Topological properties of hierarchical networks

Elena Agliari,1 Adriano Barra,1 Andrea Galluzzi,2 Francesco Guerra,1 Daniele Tantari,2 and Flavia Tavani3

1 Dipartimento di Fisica, Sapienza Università di Roma, P.le A. Moro 2, 00185, Roma, Italy.
2 Dipartimento di Matematica, Sapienza Università di Roma, P.le Aldo Moro 2, 00185, Roma, Italy.
3 Dipartimento SBAI (Ingegneria), Sapienza Università di Roma, P.le Aldo Moro 2, 00185, Roma, Italy.

(Dated: December 19, 2014)

Hierarchical networks are attracting a renewal interest for modelling the organization of a number of biological systems and for tackling the complexity of statistical mechanical models beyond mean-field limitations. Here we consider the Dyson hierarchical construction for ferromagnets, neural networks and spin-glasses, recently analyzed from a statistical-mechanics perspective, and we focus on the topological properties of the underlying structures. In particular, we find that such structures are weighted graphs that exhibit high degree of clustering and of modularity, with small spectral gap; the robustness of such features with respect to link removal is also studied. These outcomes are then discussed and related to the statistical mechanics scenario in full consistency. Lastly, we look at these weighted graphs as Markov chains and we show that in the limit of infinite size, the emergence of ergodicity breakdown for the stochastic process mirrors the emergence of meta-stabilities in the corresponding statistical mechanical analysis.

PACS numbers:

I. INTRODUCTION

When dealing with statistical-mechanics models (e.g. spin systems) overcoming the mean-field approximation is extremely challenging. Basically, the mean-field approximation lies in assuming that each spin $S_i$ ($i = 1, ..., N$) in an embedding space does interact with all the other $N - 1$ spins with the same strength, notwithstanding their mutual distance, as if spins occupied the $N$ vertices of an hyper-tetrahedron. As a notion of distance is introduced and couplings among spins are accordingly rescaled, the exact solution is, in most cases, out of reach.

In the 60’s, a hierarchical model for ferromagnetic systems was introduced to describe non-mean-field spin systems [14], and is known as the Hierarchical Ferromagnet. More recently, also the Sherrington-Kirkpatrick model for spin-glasses [9, 11, 12] and the Hopfield model for neural networks [1, 3] defined on such a hierarchical topology have been investigated.

The hierarchical network exploited in all these cases is endowed with a metric and it is explicitly not-mean-field since the coupling between two nodes at a distance $d$ scales as $d^{-\rho}$, where $\rho$ is a proper tuneable parameter. As a result, the spins can be thought of as placed on the vertices of a fully-connected weighted graph, where the coupling pattern mirrors the mutual distance among spins. This graph exhibits peculiar features (e.g., high degree of modularity), which play a crucial role in the statistical-mechanics treatability as well as in the emergent behavior of the above mentioned models. Also, the knowledge of the specific architecture considered allows to figure out the class of real-world systems where theoretical results can properly be applied. However, only marginal attention has been devoted to such topological properties in the past and in this work we just aim to deepen these aspects.

In the following we first provide a streamlined and general introduction to the statistical-mechanics models considered (i.e., the hierarchical ferromagnet, the hierarchical neural network and the hierarchical spin glass), then we move to the analysis of the underlying network by studying the degree of clustering, the modularity, the ergodicity and the spectral properties. Finally, a section with outlooks and conclusions closes the paper.

II. DEFINITION OF MODELS AND RELATED HAMILTONIANS

The three statistical-mechanics models which we adapt to live on a hierarchical network are the Curie-Weiss model, the Hopfield model and the Sherrington-Kirkpatrick model, which are the prototypes for ferromagnetism, associative neural networks and spin-glasses, respectively.

Before providing the Hamiltonians of these models when defined in a hierarchical structure, we outline how they can be built up recursively. One starts from a set of two spins properly coupled (the kind of coupling depending on the particular model considered). Then, one takes two of such dimers and makes two operations: update the existing links and introduce new links to couple spins belonging to different dimers. This constitutes the system at the first iteration. At the next iteration, one takes two replicas of such a system and, again, updates the existing links and introduces new connections among spins from different replica and so on up to the $k$-th iteration. In this way one can immediately see that a notion of distance emerges straightforwardly as two spins can be considered at a distance $d$ if they are first connected at the $d$-th iteration (see Figs. 1 and 2).

More formally, the hierarchical ferromagnet (HFM) with $K$ levels of iterations is described by the Hamil-
tonian \( H_{K}^{HFM} \), defined recursively as

\[
H_{K}^{HFM}(\{S\}|J,\sigma) = H_{K-1}^{HFM}(\{S_1\}|J,\sigma) + H_{K-1}^{HFM}(\{S_2\}|J,\sigma) - \frac{J}{2\sigma K} \sum_{i<j} S_i S_j,
\]

where \( \{S\} \) is the set of \( N = 2^K \) spins making up the system, each labeled as \( i = 1, \ldots, N \), while \( \{S_1\} \) and \( \{S_2\} \) are the sets of spins related to the two smaller copies of sizes \( 2^{K-1} \) that are merged up. Spins are binary and can take values +1 or −1. The parameters \( J \) and \( \sigma \) are bounded as \( J > 0 \) and \( \sigma \in (1/2, 1) \): the former trivially arises from the ferromagnetic nature of the model which makes neighboring spin to “imitate” each other, while the latter can be understood by noticing that for \( \sigma > 1 \) the interaction energy goes to zero in the thermodynamic limit, while for \( \sigma < 1/2 \) the interaction energy diverges in the same limit.

Next, the Hopfield model requires for its definition the set of \( N \) quenched vectors \( \{\xi^\mu\}, i = 1, \ldots, N \), of length \( \mu \) and whose entries are drawn from the distribution

\[
P(\xi^\mu) = \frac{1}{2}(\xi^\mu - 1) + \frac{1}{2} \delta(\xi^\mu + 1),
\]

being \( \mu = 1, \ldots, P \). By applying the Mattis gauge \( S_i \rightarrow -S_i \xi_i^\mu \), and summing over the \( P \) patterns, the Hamiltonian \( H_{K}^{HNN} \) for the hierarchical neural network (HNN), at the \( K \) level of iteration, reads as

\[
H_{K}^{HNN}(\{S\} | \xi, \sigma) = H_{K-1}^{HNN}(\{S_1\} | \xi, \sigma) + H_{K-1}^{HNN}(\{S_2\} | \xi, \sigma) - \frac{1}{2} \frac{1}{2^{2\sigma K}} \sum_{i<j} \sum_{\mu=1}^{P} \xi_i^\mu \xi_j^\mu S_i S_j,
\]

with \( H_{0}^{HNN} \equiv 0 \) and \( \sigma \) still within the previous bounds, i.e. \( \sigma \in (1/2, 1) \).

