Topological properties of hierarchical networks

Elena Agliari,[1] Adriano Barra,[2] Andrea Galluzzi,[3] Francesco Guerra,[2, 4] Daniele Tantari,[5] and Flavia Tavani[6]

[1]*Dipartimento di Matematica, Sapienza Università di Roma, P.le A. Moro 2, 00185, Roma, Italy.*
[2]*Dipartimento di Fisica, Sapienza Università di Roma, P.le A. Moro 2, 00185, Roma, Italy.*
[3]*Dipartimento di Matematica, Sapienza Università di Roma, P.le Aldo Moro 5, 00185, Roma, Italy.*
[4]*Istituto Nazionale di Fisica Nucleare, Sezione di Roma*
[5]*Centro di Ricerca Matematica Ennio De Giorgi,*
Scuola Normale Superiore di Pisa, Piazza dei Cavalieri, 3, Pisa.
[6]*Dipartimento di Scienze di Base e Applicate per l'Ingegneria,*
Sapienza Università di Roma, Via A. Scarpa 16, 00185, Roma, Italy.
(Dated: June 16, 2021)

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 the presence of thermal noise 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 the assumption 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 it 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 $\sim 4^{-\sigma d}$, where σ 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
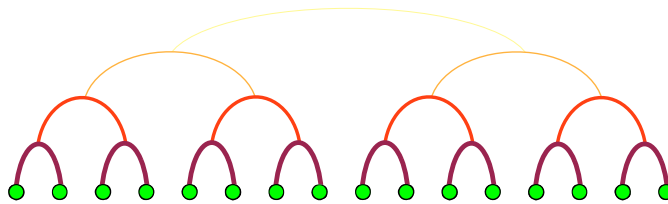
FIG. 1: (Color online) Schematic representation of the hierarchical topology that underlies the system under study: spots represent nodes where spins/neurons live, while different colors and thickness for the links mimic different intensities in their mutual interactions: the brighter and thinner the link, the smaller the related coupling.

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 Hamiltonian H_K^{HFM}, defined recursively as

$$H_K^{\mathrm{HFM}}(\{S\}|\sigma) = H_{K-1}^{\mathrm{HFM}}(\{S_1\}|\sigma) + H_{K-1}^{\mathrm{HFM}}(\{S_2\}|\sigma)$$
$$- \frac{1}{2^{2\sigma K}} \sum_{i<j}^{2^K} S_i S_j, \qquad (1)$$

where $\{S\}$ is the set of $N = 2^K$ spins making up the system, each labeled as $i = 1, ..., N$, 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 parameter σ is bounded as $\sigma \in (1/2, \ 1]$: 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. Also, notice that the coupling among spins is positive due to the ferromagnetic nature of the model which makes neighboring spins to "imitate" each other.

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

$$P(\xi_i^\mu) = \frac{1}{2}\delta(\xi_i^\mu - 1) + \frac{1}{2}\delta(\xi_i^\mu + 1), \qquad (2)$$

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

$$H_K^{\mathrm{HNN}}(\{S\}|\xi,\sigma) = H_{K-1}^{\mathrm{HNN}}(\{S_1\}|\xi,\sigma) + H_{K-1}^{\mathrm{HNN}}(\{S_2\}|\xi,\sigma)$$
$$- \frac{1}{2}\frac{1}{2^{2\sigma K}} \sum_{\mu=1}^{P} \sum_{i,j=1}^{2^K} \xi_i^\mu \xi_j^\mu S_i S_j, \qquad (3)$$

with $H_0^{\mathrm{HNN}} \equiv 0$ and σ still within the previous bounds, i.e. $\sigma \in (\frac{1}{2}, 1]$.

Finally, the hierarchical spin-glass (HSG) requires for its definition the set of $N(N-1)/2$ quenched variables χ_{ij}

drawn from a standard centered Gaussian distribution $\mathcal{N}[0,1]$ such that the related Hamiltonian H_K^{HSG}, at the K-th level of iteration, reads as

$$H_K^{\mathrm{HSG}}(\{S\}|\chi,\sigma) = H_{K-1}^{\mathrm{HSG}}(\{S_1\}|\chi,\sigma) + H_{K-1}^{\mathrm{HSG}}(\{S_2\}|\chi,\sigma)$$
$$- \frac{1}{2^{2\sigma K}} \sum_{i<j}^{2^K} \chi_{ij} S_i S_j, \qquad (4)$$

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 $i = 1, \cdots, 2^K$ 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

$$H_K^{(\mathrm{model})}(\{S\}|\sigma) = \sum_{ij} J_{ij}^{(\mathrm{model})} S_i S_j. \qquad (5)$$

