as relevant (or more so) as the possibility of every element behaving in the same way. In the case of real brain cells, experiments suggest that hub neurons play key functional roles [39, 40]. From this point of view, there may be a selective pressure for brain networks to become assortative – although, admittedly, this organ engages in such complex behavior that there must be many more functional constraints on its structure than just a high robustness to noise. Nevertheless, it would be interesting to investigate this aspect of biological systems experimentally. For this, it should be borne in mind that heterogeneous networks have a natural tendency to become disassortative, so it is against the expected value of correlations discussed in Ref. [32] that empirical data should be contrasted in order to look for meaningful deviations towards assortativity.

Acknowledgments

This work was supported by Junta de Andalucía

projects FQM-01505 and P09-FQM4682, and by Spanish MEC-FEDER project FIS2009-08451. Many thanks to Omar Al Hammal for fruitful suggestions, and to An-

tani Tarapiatapioca for inspiration as well as practical advice.

-
- [1] M.E.J. Newman, “The structure and function of complex networks”, *SIAM Reviews* **45**, 167 (2003)
- [2] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, “Complex networks: structure and dynamics”, *Phys. Rep.* **424**, 175 (2006)
- [3] D.J. Watts and S.H. Strogatz, “Collective dynamics of ‘small-world’ networks” *Nature* **395**, 440–2 (1998)
- [4] G.M. Süel, J. Garcia-Ojalvo, L.M. Liberman, and M.B. Elowitz “An excitable gene regulatory circuit induces transient cellular differentiation”, *Nature* **440**, 545–50 (2006)
- [5] S.V. Buldyrev, R. Parshani, G. Paul, H.E. Stanley, and S. Havlin, “Catastrophic cascade of failures in interdependent networks”, *Nature* **464**, 1025–8 (2010)
- [6] M. Boguñá, F. Papadopoulos, and D. Krioukov, “Sustaining the Internet with hyperbolic mapping”, *Nature Communications* **1** (2010)
- [7] A. Arenas, A. Díaz-Guilera, J. Kurths, Y. Moreno and C. Zhou, “Synchronization in complex networks”, *Phys. Rep.* **469**, 93-153 (2008)
- [8] S. Ramón y Cajal, *Histology of the Nervous System of Man and Vertebrates*, Oxford University Press, 1995.
- [9] D.J. Amit, *Modeling Brain Function*, Cambridge Univ. Press, Cambridge, 1989
- [10] L.F. Abbott and T.B. Kepler, “From Hodgkin-Huxley to Hopfield”, *Statistical mechanics of neural networks*, Springer-Verlag, Berlin, 1990
- [11] J.J. Torres, and P. Varona, “Modeling Biological Neural Networks”, *Handbook of Natural Computing*, Springer-Verlag, Berlin, 2010
- [12] R.J. Baxter, *Exactly Solved Models in Statistical Mechanics*, Academic Press, London, 1982
- [13] D.O. Hebb, *The Organization of Behavior*, Wiley, New York, 1949
- [14] S. Amari, “Characteristics of random nets of analog neuron-like elements”, *IEEE Trans. Syst. Man. Cybern.*, **2** 643–657 (1972)
- [15] J.J. Hopfield, “Neural networks and physical systems with emergent collective computational abilities”, *Proc. Natl. Acad. Sci. USA* **79** 2554–8 (1982)
- [16] D.J. Amit, “The Hebbian paradigm reintegrated: Local reverberations as internal representations”, *Behavioral and Brain Sciences* **18**, 617-657 (1995)
- [17] T.P. Vogels, K. Rajan, and L.F. Abbott “Neural network dynamics”, *Annu. Rev. Neurosci.* **28** 357–76 (2005)
- [18] J.J. Torres, J. M. Cortes, J. Marro, and H.J. Kappen “Competition between synaptic depression and facilitation in attractor neural networks”, *Neural Computation*, **19**, 2739–55 (2007)
- [19] L.A.N. Amaral, A. Scala, M. Barthélemy, and H.E. Stanley, “Classes of small-world networks”, *Proc. Natl. Acad. Sci. USA* **97** 11149–52 (2000)
- [20] O. Sporns, D.R. Chialvo, M. Kaiser and C.C. Hilgetag, “Organization, development and function of complex brain networks”, *Trends Cogn. Sci.* **8** 418–25 (2004)
- [21] V.M. Eguíluz, D.R. Chialvo, G.A. Cecchi, M. Baliki, and A. V. Apkarian, “Scale-free brain functional networks”, *Phys. Rev. Lett.* **94**, 018102 (2005)
- [22] A. Arenas, A. Fernández and S. Gómez, “A complex network approach to the determination of functional groups in the neural system of *C. Elegans*”, *Lect. Notes Comp. Sci.* **5151**, 9-18 (2008)
- [23] E Bullmore, O Sporns “Complex brain networks: graph theoretical analysis of structural and functional systems”, *Nature Reviews Neuroscience* **10**, 186-198 (2009)
- [24] S. Johnson, J. Marro, and J.J. Torres, “Evolving networks and the development of neural systems”, *J. Stat. Mech.* (2010) P03003
- [25] P. Peretto “An Introduction to the Modeling of Neural Networks”, Cambridge Univ. Press, 1992
- [26] A.L. Barabási and Z.N. Oltvai “Network biology: understanding the cell’s functional organization”, *Nature Reviews Genetics* **5**, 101–3 (2004)
- [27] J.J. Torres, M.A. Muñoz, J. Marro, and P.L. Garrido, “Influence of topology on a neural network performance”, *Neurocomputing*, **58-60** 229-234 (2004)
- [28] S. Johnson, J. Marro, and J.J. Torres, “Functional optimization in complex excitable networks”, *EPL* **83**, 46006 (2008).
- [29] M. Barahona and L.M. Pecora, “Synchronization in small-world systems”, *Phys. Rev. Lett.* **89**, 054101 (2002)
- [30] R. Pastor-Satorras, A. Vázquez, and A. Vespignani, *Phys. Rev. Lett.* **87**, 258701 (2001)
- [31] M.E.J. Newman, “Mixing patterns in networks”, *Phys. Rev. Lett.*, **89**, 208701 (2002); “Assortative mixing in networks”, *Phys. Rev. E* **67**, 026126 (2003)
- [32] S. Johnson, J.J. Torres, J. Marro, and M.A. Muñoz, “Entropic origin of disassortativity in complex networks”, *Phys. Rev. Lett.* **104**, 108702 (2010)
- [33] M. Brede and S. Sinha, “Assortative mixing by degree makes a network more unstable”, arXiv:cond-mat/0507710
- [34] S. Maslov, K. Sneppen, A. Zaliznyak, “Pattern detection in complex networks: correlation profile of the Internet”, *Physica A* **333**, 529-40 (2004)
- [35] G. Caldarelli, A. Capocci, P. De Los Rios, and M.A. Muñoz, “Scale-free networks from varying vertex intrinsic fitness”, *Phys. Rev. Lett.* **89**, 258702 (2002)
- [36] B. Söderberg, “General formalism for inhomogeneous random graphs”, *Phys. Rev. E* **66**,066121 (2002)
- [37] M. Boguñá and R. Pastor-Satorras, “Class of correlated random networks with hidden variables”, *Phys. Rev. E* **68**, 036112 (2003)
- [38] A. Fronczak and P. Fronczak, “Networks with given two-point correlations: hidden correlations from degree correlations”, *Phys. Rev. E* **74**, 026121 (2006)
- [39] R.J. Morgan and I Soltesz, “Nonrandom connectivity of the epileptic dentate gyrus predicts a major role for neuronal hubs in seizures” *PNAS* **105**, 6179 (2008)
- [40] P. Bonifazi, M. Goldin, M.A. Picardo, I. Jorquera, A. Cattani, G. Bianconi, A. Represa, Y. Ben-Ari, and R. Cossart, “GABAergic hub neurons orchestrate synchrony