Finally, the hierarchical spin-glass (HSG) requires for its definition the set of \( N(N-1)/2 \) quenched variables \( \chi_{ij} \) drawn from a standard centered Gaussian distribution \( N[0,1] \) such that the related Hamiltonian \( H_{K}^{HSG} \), at the \( K \) level of iteration, reads as

\[
H_{K}^{HSG}(\{\chi\} | \chi, \sigma) = H_{K-1}^{HSG}(\{\chi_1\} | \chi, \sigma) + H_{K-1}^{HSG}(\{\chi_2\} | \chi, \sigma) - \frac{1}{2} \frac{1}{2^{2\sigma K}} \sum_{i<j} \chi_{ij} S_i S_j,
\]

All these models (i.e., HFM, HNN, HSG) can be thought of as spin systems embedded on a weighted graph \( \mathcal{G} = (V, E) \), where \( V \) is the set of nodes labeled as \( i = 1, \ldots, 2^K \) and \( E \) is the set of links whose cardinality is \( |E| = 2^{K-1}(2^K - 1) \). Each spin \( S_i \) occupies the vertex \( i \in V \) and each link \( (i,j) \in E \) is associated to a weight \( J_{ij} \) capturing the effective coupling among spins. Then, in general, the Hamiltonians in (1), (3), and (4) can all be written in the compact form

\[
H_{K}^{(\text{model})}(\{S\}|J,\sigma) = \sum_{ij} J_{ij}^{(\text{model})} S_i S_j.
\]

III. GRAPH GENERATION IN THE HIERARCHICAL FERROMAGNET

In this section we focus on the generation of the weighted graph \( \mathcal{G} \) underlying the Hamiltonian \( H_{K}^{HFM}(\{S\}|J,\sigma) \) and in the next subsections we will analyze its properties.

The iterative construction outlined in the previous section can be adopted to build up \( \mathcal{G} \) (see Fig. 2): we start from a couple of nodes \( i \) and \( j \), connected by a link carrying a weight \( J_{ij} = J_1(1) = 4^{-\sigma} \), and we refer to this graph as \( \mathcal{G}_1 \). Then, we take two replicas of \( \mathcal{G}_1 \) and we connect nodes belonging to different replicas with links carrying a weight \( J_2(2) = 4^{-2\sigma} \), while existing links are updated as \( J_2(1) = J_2(1) + J_2(2) = 4^{-\sigma} + 4^{-2\sigma} \). The graph \( \mathcal{G}_2 \) therefore counts \( 2^2 \) nodes. We proceed iteratively in such a way that at the \( K \)-th iteration new links connecting nodes belonging to different replicas are associated to a weight \( J_{K}(K) = 4^{-K\sigma} \), while existing links in each replica \( \mathcal{G}_{K-1} \) are all increased by the same value \( J_{K}(K) \), i.e.

\[
J_{K}(d) = \sum_{l=d}^{K} 4^{-l\sigma} = \frac{4^{\sigma(1-d)} - 4^{-K\sigma}}{4^\sigma - 1}.
\]

The resulting graph \( \mathcal{G}_K \) (simply referred to as \( \mathcal{G} \) to lighten the notation), is undirected and fully connected. Its nodes make up a set \( V \) of size \( N = 2^K \) and are labeled as \( i = 1, \ldots, N \). Also, the set of links \( E \) contains all possible \( \binom{N}{2} \) connections as the graph is fully connected, and each link \( (i,j) \in E \) is associated to a weight \( J_{ij} \) which can be
defined in terms of the distance between nodes $i$ and $j$, once a proper metric has been introduced.

In fact, the procedure described above provides a notion of distance $d$ which we recall here: two nodes are said to be at distance $d$ if they are first connected at the $d$-th iteration. For completeness, we also fix $G_0$ as the graph consisting in a single node.

As a result, this metric is intrinsically ultrametric as, for any pair $i, j \in V$, we have

- $d_{ij} \geq 0$;
- $d_{ij} = 0$ iff $i = j$;
- $d_{ij} = d_{ji}$, that is, the metric is symmetric;
- $d_{ij} \leq \max(d_{iz}, d_{jz})$ (this is the so-called ultrametric inequality).

Beyond the definition of distance $d_{ij}$ given in the previous section and based on recursivity, we can straightforwardly adopt the $p$-adic metric [24], and measures the distance $\rho_{ij}$ between nodes $i$ and $j$, as (here $p$ is set equal to 2)

$$\rho_{ij} = ||i - j||_2 = 2^{-\text{ord}_2(i-j)},$$

being $\text{ord}_2(i - j)$ the exponent of the largest power of 2 that divides $(i - j)$. As a result, $\rho_{ij} \in \{2^{-K+1}, 2^{-K+2}, \ldots, 2, 1\}$. Then, it is reasonable to take coupling strengths decaying with the distance and, in general, we can pose

$$J_{ij} = \frac{A}{\rho_{ij}^{2\sigma}} + B.$$  \hspace{1cm} (8)

By fixing $A = \frac{2^{-2\sigma K}}{2^{2\sigma}-1} 2^{2\sigma}$ and $B = -A^{-2\sigma}$, we recover the definition in [6]. In fact, we can rearrange Eq. (8) and get the weights as

$$J_{ij} = \frac{2^{-2\sigma K}}{4^{\sigma} - 1} \left[ \left( \frac{2}{\rho_{ij}} \right)^{2\sigma} - 1 \right] \sim \frac{1}{(N\rho_{ij})^{2\sigma}}.$$  \hspace{1cm} (9)

The two extrema for $\sigma$, i.e. $\sigma = 1/2$ and $\sigma = 1$, therefore correspond to a coupling strength scaling linearly and quadratically, respectively, with the (2-adic) distance between nodes. The metric $\rho$ is connected with $d$ by $d_{ij} = K - \text{ord}_2(i - j)$.

As anticipated, the HFM in [1], is obtained by pasting on each vertex $i$ a spin $S_i$ and letting spins interact with a coupling $J_{ij}$.

The formalization just described can be properly extended to allow for a degree of stochasticity, e.g. the set of labels can be extracted from a suitable distribution, and/or $p$ can be varied hence generating structures based on $p$-plets rather than on couples, that is, ultimately hierarchical $p$-spin models [10] or their $p \to \infty$ limit known as hierarchical random energy model [11]. Here we focus on the deterministic case depicted in Fig. 2, which holds for pairwise interactions only.