III. GRAPH GENERATION IN THE HIERARCHICAL FERROMAGNET

In this section we focus on the generation of the weighted graph $\mathcal{G}$ underlying the Hamiltonian $H_K^{\mathrm{HFM}}(\{S\}|J,\sigma)$ 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
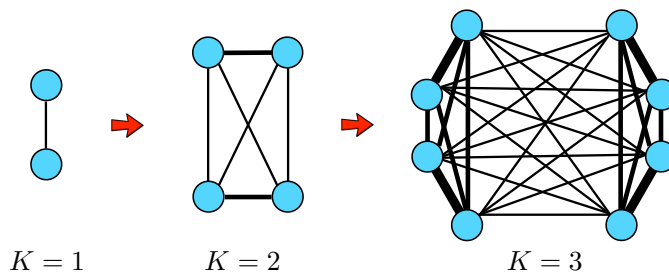
FIG. 2: (Color online) Iterative construction of the hierarchical structure up to generation $K = 3$, corresponding to $N = 2^3$ vertices. Links display different thickness according to their weight.

$J_K(K)$, in such a way that

$$J_K(d) = \sum_{l=d}^{K} J_K(l) = \sum_{l=d}^{K} 4^{-l\sigma} = \frac{4^{\sigma(1-d)} - 4^{-K\sigma}}{4^\sigma - 1}. \quad (6)$$

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 $i = 1, ..., N$. 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 $\mathcal{G}_0$ 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;
- $d_{ij} \leq \max(d_{iz}, d_{zj})$ (this is the so-called ultrametric inequality).

Beyond the definition of distance d_{ij} based on recursivity, we can straightforwardly adopt the p-adic metric [27], and measure the p-adic distance ρ_{ij} between nodes i and j, as (here p is set equal to 2)

$$\rho_{ij} = ||i - j||_2 = 2^{-\text{ord}_2(i-j)}, \quad (7)$$

being $\text{ord}_2(i - j)$ the exponent of the largest power of 2 that divides $(i - j)$ [28]. Notice that ρ is connected with d by $d_{ij} = K - \text{ord}_2(i - j)$. As a result, $\rho_{ij} \in \{2^{-K+1}, 2^{-K+2}, ..., 2, 1\}$. Then, using the 2-adic metric, one can see that the coupling strength turns out to decay algebraically with the (2-adic) distance, as typical for long-range interactions, that is

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

In fact, by posing $A = \frac{4^{-\sigma(K-1)}}{4^\sigma - 1}$ and $B = -4^{-\sigma} A$, we recover the definition in (6). Moreover, we can rearrange
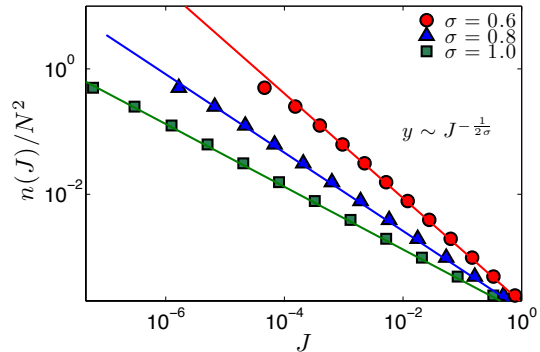

FIG. 3: (Color online) Distribution $n_k(J)$ for $K = 12$ and $\sigma = 0.6, \sigma = 0.8$, and $\sigma = 1.0$, as reported by the legend. Straight lines correspond to $y = J^{-1/(2\sigma)}/(2N)$, see Eq. 13.

Eq. 6 and, in the limit of large size, we get

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

The two extrema for σ, 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.

As anticipated, the hierarchical ferromagnet 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

$$J_{\max} = J(1) = \frac{1 - 4^{-K\sigma}}{4^\sigma - 1}, \qquad (10)$$

$$J_{\min} = J(K) = 4^{-K\sigma}. \qquad (11)$$

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

$$n_K(d) = 2^{d-1}\frac{N}{2} = 2^{K+d-2}, \qquad (12)$$

and, of course, $\sum_{d=1}^{K} n(d)/\binom{N}{2} = 2^{K-1}(2^K - 1)/\binom{N}{2} = 1$.

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

$$n_K(J) = \frac{N^2}{2}\left[1 + JN^{2\sigma}(2^{2\sigma} - 1)\right]^{-\frac{1}{2\sigma}} \approx \frac{N}{2}J^{-\frac{1}{2\sigma}}, \qquad (13)$$

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). Otherwise stated, this model can be looked at as a "scale-free Curie-Weiss". The distribution $n_K(J)$ is depicted in Fig. 3, where different choices of σ are as well compared, while in Fig. 4 the overall pattern of weights $\mathbf{J}$ is shown.