- in developing hippocampal networks”, *Science* **326**, 1419 (2009)
- [41] O. Paulsen and T.J. Sejnowski, “Natural patterns of activity and long-term synaptic plasticity”, *Curr. Opin. Neurobiol.* **10** 172–9 (2000)
- [42] A. Gruart, M.D. Muñoz, and J.M. Delgado-García, “Involvement of the CA3-CA1 synapse in the acquisition of associative learning in behaving mice”, *J. Neurosci.* **26**, 1077–87 (2006).
- [43] In directed networks the mean *in* degree and the mean *out* degree necessarily coincide, whatever the forms of the *in* and *out* distributions.
- [44] As in statistical physics, one can consider the *micro-canonical* ensemble, in which each element (network) satisfies the constraints exactly, or the *canonical* ensemble, where the constraints are satisfied on average [45]. Throughout this work, we shall refer to canonical ensembles.
- [45] G. Bianconi, “Entropy of network ensembles”, *Phys. Rev. E* **79**, 036114 (2009)
- [46] G. Bianconi, “Mean-field solution of the Ising model on a Barabasi-Albert network”, *Phys. Lett. A* **303** 166–8 (2002)
- [47] S.N. Dorogovtsev, A.L. Ferreira, A.V. Goltsev, and J.F.F. Mendes, “Zero pearson coefficient for strongly correlated growing trees”, *Phys. Rev. E* **81**, 031135 (2010)
- [48] G. Bianconi, “The entropy of randomized network ensembles”, *Europhys. Lett.* **81**, 28005 (2008)
- [49] K. Suchecki, V.M. Eguíluz, and M. San Miguel “Conservation laws for the voter model in complex networks”, *EPL*, **69**, 228–34 (2005)
- [50] T.P. Peixoto, “Redundancy and error resilience in Boolean Networks”, *Phys. Rev. Lett.* **104**, 048701 (2010)