We proceed the investigation by deriving the number $n_K(J)$ of links carrying weight $J$, which provides a picture of how weights are distributed in between the two extrema

$$J_{\text{max}} = J(1) = \frac{1 - 4^{-K\sigma}}{4^{\sigma} - 1},$$

$$J_{\text{min}} = J(K) = 4^{-K\sigma}.$$  \hspace{1cm} (10)

To this aim it is convenient to count the number $n(d)$ of couples $(i,j)$ such that $d_{ij} = d$ and which are therefore connected by a link with weight $J_{ij} = J(d_{ij}) = J(d)$. In fact, we have

$$n_K(d) = 2^{d-1} \frac{N}{2} = 2^{K+d-2},$$  \hspace{1cm} (11)

$$\frac{n_K(d)}{n_K(d-1)} = \frac{2^{2\sigma}}{2^{2\sigma} - 1}.$$  \hspace{1cm} (12)
and, of course, \( \sum_{d=1}^{K} n(d)/(N^2) = 2^{K-1}(2^K - 1)/(N^2) = 1 \).

Moreover, by inverting the formula in Eq. (6) i.e. \( d = -1/(2\sigma) \log_2((2N)^{-2\sigma} + J(1-2^{-2\sigma})) \), we can express \( n(d) \) in terms of \( J \), namely

\[
n_K(J) = \frac{N^2}{2} \left[ 1 + JN^{2\sigma}(2^{2\sigma} - 1) \right]^{-\frac{1}{2\sigma}} \approx \frac{N}{2} J^{-\frac{1}{2\sigma}},
\]

where the last approximation holds for \( N \) large and highlights that the distribution is power-law (although with cut-offs given by Eqs. (10) and (11)). The distribution \( n_K(J) \) is depicted in Fig. 3, where different choices of \( \sigma \) are as well compared, while in Fig. 4 (left panel) the overall pattern of weights \( J \) is shown.

Another observable, closely related to the coupling matrix \( J \) is the weighted degree \( w \) [6]. Differently from the (bare) degree \( d \), which simply counts the number of links stemming from a node, the weighted degree also accounts for the weights associated to stemming links. More precisely, the weighted degree \( w_i \) of node \( i \) is defined as

\[
w_i = \sum_{j=1,j\neq i}^{N} J_{ij};
\]

of course, since there is perfect homogeneity within this system \( w_i \equiv w, \forall i \). Thus, from a statistical mechanics perspective, \( w_i \) (respectively \(-w_i \)) represents the field acting on the \( i \)-th spins when all the remaining spins are pointing upwards (respectively downwards). Recalling Eqs. (6) and (12) we get

\[
w_K(\sigma) = \sum_{d=1}^{K} 2^{d-1} J(d)
\]

\[
= \frac{2^{K-2}}{4^\sigma - 1} \left[ 4^\sigma \sum_{d=1}^{K} 2^{d(1-2\sigma)} - 4^{-K\sigma} \sum_{d=1}^{K} 2^{d} \right]
\]

\[
= \frac{1}{1 - 4^{-\sigma}} \left[ 1 - t^{-K} t^{-1} - 2^{-2\sigma}(K+1)(2^{K} - 1) \right]
\]

\[
= \frac{1}{1 - 4^{-\sigma}} \left[ 1 + t^{-K} (2^{K} - 1) - \frac{N-1}{2N(2^{2\sigma})} \right],
\]

where in the first line \( 2^{d-1} \) is the number of neighbors at distance \( d \) and we posed \( t = 2^{2\sigma-1} > 1 \). Notice that, when \( \sigma > 1/2 \), the second term in the square brackets in Eq. (15) gets negligible in the thermodynamic limit and similarly for \( t^{-K} \), being \( 1 - 2\sigma < 0 \), whence we have

\[
w_K(\sigma > 1/2) \approx 1/\left(4^\sigma - 1(4^\sigma - 2)\right).
\]

It is worth stressing that, in the thermodynamic limit, the weighted degree \( w_K(\sigma > 1/2) \) remains finite, although the bare degree of any node goes to infinity. On the other hand, when \( \sigma = 1/2 \) we have \( t = 1 \) and, using [15], the first term in square brackets converges to \( K-1 \), while the second term converges to \( 1/2 \), whence we have

\[
w_K(\sigma = 1/2) = 2 \left[ K - 1 - \frac{N - 1}{2N} \right] \sim K,
\]

that is, in the thermodynamic limit, the weighted degree has a logarithmic divergence with \( N \) (we recall that \( N = 2^K \)); coherently, the case \( \sigma = 1/2 \) is excluded from the statistical-mechanics investigations [2, 3].

The last part of this section is devoted to the study of the network modularity and clustering. Of course, when looking at the bare topology of the hierarchical network we have a fully-connected graph with no community structure and a trivial, unitary clustering coefficient. However, when weights on links are also taken into account one can highlight the emergence of a high degree of modularity and of clustering. The former can be quantified in terms of the Generalized Topological Overlapping Matrix (GTOM) [19], while the latter can be quantified by properly extending the standard formula in order to account for the presence of weights.

Let us start with the degree of modularity. The hierarchical graph exhibits, by construction, a modular structure as the \( N/2 \) sub-graphs \( G_1 \), the \( N/4 \) subgraphs \( G_2 \) and so on are tightly connected. In order to study this kind of structure we use the GTOM algorithm, that outlines the degree of similarity between a couple of nodes: the higher this value, the closer are nodes. In particular, we can define the similarity between nodes considering a measure based on the number of shared direct neighbors of a node \( i \): in this case, since we are dealing with weighted matrix, the algorithm considers the following

\[
\text{GTOM1}(i, j) = \frac{|N_1(i) \cap N_1(j)| + J_{ij}}{|N_1(i) \cap N_1(j)| - J_{ij}},
\]

where \( N_1(i) \) and \( N_1(j) \) are the neighbors connected respectively to the nodes \( i \) and \( j \) by a single edge, and \( J_{ij} \) is the weighted adjacency matrix (i.e., the couplings). As a result, we obtain a matrix whose elements represent the degree of overlap between clusters of nodes (dimers, squares, and so on): since \( J \) is a block matrix, we expect that this same structure is preserved in the computation of the GTOM value. As we can see in the lower panel of Fig. 5, this is the case, and the matricial representation of GTOM mirrors the ultrametric structure of the graph. It is worth stressing that we can compute the degree of dissimilarity as

\[
\text{disGTOM1}(i, j) = 1 - \text{GTOM1}(i, j),
\]
is node independent and can be simplified as

\[cw = \frac{1}{w(N-2)} \left[ \sum_{d=1}^{K} \sum_{d' = 1}^{K} J(d) + J(d') \frac{2^{d-1}2^{d'-1}}{2} + 2 \sum_{d=1}^{K} J(d) \left( \frac{2^{d-1}}{2} \right) \right] = 1.\]  (22)

The result in eq. (22) derives from the fact that the hierarchical graph is fully connected, thus, as only weights of adjacent links are counted, the summation simply returns the weighted degree times the number of triangles including a given edge.