Another observable, closely related to the coupling matrix $\mathbf{J}$ is the weighted degree w [6]. Differently from the (bare) degree z, 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

$$w_i = \sum_{j=1, j\neq i}^{N} J_{ij}; \qquad (14)$$

of course, since there is perfect homogeneity within this system $w_i \equiv w, \forall i$. 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

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

$$= \frac{1}{2(4^\sigma - 1)}\left[4^\sigma \sum_{d=1}^{K} 2^{d(1-2\sigma)} - 4^{-K\sigma}\sum_{d=1}^{K} 2^d\right] \qquad (15)$$

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

$$= \frac{4^\sigma}{2(4^\sigma - 1)}\left[\frac{(2N)^{2\sigma} - N(3\times 2^{2\sigma-1} - 1) + 2^{2\sigma} - 1}{(2N)^{2\sigma}(2^{2\sigma-1} - 1)}\right], \qquad (16)$$
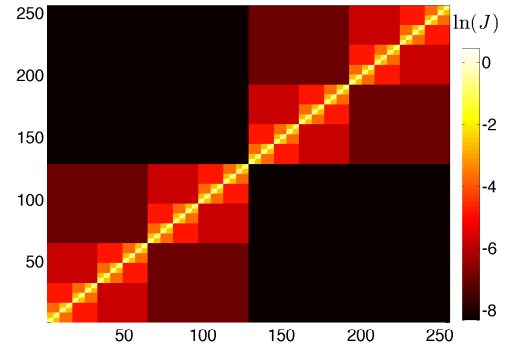

FIG. 4: (Color online) The pattern of weights J_{ij} for $K = 8$ is represented in a logarithmic scale. Notice that these patterns mirror the ultrametric structure of the graph.

where in the first line 2^{d-1} is the number of neighbors at distance d. When $\sigma > 1/2$, in the thermodynamic limit, we get

$$w_K(\sigma > 1/2) \xrightarrow{K \gg 1} \frac{4^\sigma}{(4^\sigma - 1)(4^\sigma - 2)}. \qquad (17)$$

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 $\sigma = 1/2$, using (15), the first term in square brackets converges to $K - 1$, while the second term converges to $1/2$, whence we have

$$w_K(\sigma = 1/2) = \left[K - 1 - \frac{N-2}{2N}\right] \sim K, \qquad (18)$$

that is, in the thermodynamic limit, the weighted degree has a logarithmical 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 by properly extending the formula meant for unweighted networks. In particular, modularity can be quantified in terms of the generalized topological overlap matrix $\mathbf{O}$ [19], whose entry O_{ij} measures the degree of similarity displayed by the couple of nodes (i,j) in terms of the number of shared neighbours, namely

$$O_{ij} = \frac{|N(i) \cap N(j)| + A_{ij}}{\min\{|N(i)|, |N(j)|\} - A_{ij} + 1}, \qquad (19)$$

where $N(i)$ and $N(j)$ are the sets of nearest neighbors of i and j, respectively, $|N(i) \cap N(j)|$ represents the number of common neighbours that nodes i and j share, and $\mathbf{A}$ is the adjacency matrix. Now, the presence of weights
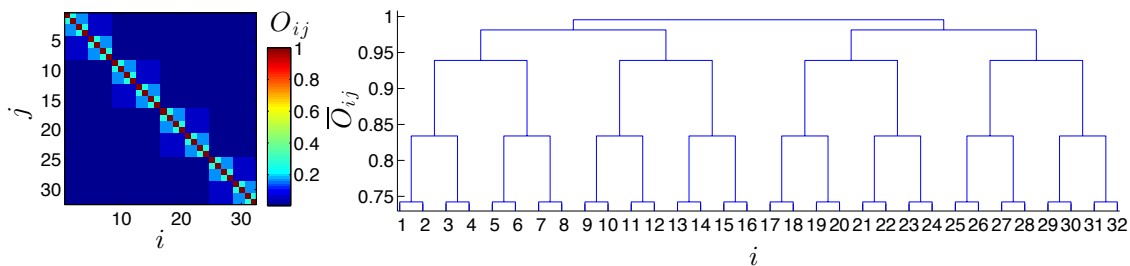
FIG. 5: (Color online) Modularity of the graph $\mathcal{G}$ with $K = 5$, $\sigma = 0.9$, represented by means of the overlap matrix $\mathbf{O}$ (left panel). Its regular structure mirrors the deterministic modularity of the network. The dissimilarity of the graph is depicted through a dendrogram (right panel), where nodes at distance $d_{ij} = 1$ (see for instance node $i = 1$ and $j = 2$, at the lower level of the dendrogram) have higher values of overlap (hence have lower values of dissimilarity) with respect to those at distance $d_{ij} = 2$ (second level from the bottom of the dendrogram), up to the maximum distance $d_{ij} = 5$ (first level on the top), underlying again the ultrametric structure of the network.

can be accounted for by modifying Eq. 19 as proposed in [20]

$$O_{ij} = \frac{1}{J_{\max}} \frac{\sum_{k=1}^{N} J_{ik} J_{kj} + J_{ij} J_{\max}}{\min\{w_i, w_j\} - J_{ij} + J_{\max}}. \quad (20)$$

Of course, Eq. 20 recovers Eq. 19 as long as we replace the adjacency matrix with the normalized coupling matrix ($0 \leq J_{ij}/J_{\max} \leq 1$). The generalized topological overlap matrix for the graph under study is shown in the left panel of Fig. 5, where one can see that $\mathbf{O}$ mirrors the ultrametric structure of the graph.

Moreover, we can compute the degree of dissimilarity as

$$\overline{O}_{ij} = 1 - O_{ij}, \quad (21)$$

which is shown through a dendrogram plot in the right panel of Fig. 5. Again, the ultrametric structure of the graph emerges markedly. Further details on modularity can be found in Appendix A.

As for the clustering coefficient, several definitions of weighted clustering coefficient cw have appeared in the literature, as summarised and compared in [25]. Since there is not any ultimate formulation, we consider two definitions, introduced in [8] and in [26], respectively, which can be seen as limiting cases. According to the formula given in [8], we get

$$cw_i^{(1)} = \frac{1}{w_i(z_i - 1)} \sum_{j,h \in \mathcal{T}_i} \frac{J_{ij} + J_{ih}}{2}, \quad (22)$$

where $\mathcal{T}_i$ is the number of triangles including node i, w_i is the weighted degree of node i, z_i is the number of nearest-neighbors of i (i.e., its bare degree), and the normalization factor $w_i(z_i - 1)$ ensures that $0 \leq cw_i^{(1)} \leq 1$. This definition of weighted clustering coefficient considers only weights of edges adjacent to node i, but not the weights of edges between the neighbors of the node i (i.e. J_{jh} in the previous formula).

Of course, the formula (22) recovers the standard definition of clustering coefficient c_i for unweighted graphs,

namely $cw_i^{(1)} \to c_i$ as long as $J_{ij} \to 1$. Also, for the hierarchical graph considered here, due to homogeneity, $cw_i^{(1)}$ is node independent and can be simplified as

$$cw^{(1)} = \frac{1}{w(N-2)} \left[\sum_{d=1}^{K} \sum_{\substack{d'=1 \\ d' \neq d}}^{K} \frac{J(d) + J(d')}{2} 2^{d-1} 2^{d'-1} \right.$$
$$\left. + 2 \sum_{d=1}^{K} J(d) \binom{2^{d-1}}{2} \right] = 1. \quad (23)$$

The result in eq. (23) 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.

According to the definition given in [26], we have

$$cw_i^{(2)} = \frac{1}{(N-1)(N-2)} \frac{1}{J_{\max}} \sum_{j,h} (J_{ij} J_{ih} J_{jh})^{1/3}, \quad (24)$$

which is again normalized, i.e., $0 \leq cw_i^{(2)} \leq 1$, but, different from Eq. 22, takes into account the weights of all edges making up a triangle and is invariant to weight permutation for one triangle. As already noticed, due to the homogeneity of the graph under study, the clustering is node independent and hereafter we shall simply refer to $cw^{(2)}$, dropping the index i.

With some algebra, we can rewrite the previous formula in terms of distances between nodes as

$$cw^{(2)} = \frac{2}{(N-1)(N-2)} \frac{1}{J_K(1)} \sum_{d=1}^{K-1} 3 \times 2^{d-3} [J_K(d)]^{\frac{1}{3}} \times$$
$$\times \sum_{d'=d+1}^{K} 2^{d'} J_K(d')^{\frac{2}{3}},$$

where $3 \times 2^{d+d'-3}$ is the number of triangles having two nodes at distance d' from a fixed node, and being themselves at distance d each other. Substituting $J_K(d)$ and