A. The Hierarchical Ferromagnet with noise: deterministic dilution

We can allow for the presence of noise within the system by assuming that links, whose weight is smaller than the noise level \(T\), are ineffective (this mimics e.g., the fail or the unreliability of the link itself). Therefore, despite the network we are considering is fully connected, when noise is present weaker weights, with \(J_{ij} < T\), basically do not play any longer, as if they were missing [7]. Since in the statistical mechanical analysis the noise level can be tuned arbitrarily [23], it is crucial to understand how the overall network connection and clustering are accordingly modified.

The analysis described in the previous section can be generalized in these terms. For instance, the distribution \(n_K(J)\) will exhibit a lower cutoff, being \(n_K(J) = 0\) for any \(J < T\). As for the weighted degree, \(w_K(\sigma)\) (see Eq. [15]) can be generalized to \(w_K(\sigma, k)\) reading as

\[w_K(\sigma, k) = \sum_{d=1}^{k} 2^{d-1} J(d) = \frac{(2t)^{-k} \Gamma(t, k)}{(t-1)(2t-1)},\]  (23)

where \(k = k(T) = 1 - \frac{1}{\sigma} \log_2[T(4^\sigma - 1) + 4^{-K\sigma}]\), namely \(k = \min_{i \in [1, K]} \{ J(i) < T \}\), and \(\Gamma(t, k) = 2^t - 1 + t - 2^{t+1} t + 2^{t+1} j, j \in [1, K]\). Of course, by definition, \(w_K(\sigma, k) \equiv w_K(\sigma)\). These results are summarized in Fig. 6, where the behaviour of the weight of nodes is computed, as the level of noise \(T\) and the parameter \(\sigma\) are varied.

As for the clustering coefficient, we are interested in understanding whether, as the level of noise is increased, the giant component breaks into structure-less parts or it retains a large degree of clustering. The expression for the weighted clustering coefficient can be generalized into

\[cw_i = \frac{1}{w_i(z_i - 1)} \sum_{j, h \in T_i} \frac{J_{ij} + J_{jh}}{2},\]  (21)

where the normalization factor \(w_i(z_i - 1)\) ensures that \(0 \leq cw_i \leq 1\). This definition of weighted clustering coefficient considers only weights of edges adjacent to node \(i\), but not the weights of edges between neighbors of the node \(i\) (i.e. \(J_{jh}\) in the previous formula).

Of course, the formula (21) recovers the standard definition of clustering coefficient \(c_i\) for unweighted graphs, namely \(cw_i \to c_i\) as long as \(J_{ij} \to 1\). Also, for the hierarchical graph considered here, due to homogeneity, \(cw_i\) and in the upper panel of Fig. 5 a dendrogram showing the degree of dissimilarity is plotted: \(\text{disGTOM1}(i,j)\) is regularly crescent as \(d_{ij}\) grows, underlying again the symmetrical structure of the graph. Further details on modularity can be found in Appendix A.

As for the clustering coefficient, we adopt the formula given in [8], that is

\[cw_i = \frac{1}{w_i(z_i - 1)} \sum_{j, h \in T_i} \frac{J_{ij} + J_{jh}}{2},\]  (21)
cw(k) to account for the presence of some noise that impairs weak links. When \( k(T) = K - 1 \), the \( n(k) = N^2/4 \) weakest links are neglected and the graph breaks down in two equal components of size \( N/2 \) which are a rescaled version of the original graph. Hence, for any node of each component \( cw(k = K - 1) \) is still unitary. As noise is raised each component of the graph is further split and the resulting components all form clique.

Therefore, the stationary distribution, referred to as \( \pi \), satisfies \( \pi = W \pi \), that is, \( \pi \) coincides with the eigenvector \( \phi_{\lambda_0} \) of \( W \) corresponding to eigenvalue \( \lambda_0 = 1 \). Due to the stochasticity of \( W \), \( \lambda_0 = 1 \) is just the Perron-Frobenius eigenvalue of \( W \) and \( \pi = e/\sqrt{N} \), where all the \( N \) entries of the vector \( e \) are equal to 1.

The particular symmetry of \( W \) allows to see that the states

\[
\phi_{\lambda_1} = \frac{1,1,\ldots,1,-1,\ldots,-1}{\sqrt{N}},
\phi_{\lambda_2} = \frac{1,1,\ldots,1,-1,\ldots,-1}{\sqrt{N}},
\phi_{\lambda_3} = \frac{1,1,\ldots,1,-1,\ldots,-1}{\sqrt{N}},
\]

(and so on), are also eigenstates of \( W \) and the related eigenvalues are

\[
\lambda_1 = \frac{N/2}{\sum_{j=1}^{N/2} W_{ij} - \frac{N}{2} W_{iN}} = \frac{1}{w} \left( \sum_{j=1}^{N/2} J_{ij} - \frac{N}{2} J_{iN} \right)
\]

\[
\lambda_2 = \frac{N/4}{\sum_{j=1}^{N/4} W_{ij} - \frac{N}{4} W_{iN}} = \frac{1}{w} \left( \sum_{j=1}^{N/4} J_{ij} - \frac{N}{4} J_{iN} \right)
\]

\[
\lambda_3 = \frac{N/8}{\sum_{j=1}^{N/8} W_{ij} - \frac{N}{8} W_{iN}} = \frac{1}{w} \left( \sum_{j=1}^{N/8} J_{ij} - \frac{N}{8} J_{iN} \right)
\]

Here we adopt the convention \( 1 = \lambda_0 \leq \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N \).

Incidentally, we notice that \( \lambda_1 \) is exactly the difference between the external fields acting on spins when their state is fixed as \( S_i = 1, \forall i \) and as \( S_i = 1, \forall i \leq N/2 \wedge S_i = -1, \forall i > N/2 \), respectively (as clearly the field acting on a node \( i \) is \( h_i = \sum_j J_{ij} S_j \)).

Moreover, as one can see from Eq. 25 \( \lambda_1 \) converges to 1 in the thermodynamic limit and this suggests an ergodicity breaking for the stochastic process (which in turn mirrors ergodicity breaking in statistical mechanics too and hides the presence of several metastable states in the model thermodynamics [11, 2]). In fact, \( \phi_{\lambda_0} \) and \( \phi_{\lambda_1} \) generate a subspace such that any vector in this subspace (hence writable as a linear combination of \( \phi_{\lambda_0} \) and \( \phi_{\lambda_1} \)) is an eigenvector of the operator of the same eigenvalue.