$J_K(d')$ with their exact values given by (6) and then assuming $K \gg 1$

$$(J(d)J(d')^2)^{\frac{1}{3}} = \left[\frac{4^{\sigma(1-d)} - 4^{-K\sigma}}{4^\sigma - 1}\right]^{\frac{1}{3}} \left[\frac{4^{\sigma(1-d')} - 4^{-K\sigma}}{4^\sigma - 1}\right]^{\frac{2}{3}} \approx$$

$$\approx \frac{1}{4^\sigma - 1}\left[4^{\sigma(1-d)}4^{2\sigma(1-d')}\right]^{\frac{1}{3}},$$

we arrive to the following approximation of the clustering coefficient $cw^{(2)}$

$$cw^{(2)} \approx \widetilde{cw}^{(2)} = \frac{3}{4(N-1)(N-2)} \frac{4^\sigma}{(4^\sigma - 1)J_K(1)} \times$$

$$\times \sum_{d=1}^{K-1} \sum_{d'=d+1}^{K} 2^{d(1-\frac{2}{3}\sigma)} 2^{d'(1-\frac{4}{3}\sigma)} = \qquad (25)$$

$$= \frac{3(2N)^{2\sigma}[2^{\frac{2\sigma}{3}+1} - 4 + 2^{\frac{2\sigma}{3}}N^{2(1-\sigma)}(2^{\frac{4\sigma}{3}} - 2)]}{(2^{\frac{2\sigma}{3}} - 2)(2^{\frac{4\sigma}{3}} - 2)(2 - 3N + N^2)(2^{2\sigma} - 4)(N^{2\sigma} - 1)} +$$

$$- \frac{3N^{\frac{2}{3}\sigma+1}2^{2\sigma}(2^{2\sigma} - 4)}{(2^{\frac{2\sigma}{3}} - 2)(2^{\frac{4\sigma}{3}} - 2)(2 - 3N + N^2)(2^{2\sigma} - 4)(N^{2\sigma} - 1)}.$$

This approximation provides the leading behavior for $cw^{(2)}$ in the limit of large size. It is worth noticing that, differently from the previous definition (22), here $cw^{(2)}$ is always close to zero, due to presence in the graph of a high number of triangles constituted by distant nodes.

The dependence on σ of $cw^{(2)}$ and the goodness of the approximation provided by $\widetilde{cw}^{(2)}$ are visualised in Fig. 6. In particular, $cw^{(2)}$ is relatively low and decreasing with σ. In fact, the definition (23) takes into account the weights of all the links making up a triangle and the number of links between distant nodes (i.e., nodes loosely connected) is much larger than the number of links between close nodes (i.e., nodes tightly connected). Moreover, any weight is decreasing in σ and, as a result, the overall clustering coefficient $cw^{(2)}$ is also decreasing in σ.

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 [2, 3], 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
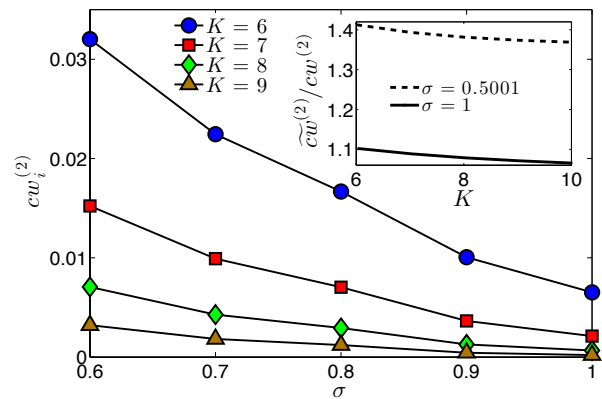
FIG. 6: (Color online) Main plot: clustering $cw^{(2)}$ as a function of σ and for different choices of K, as explained in the legend. The value of $cw^{(2)}$ is monotonically decreasing with K and with σ. Inset: ratio between the approximated value $\widetilde{cw}^{(2)}$ (calculated via Eq. 25) and the exact values $cw^{(2)}$ (calculated numerically). More precisely, $\widetilde{cw}^{(2)}$ provides an upper bound for $cw^{(2)}$ and the approximation is better for large σ.

Eq. 15) can be extended to $w_K(\sigma, k)$ reading as

$$w_K(\sigma, k) = \sum_{d=1}^{k} 2^{d-1}J(d) = \qquad (26)$$

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

$$= \frac{4^{-\sigma k}\Gamma(2^{2\sigma-1}, k)}{(2^{2\sigma-1} - 1)(2^{2\sigma} - 1)},$$

where $k = k(T) = 1 - \frac{1}{2\sigma}\log_2[T(4^\sigma - 1) + 4^{-K\sigma}]$, namely $k = \min_{i\in[1,K]}\{J(i) < T\}$, and $\Gamma(t, j) = 2^j - 1 + t - 2^{j+1}t + 2^jt^{j+1}, j \in [1, K]$. Of course, by definition, $w_K(\sigma, K) \equiv w_K(\sigma)$. These results are summarized in Fig. 7, where the behaviour of the weight of nodes is computed, as the level of noise T and the parameter σ 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 coefficients can be generalized into $cw^{(1,2)}(k)$ 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 $cw^{(1)}(k = K - 1)$ is still unitary. As noise is raised each component of the graph is further split and the resulting components all form weighted cliques. Analogous arguments also hold for the degree of modularity.

On the other hand $cw^{(2)}(k)$, is quantitatively affected by the level of noise which further reduce its value.
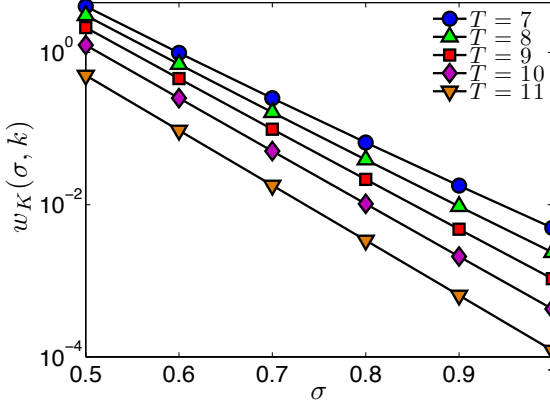
FIG. 7: (Color online) Weighted degree $w_K(\sigma, k)$ of nodes for the hierarchical graph where the presence of noise is mimicked by neglecting links displaying a weight smaller than $J(d)$, namely links connecting nodes at a distance larger than d. In this case $K = 12$, $\sigma \in (0.5, 1]$ and T is taken varying in the interval $[7, 11]$ (that means to neglet links such that $J(d) < T$). Notice that the higher values of w are obtained in correspondence of low σ and d; when $d = K = 12$ all links are neglected.

Therefore, even in the presence of noise, we can look at $\mathcal{G}$ 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. $J_{\min} \leq J_{ij} \leq J_{\max}$, for any couple (i, j). Given such properties, upon proper normalization of weights $J_{ij} \rightarrow W_{ij} = J_{ij}/w_i$, the graph $\mathcal{G}(V, E)$ describes a Markov chain, where V is the state space (each node i represents a state) and $\mathbf{W}$ is the transition matrix (see e.g., [18]).

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, $\mathbf{W}$ is not only stochastic, but even doubly stochastic. We also introduce a distribution [29] $p = (p_i : i \in V)$ 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

$$p(t + 1) = W p(t) \rightarrow \dot{p}(t) = W p(t) - p(t). \quad (27)$$

Therefore, the stationary distribution, referred to as π, satisfies $\pi = \mathbf{W}\pi$, that is, π coincides with the eigenvector ϕ_{λ_0} of $\mathbf{W}$ corresponding to eigenvalue $\lambda_0 = 1$. Due

to the stochasticity of $\mathbf{W}$, $\lambda_0 = 1$ is just the Perron-Frobenius eigenvalue of $\mathbf{W}$ and $\pi = \mathbf{e}/\sqrt{N}$, where all the N entries of the vector $\mathbf{e}$ are equal to 1.

The particular symmetry of $\mathbf{W}$ allows to see that the states

$$\phi_{\lambda_1} = (\underbrace{1, 1, \ldots, 1}_{N/2}, \underbrace{-1, -1, \ldots, -1}_{N/2})/\sqrt{N},$$

$$\phi_{\lambda_2} = (\underbrace{1, \ldots, 1}_{N/4}, \underbrace{-1, \ldots, -1}_{N/4}, \underbrace{1, \ldots, 1}_{N/4}, \underbrace{-1, \ldots, -1}_{N/4})/\sqrt{N},$$

$$\phi_{\lambda_3} = (\underbrace{1, \ldots, 1}_{N/8}, \ldots, \underbrace{-1, \ldots, -1}_{N/8})/\sqrt{N},$$