FIG. 6: Weighted degree of nodes for the hierarchical graph where the presence of noise is mimicked by neglecting links displaying a weight smaller than \( J(d) \), namely links connecting nodes at a distance larger than \( d \). In this case \( K = 12 \), \( \sigma \in [0.5, 1] \) and \( T \) is taken varying in the interval [2, 12] (that means to neglect links such that \( J(d < T) \). We see that the higher values of \( w \) are obtained in correspondence of low \( \sigma \) and \( d \); when \( d = K = 12 \) all links are neglected.

Analogous arguments also hold for the degree of modularity. Therefore, even in the presence of noise, we can look at \( G \) as a clustered structure with a large degree of redundancy.

B. The Hierarchical Ferromagnet as a Markov chain

The graph modeling the Hierarchical Ferromagnet displays a countable set of nodes and finite weights, i.e. \( J_{\text{min}} \leq J_{ij} \leq J_{\text{max}} \), for any couple \((i,j)\). Given such properties, upon proper normalization of weights \( J_{ij} \rightarrow W_{ij} = J_{ij}/w_i \), the graph \( G(V,E) \) describes a Markov chain, where \( V \) is the state space (each node \( i \) represents a state) and \( W \) is the transition matrix (see e.g., [13]).

We now focus on the pattern of couplings and check the stationary states, without any (direct) concern about spin dynamics: we will see, however, that the latter share several properties with those of this Markov chain [2].

Due to the symmetry among rows and columns (the summation over the rows equals 1 as the summation over the columns) that the graph implicitly has, \( W \) is not only stochastic, but even doubly stochastic. We also introduce a distribution \( p = (p_i : i \in V) \) on \( V \) in such a way that the probability to find the random process in a state \( i \) is given by \( p_i \). The evolution of the stochastic process is then provided by the following master equation

\[
p(t + 1) = Wp(t) \rightarrow \dot{p}(t) = Wp(t) - p(t).
\]

(24)
satisfies

\[ 0 = \lambda_0 \leq \lambda_1 \leq \lambda_2 = \lambda_3 \leq \lambda_4 = \lambda_5 = \lambda_6 = \lambda_7 \leq \lambda_8 = \ldots = \lambda_{15} \leq \ldots \leq 2w, \]

and we call spectral gap \( \lambda \) of \( L \) the smallest non trivial eigenvalue, \( \lambda_1 \), of \( L \) [13]. In particular, the smaller \( \lambda_1 \) and the lower the number of links we need to cut so that the graph is divided in two independent blocks. In the HFM, we expect that this value tends to zero when the size of the system increases (see Fig. 7), obtaining the division of the network in two independent subgraphs, not interconnected. As depicted in Fig. 7, \( \lambda \) goes to zero exponentially with \( K \) according to \( f(K) = e^{-a_\sigma K} \). By fitting numerical data we find that the rate \( a_\sigma \) decreases with \( \sigma \) meaning that the higher the value of \( \sigma \) and the less structured the graph is as the system size increases.

![Graph showing the behavior of the spectral gap \( \lambda \) with \( K \) varying in [7, 12], and \( \sigma \) with values 0.8, 0.9, 1, represented by symbols (square, triangles and circles respectively). We notice that the parameter \( \sigma \), tuning the interaction strength between nodes, plays a crucial role in the computation of the spectral gap; in fact, the higher is \( \sigma \) and the lower is \( \lambda \) (for every \( K \) fixed). The continuous lines represent data fitting functions, with \( a_1 = 0.6931 \), \( a_{0.9} = 0.5545 \) and \( a_{0.8} = 0.4159 \), that underline the monotonically decrecent behavior of the data (symbols). In the inset we plot \( \lambda \) as function of both \( K \) and \( \sigma \): the brighter the color, the higher \( \lambda(\sigma, K) \), as evidenced by the colorbar.](image)

We close with a remark. Once the size of the network fixed, the degree of modularity grows with \( \lambda \). Accordingly, we expect that the mean time for the Markov process (e.g. a random walker) to get broadened over the whole system grows with \( \lambda \). Therefore, consistently with [14], we find that modularity has a role in slowing down the transport process on a network.

### IV. Graph Generation in the Hierarchical Neural Network

Let us consider the hierarchical weighted graph \( G \) and let us generalize its coupling matrix \( J \) in order to account for the Hebbian prescription. This can be done following the so-called attribute approach: each node
\( i \in V \) is endowed with a set of attributes \( \xi_i \) encoded by a vector of length \( P \) whose entries are dichotomous and defined stochastically (see Eq. 2). The coupling \( X_{ij} \), arising by comparing \( \xi_i \) and \( \xi_j \), is meant to mimic a learning process hence correlating/uncorrelating (i.e. strongly/poorly connecting).

The coupling matrix \( X \) is then used to modulate the former \( J \) in such a way that the final coupling matrix \( Q \) is given by the element-wise product

\[
Q_{ij} = X_{ij} J_{ij}, \quad (30)
\]

for any couple \((i,j)\). More precisely, recalling Eq. 3 we have

\[
X_{ij} = \sum_{\mu=1}^P \xi_i^\mu \xi_j^\mu, \quad (31)
\]

which is also known as Hebbian rule in the neural-network context [16]. In this way, even close (according to the ultrametric distance) couples may possible exhibit an overall null coupling if it occurs that the related entry in \( X \) is null. Basically, \( J \) favors couples which are close according to the ultrametric distance (defined on the set \( \{\xi_i\} \)) [4].

Notice that \( X_{ij} \) is a stochastic variable fulfilling a binomial distribution peaked at zero and with variance scaling linearly with \( P \) [26]. As both \( X_{ij} \) and \( J_{ij} \) are bounded we have

\[
Q_{\max} = J_{\max} X_{\max} = J(1) \times P = \frac{P(1 - 4^{-K\sigma})}{4^\sigma - 1}, \quad (32)
\]

\[
Q_{\min} = -Q_{\max}, \quad (33)
\]

\[
|Q|_{\min} = |J|_{\min} |X|_{\min} = 0. \quad (34)
\]

where the third line derives from the fact that \( Q \) is symmetrically distributed around 0.