(and so on), are also eigenstates of $\mathbf{W}$ and the related eigenvalues are

$$\lambda_1 = \sum_{j=1}^{N/2} W_{1j} - \frac{N}{2} W_{1N} = \frac{1}{w}\left(\sum_{j=1}^{N/2} J_{1j} - \frac{N}{2} J_{1N}\right)$$
$$= 1 - \frac{N(2^{2\sigma} - 2)(2^{2\sigma} - 1)}{(2N)^{2\sigma} - 2N(2^\sigma - 1) + 2^{2\sigma} - 2} \approx 1 - \frac{2^{4\sigma}}{N^{2\sigma - 1}} \quad (28)$$

$$\lambda_2 = \sum_{j=1}^{N/4} W_{1j} - \sum_{j=N/4+1}^{N/2} W_{1j} = \frac{1}{w}\left(\sum_{j=1}^{N/4} J_{1j} - \sum_{j=N/4+1}^{N/2} J_{1j}\right)$$
$$= 1 - \frac{N(2^{2\sigma} - 1)(2^{2\sigma-1} - 1)(2^{2\sigma} + 2)}{(2N)^{2\sigma} - 2N(2^\sigma - 1) + 2^{2\sigma} - 2} \approx 1 - \frac{2^{6\sigma - 1}}{N^{2\sigma - 1}} \quad (29)$$

$$\lambda_3 = \sum_{j=1}^{N/8} W_{1j} - \sum_{j=N/8+1}^{N/4} W_{1j} = \frac{1}{w}\left(\sum_{j=1}^{N/8} J_{1j} - \sum_{j=N/8+1}^{N/4} J_{1j}\right)$$
$$= 1 - \frac{N(2^{2\sigma} - 1)(2^{6\sigma-2} - 2)}{(2N)^{2\sigma} - 2N(2^{2\sigma} - 1) + 2^{2\sigma} - 1} \approx 1 - \frac{2^{8\sigma - 2}}{N^{2\sigma - 1}} \quad (30)$$

where the approximation in the last passages holds in the thermodynamic limit and we adopted the convention $1 = \lambda_0 \geq \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$. In general, one can see that $\lambda_l \approx 1 - 2^{2\sigma(l+1)-(l-1)}/N^{2\sigma-1}$.

Incidentally, we notice that λ_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 $h_i = \sum_j J_{ij} S_j$). More generally, λ_l corresponds to the field acting on spins in the l-th metastable state of the model [1, 2].

Moreover, as one can see from Eq. 28, λ_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 [1, 2]). In fact, ϕ_{λ_0} and ϕ_{λ_1} generate a subspace such that any vector in this subspace (hence writable as a linear combination of ϕ_{λ_0} and ϕ_{λ_1}) is an eigenvector of $\mathbf{W}$ of the same eigenvalue $\lambda = 1$. In

particular,

$$\phi_{\lambda_0} + \phi_{\lambda_1} = (\underbrace{1,1,\ldots,1}_{N/2},\underbrace{0,0,\ldots,0}_{N/2})\sqrt{2/N}, \qquad (31)$$

$$\phi_{\lambda_0} - \phi_{\lambda_1} = (\underbrace{0,0,\ldots,0}_{N/2};\underbrace{1,1,\ldots,1}_{N/2})\sqrt{2/N}, \qquad (32)$$

corresponds to a stationary state localized on the left and on the right branches of the graph, respectively.

Actually, it is easy to see that even λ_2 and λ_3 converge to 1 in the thermodynamic limit, although with slower rate. The degenerate eigenstate therefore allows for stationary distributions localized on smaller portions of the structure. Indeed, this can be proved to hold iteratively by including eigenstates of higher and higher order which allow eigenstates localized in smaller and smaller portions of the graph, provided that the hierarchical symmetry is fulfilled (thus this decomposition can not be pushed up indefinitely so to reach the lowest structures as dimers, because it would be unstable [2, 3]).

This means that if we initialize the stochastic process in any node i, as long as K is finite, the distribution describing the state of the graph will reach a stationary state broadened over the whole set of states V with equal probability, that is, the dynamic process on $\mathcal{G}$ is ergodic. But, as $K \to \infty$ the system tends to be localized only on the subset of nodes $V_1 = \{1, 2, ..., N/2\}$ if $i \le N/2$ or on the subset $V_2 = V \setminus V_1$ if $i > N/2$. More precisely, if the system is initially prepared according to the distribution $\phi_{\lambda_0} + \phi_{\lambda_1}$, in the thermodynamic limit it will never reach any node $j > N/2$, hence ergodicity is broken (and, correspondingly in statistical mechanics, metastable states become stable [1, 2]). If we simply assume that $K \gg 1$, then any localized state belonging to, say, the portion V_1, displays a characteristic time scale before broadening over the whole structure. The larger the subspace and the longer the time-scale; the larger K and the longer the time-scales. In the thermodynamic limit timescale diverges, conferring even to this (non-mean-field) ferromagnet a glassy flavour.