Moreover, as long as \( P \) is large enough, we can write the following distribution for the coupling \( Q_{ij} \)

\[
P_{K,P}(Q_{ij} = q; \sigma) = P_{K,P}(X_{ij} = q/J_{ij}; \sigma) = \frac{1}{\sqrt{2\pi P}} \exp \left\{ -\frac{q^2(4^\sigma - 1)}{2P[4^{\sigma(1-d_{ij})} - 4^{-K\sigma}]} \right\}, \quad (35)
\]

where, exploiting the central limit theorem, we replaced the binomial distribution with a Gaussian distribution 27.

The formalization just described can be properly extended to allow for correlation among string entries (e.g., see [3]) and for dilution in string entries (e.g., [5]).

Here we focus on the simplest case (following Eqs 2 and 31) and we start the investigation by looking at how the distribution of weights \( n(Q) \) is affected by the modulation induced by \( X \). Results for several choices of the parameters \( \sigma \) and \( P \) are shown in Fig. 8 (actually, due the symmetry of the distribution we can focus just on positive weights). With respect to the case analyzed in Sec. III and corresponding to the graph generated by the ultrametric contribution \( J \) only, here the set of possible values for \( Q_{ij} \) is \( P \) times larger, as \( P \) is the number of possible for the entries of \( X \), that is

\[
J_{ij} \in \{J_1, J_2, \ldots, J_K\} \rightarrow Q_{ij} \in \{0, \pm 2J_1, \pm 4J_1, \ldots, \pm P J_1, \pm 2J_2, \ldots, \pm PJ_K\} \quad \text{if } P \text{ even}
\]

\[
J_{ij} \in \{J_1, J_2, \ldots, J_K\} \rightarrow Q_{ij} \in \{\pm J_1, \pm 3J_1, \ldots, \pm P J_1, \pm J_2, \ldots, \pm PJ_K\} \quad \text{if } P \text{ odd}.
\]

As a result, focusing on \( P \) odd to fix ideas, \( n(|Q|) \) is enveloped by the power law \( Q^{-1/(2\sigma)} \), which matches the values \( J_1, J_2, \ldots, J_K \), and such values are also accompanied by other \( (P-1)/2 \) values whose occurrence follows a binomial distribution.

Notice that a large \( P \) implies a broader distribution; similarly, a small \( \sigma \) implies a larger support. Therefore, we expect that the pattern of \( Q \) is still reminiscent of the hierarchical underlying structure, yet it is perturbed and the extent of such perturbation is more evident when \( P \) is large (Fig. 9).

![FIG. 8: Distribution of \( J_{K}(|Q|) \) for \( K = 9 \) and different choice of the parameters \( \sigma \) and \( P \) as specified. Circles (o) represent data points for the graph generated by \( Q \) (see Eq. 30), squares (□) represent the distribution one would obtain from the ultrametric contribution \( J \) only (see Eq. 35), and straight lines correspond to \( y = J^{-1/(2\sigma)}/(2N) \) (see Eq. 13).](image-url)

We now calculate the weighted degree of node \( i \) defined as

\[
w_i = \sum_{j=1}^N Q_{ij} = \sum_{j=1}^N J_{ij} X_{ij}. \quad (36)
\]

Differently from the HFM model, here the strict homogeneity among nodes is lost and, in general, \( w_i \) is site dependent. We can therefore estimate the distribution \( n(w) \) of weighted degrees: recalling Eq. 15 and that \( X_{ij} \) is normally (at least as long as \( P \) is sufficiently large) distributed with variance \( P \), we expect that \( v = w_i/w \) (with \( w = \sum_{j=1}^N J_{ij} \)) is normally distributed with variance
random variables $\xi$. Recalling that
\[
w_i = \sum_{i \neq j} J_{ij} X_{ij}, \quad \text{with} \quad J_{ij} X_{ij} = J(d_{ij}) \sum_{\mu=1}^{P} \xi_i^{\mu} \xi_j^{\mu}, \quad (37)
\]

it is easy to see that $E(w_i) = 0$ and, computing the variance of $w_i$, we obtain
\[
\text{Var}(w_i) = \sum_{i \neq j} J(d_{ij}) \text{Var}(\sum_{\mu=1}^{P} \xi_i^{\mu} \xi_j^{\mu}) = \sum_{i \neq j} \tilde{P} \cdot J(d_{ij}) = \tilde{P} \sigma^2 (1-d) - 4^{-K} \sigma^4 - 1,
\]

where $\tilde{P}$ represents the variance of $X_{ij}$. Starting from this point we can analyze how the expected value of the average weight of the graph (i.e. $E_{\xi}[^w] = E_{\xi}[^{\sum_{i=1}^{N} w_i}]$) changes as a function of the size $N$. This is shown in Fig. 11, where one can see the behavior of the average weight as a function of the size of the system and of $\sigma$, that tunes the interaction strength between links, and so contributes to the computation of the weight. In particular, as expected, once the size is fixed, the higher values of $E_{\xi}[^w]$ are in correspondence of low values of $\sigma$: in fact the lower $\sigma$, and the higher the weighted degrees of nodes.

The next step is to understand what happens if we compute the degree of modularity. Differently from the HFM, where there is a perfect homogeneity in the weight of nodes, here we can not apply simplifications. To better understand what happens in this case, we use the GTOM, with the GTOM1 algorithm (see in the previous section the Eq. 19) to determine the level of overlap between couples or clusters of nodes, and Eq. 20 to determine the dissimilarity between them. As shown in Fig. 12, the highest values of overlap are obtained for
diners that have ultrametric distance equal to one, yet the resulting structure is not regular as for HFM and two nodes at distance 1 may in principle exhibit a relatively large dissimilarity.

We conclude this section stressing that, in the Hopfield network, the presence of \(P\) random vectors \(\xi^\mu\), \(\mu = 1, ..., P\) peaked at zero implies that we are going to have some zero-entries in the adjacency matrix, but this does not coincide with an absence of a connection between the correspondent nodes. In particular, is not possible to establish that \(Q(d_{ij}) > Q(d_{hk})\) when \(d_{ij} > d_{hk}\) and this is the cause of the loss of a regular structure in the overlap measure shown in Fig. 12.

\[
Q_{ij} = \chi_{ij} J_{ij} = \frac{4^{\sigma(1-d_{ij})} - 4^{-K\sigma}}{4^\sigma - 1}, \tag{38}
\]

where \(\chi_{ij}\) are random independent centered Gaussian variables. In this case, we have again that the weight of nodes is site dependent, since we have the contribution of random variables. Then, we can write

\[
\mathbb{E}_{\chi_{ij}}[w_{ij}] = 0 \\
\text{Var}_{\chi_{ij}}[w_{ij}] = \text{Var}_{\chi_{ij}}[\sum_{i \neq j} \chi_{ij} J_{ij}] = \\
(N - 1) \frac{4^{\sigma(1-d_{ij})} - 4^{-K\sigma}}{4^\sigma - 1},
\]

where we used that a linear combination of random Gaussian independent variables is still a Gaussian variable, with variance equal to the sum of variances of the variables. Numerical results for the average value \(\mathbb{E}_{\chi_{ij}}[\mathbb{W}] = \mathbb{E}_{\chi_{ij}}[\sum_{i=1}^{N} w_{ij}]\) are shown in Fig. 13, where, as already explained for the HNN model, once the size is fixed, high values of \(\mathbb{E}_{\chi_{ij}}[\mathbb{W}]\) correspond to low values of \(\sigma\), that, as usual, has the task to rule the interaction strength between nodes. In this case we notice that the average value of the weight is higher than the previous case of HNN, due to the different stochastic factors \(\chi_{ij}\) and \(X_{ij}\), respectively.

\[
\text{FIG. 12: Representation of modularity between nodes using the topological overlapping measure (GTOM) for fixed } K = 5, \text{ and } \sigma = 0.9. \text{ Upper panel: dendrogram showing the dissimilarity between nodes in the graph: nodes at distance } d_{ij} = 1 \text{ (e.g., node } i = 3 \text{ and } j = 4\text{) typically display high overlap (hence have low dissimilarity) with respect to those at distance } d_{ij} = 2 \text{ (e.g., } i = 13 \text{ and } j = 15\text{), up to the maximum distance } d_{ij} = 5 \text{ (e.g., } i = 1 \text{ and } j = 24\text{), underlying the ultrametric structure of the network. However, differently from the case of HFM, this case is irregular and overlaps, especially between nodes at close distance, are broadly distributed. Lower panel: Matricial representation of overlaps using Generalized Topological Overlapping Matrix (GTOM). Different colours represent different values of overlap, as explained by the colorbar on the right. Due to the presence of random variables in the construction the adjacency matrix } Q, \text{ we can see partial loss of regularity in the structure of the GTOM.}
\]

\[
\text{FIG. 13: Plot of } \mathbb{E}_{\chi_{ij}}[\mathbb{W}] = \mathbb{E}_{\chi_{ij}}[\sum_{i=1}^{N} w_{ij}] \text{ as a function of } \sigma \text{ and the size of the system } N = 2^K (K \in [2, 12]), \text{ with } \sigma \in [0.5, 1]. \text{ To realize it, 200 realization of } \chi_{ij} \text{ were produced, and, for everyone, we obtained } w_{ij}, \forall i \in [1, N]. \text{ Then, the algebraic mean value over the realizations was computed, with fixed } K. \text{ As a result, the higher values of } w \text{ are obtained in correspondence of high value of } K \text{ and low values of } \sigma: \text{ in blue we see the higher values of } \mathbb{E}_{\chi_{ij}}[\mathbb{W}].}
\]

We also checked the modularity of this networks by exploiting again the GTOM [19], with the GTOM1 algorithm, introduced for HFM, and already recalled for the HNN using Eqs. [19] [20]. In this case we obtain a more regular structure with respect to the Hierarchical Neural Network, due to the presence of random quenched Gaussian variables. This perfectly matches with the
VI. CONCLUSIONS AND OUTLOOKS

In the last decade hierarchical networks have been found to play a crucial and widespread role in natural phenomena \cite{1,2,9,12,13}, particularly in biological systems \cite{22,23}. Furthermore, these structures turn out to be also quasi-tractable in statistical mechanics \cite{14}, even when glassiness is present \cite{1,2,9,12,14}, hence triggering further studies of their properties.

In this work we discuss the topological features of three hierarchical models, each describing a different rule for generating coupling among nodes: Hierarchical Ferromagnet (HFM), Hierarchical Neural Network (HNN) and Hierarchical Spin Glass (HSG). Among the most important results we achieved in the present paper, we highlight that the subtle metastabilities exhibited by HFM (see e.g. \cite{1,2}) can also be evidenced in terms of ergodicity breakdown for Markov processes defined on the hierarchical weighted graph over which the spins insist: through this approach statistical mechanics and graph theory work synergically to better describe the features of these networks.

Further, these structures also exhibit high clustering and modularity which are two important properties well-evidenced in many real systems \cite{8,17}. Finally, such properties are robust with respect to the introduction of some degree of (quenched) noise coded in terms of Gaussian variables (as it is the case for the HSG) or in terms of Hebbian attributes (as it is the case for the HNN), the latter typically inducing a larger degree of broadness.

Starting with HFM, we showed the trend of the modularity and weight degree of nodes when varying the interaction strength between the elements. We then noticed that the graph could be considered as a Markov chain, where the state space is the set of nodes, and entries in the transition matrix are constituted by the distances between nodes, upon a proper normalization: the breakdown of ergodicity is thus depicted by the divergence of the mixing time, mirroring the results obtained via the statistical mechanical route. These same analyses were carried out for HNN and HSG. As expected (because now quenched disorder is introduced in the system), differences were shown to exist between the two of them and the HFM: the most important is the lost of the symmetric hierarchical structure of weights of the links, due to the presence of random variables that contribute to create the coupling matrix: for the HNN, the Hebbian rule leads to a binomial distribution for the couplings, peaked at zero and with variance scaling with P. It is worth noticing, however, that -whatever the weakness of any link- we always deal with a fully connected graph: low values of the ultrametric distance (favored by the matrix J) do not necessarily correspond to low values of the Hamming distance (favored by the matrix X).

Arguments can be used for the HSG, where the coupling matrix is defined by a Normal distribution. These investigations overcome mean field approximation and contribute to provide new information that relates
the capability of these networks to perform parallel retrieval with the underlying topology they obey to.

VII. APPENDIX A

This section is devoted to describe in details the network’s modularity.

In order to figure out which is the most effective partition, we can apply the formula introduced in \cite{1} (suitable for weighted graphs as well) which measures the degree of modularity $M$ for a given modular subdivision, chosen a priori, where each node $i$ is associated to a module $m_i$ out of $P$, i.e. $c_1 = 1, ..., P$. More precisely,

$$ M = \frac{1}{m} \sum_{i,j} \left[ J_{ij} - \frac{w_{ij}}{m} \right] \delta(c_i, c_j), \quad (39) $$

where $m = \sum_i w_i$. In particular, exploiting the homogeneity of the hierarchical graph we can write Eq. (39) in a simpler form as

$$ M = \frac{1}{Nw} \sum_{i<j} \left[ J_{ij} - \frac{w}{N} \right] \delta(c_i, c_j). \quad (40) $$

According to the different modular subdivision, we can calculate the resulting $M$, and, in general, with communities made of $2^l$ nodes we have

$$ M(l) = \frac{1}{Nw} \sum_{d=1}^{l} \left[ J(d) - \frac{w}{N} \right] 2^{l+d-1} \quad (41) $$

where we considered again $t = 2^{2\sigma - 1}$ and where

$$ w_l(\sigma) = \sum_{d=1}^{l} 2^{d-1}J(d); $$

$$ \Gamma(t, j) = 2^j - 1 + t - 2^{j+1}t + 2^{j+1}, j \in [1, K]. $$

As shown in Fig. 15, the function $M(l)$ exhibits a peak at a value $l$ approaching $k/2$ as $\sigma \to 1$. This means that the most effective modular partition (according to Eq. 39) is the one where the graph is divided in a relatively small number of clusters, but for large $\sigma$ (namely where the hierarchy is less important, see Eq. 8) this number gets smaller.

Acknowledgments

GNFM-INdAM via Progetto Giovani Agliari2014 and INFN are acknowledged for partial financial support.

[1] E. Agliari, A. Barra, A. Galluzzi, F. Guerra, D. Tantari, F. Tavani From Dyson to Hopfield: Processing on hierarchical networks, Phys. Rev. Letts. (in press), available at http://arxiv.org/abs/1407.5019

[2] E. Agliari, A. Barra, A. Galluzzi, F. Guerra, D. Tantari, F. Tavani Metastable states in the hierarchical Dyson model drive parallel processing in the hierarchical Hopfield network, J. Phys. A: Math. Theor. 48 (2015).

[3] E. Agliari, A. Barra, A. Galluzzi, F. Guerra, D. Tantari, F. Tavani Hierarchical neural networks perform both serial and parallel processing, submitted, available at http://arxiv.org/abs/1409.0227

[4] E. Agliari, A. Barra, A Hebbian approach to complex-network generation, Europhys. Lett. 94:1: 10002 (2011).

[5] E. Agliari, A. Barra, A. Galluzzi, F. Guerra, F. Moauro, Multitasking associative networks, Phys. Rev. Lett. 109(26), 268101 (2012).

[6] E. Agliari, A. Barra, G. Del Ferraro, F. Guerra, D. Tantari, Anergy in self-directed B lymphocytes: A statistical mechanics perspective, to appear in J. Theor. Biol. doi.org/10.1016/j.jtbi.2014.05.006, (2014).

[7] E. Agliari, C. Cioli, E. Guadagnini, Percolation on correlated random networks, Phys. Rev. E 84, 031120 (2011).

[8] A. Barrat, M. Barthélémy, R. Pastor-Satorras, A. Vespignani, The architecture of complex weighted networks Proc. Natl. Acad. Sci. USA 101, 3747-3752, (2004).

[9] M. Castellana, A. Barra, F. Guerra, Free-energy bounds for hierarchical spin models, J. Stat. Phys. 155, 211, (2014).

[10] M. Castellana, Rigorous Results for Hierarchical Models of Structural Glasses, J. Stat. Phys. 157, 219, (2014).

[11] M. Castellana, A. Decelle, S. Franz, M. Mezard, G.
[12] M. Castellana, G. Parisi, A renormalization group computation of the critical exponents of hierarchical spin glasses, Phys. Rev. E 83, 041134, (2011).

[13] L. Donetti, F. Neri, M.A. Muñoz, Optimal network topologies: expanders, cages, Ramanujan graphs, entangled networks and all that, JSTAT 2006.08:P08007, (2006).

[14] F.J. Dyson, Existence of a Phase-Transition in a One-Dimensional Ising Ferromagnet, Comm. Math. Phys. 12, 91-107, (1969).

[15] L. K. Gallos, C. Song, S. Havlin, H.A. Makse, Scaling theory of transport in complex biological networks, Proc. Natl. Acad. Sci. USA 104(19):7746-7751 (2007).

[16] D.O. Hebb, The organization of behavior: A neuropsychological theory, Psychology Press (1940).

[17] M.E.J. Newman, Networks: An introduction, Oxford University Press, Oxford (2010).

[18] J.R. Norris, Markov Chains, Cambridge Series in Statistical and Probabilistic Mathematics, New York (1994).

[19] A. Yip, S. Horvath, Gene network interconnectedness and the generalized topological overlap measure, BMC Bioinformatics, 8, 22 (2007).

[20] E. Ravasz, A.L. Somera, D.A. Mongru, Z.N. Oltvai, A.L. Barabasi, Hierarchical organization of modularity in metabolic networks, Science 297, 1551, (2002).

[21] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.U. Hwang, Complex networks: Structure and dynamics, Phys. Rep. 424, 175-308, (2006).

[22] P. Moretti, A.M. Munoz, Griffiths phases and the stretching of criticality in brain networks, Nature Comm. 4, 2521, (2013).

[23] S. Maslov, K. Sneppen, Specificity and Stability in Topology of Protein Networks, Science 293, 910-913, (2002).

[24] We define the \( p \)-adic metric in terms of the \( p \)-adic norm exactly the way that we defined Euclidean distance in terms of the absolute value norm. In the \( p \)-adic integers, the norm of a number is based around the largest power of the base that’s a factor of that number: for an integer \( x \), if \( p^n \) is the largest power of \( p \) that is a factor of \( x \), then the \( p \)-adic norm of \( x \) (written \( ||x||_p \)) is \( p^{-n} \). So the more times you multiply a number by the \( p \)-adic base, the smaller the \( p \)-adic norm of that number is. The way we apply that to the rationales is to extend the definition of \( p \)-factoring: if is our \( p \)-adic base, then we can define the \( p \)-adic norm of a rational number as: i. \( ||0||_p = 0 \), ii. For other rational numbers \( x \): \( ||x||_p = p^{-\alpha(x)} \); where, if \( x \) is a natural number, then \( \alpha(x) \) is the exponent of the largest power of \( p \) that divides \( x \). Therefore, two \( p \)-adic numbers \( x \) and \( y \) are close together if \( x - y \) is divisible by a large power of \( p \). Closest nodes are at distance \( 2^{-K+1} \), farthest nodes are at a distance \( 1 \).

[25] As such, \( p \) must satisfy positivity, i.e. \( \forall i \in V, 0 \leq p_i \), and normalization, i.e. \( \sum_{i \in V} p_i \).\( = 1 \).

[26] \( X_{ij} \) can be looked at as the position reached by a one-dimensional simple random walk after \( P \) steps.

[27] Strictly speaking the convergence to a Gaussian distribution is better performing in the high storage regime only, namely where \( P \sim N \); however, as we want to approach -through this perspective- the topology of the hierarchical spin glass too, we allow ourselves in taking such an approximation.