C. The spectral gap of the Hierarchical Ferromagnet

A close way to see the breakdown of ergodicity is to look at the spectral gap of the related Laplacian matrix. We already observed that the coupling matrix $\mathbf{J}$ is a block matrix, and every element J_{ij} represents the weight of the link connecting the nodes i and j. We can now introduce the Laplacian matrix, defined as $\mathbf{L} = \mathbf{Z} - \mathbf{J}$, where $\mathbf{Z}$ is a diagonal connectivity matrix such that $Z_{ii} = w, \forall i \in [1, N]$. The eigenvalues μ_i of $\mathbf{L}$ satisfy

$$0 = \mu_0 \le \underbrace{\mu_1}_{1} \le \underbrace{\mu_2 = \mu_3}_{2} \le \underbrace{\mu_4 = \mu_5 = \mu_6 = \mu_7}_{4} \le$$

$$\le \underbrace{\mu_8 = ... = \mu_{15}}_{8} \le ... \le \underbrace{...}_{N/2} \le 2w,$$

and we call spectral gap μ of $\mathbf{L}$ the smallest non trivial eigenvalue of $\mathbf{L}$ [13]. In particular, the smaller μ and the lower the number of links we need to cut so that the graph is divided in two independent blocks. In the hierarchical ferromagnet, we expect that this value tends to zero when the size of the system increases, obtaining the division of the network in two independent subgraphs, not interconnected. As depicted in Fig. 8, μ goes to zero exponentially with K according to $f(K) = e^{-a_\sigma K}$. By fitting numerical data we find that the rate a_σ decreases with σ meaning that the higher the value of σ and the lower the cost to fragment the graph.
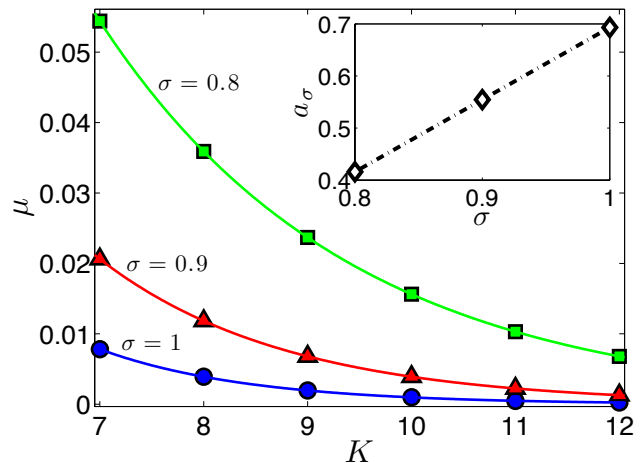

FIG. 8: (Color online) Spectral gap μ as a function of K (varying in $[7, 12]$) and of σ (with values 0.8, 0.9, 1). As expected, the spectral gap decreases with the system size and with σ. Data from numerical calculations (symbols) are fitted (solid lines) via the function $y = \exp(-a_\sigma x)$. The dependence on σ of the parameter a_σ, obtained from fitting procedures, is shown in the inset where the monotonic behaviour is highlighted.

We close with a remark. Once the size of the network fixed, the degree of modularity grows with μ. 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 μ. Therefore, consistently with [15], 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 $\mathcal{G}$ and let us generalize its coupling matrix $\mathbf{J}$ 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 ξ_i encoded by a vector of length P whose entries are dichotomic and defined stochastically (see Eq. 2). The coupling X_{ij}, arising by comparing ξ_i and ξ_j, is meant to mimic

a learning process hence correlating/uncorrelating (i.e. strongly/poorly connecting).

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

$$Q_{ij} = X_{ij}J_{ij}, \qquad (33)$$

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

$$X_{ij} = \sum_{\mu=1}^{P} \xi_i^\mu \xi_j^\mu, \qquad (34)$$

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 $\mathbf{X}$ is null. Basically, $\mathbf{J}$ favors couples which are close according to the *ultrametric* distance (defined on the set $\{i\}$), while $\mathbf{X}$ favors couples which are close according to the *Hamming* distance (defined on the set $\{\xi\}$) [4].

Notice that X_{ij} is a stochastic variable fulfilling a binomial distribution peaked at zero and with variance scaling linearly with P [30]. As both X_{ij} and J_{ij} are bounded we have

$$Q_{\max} = J_{\max}X_{\max} = J(1) \times P = \frac{P(1 - 4^{-K\sigma})}{4^\sigma - 1}(35)$$

$$Q_{\min} = -Q_{\max}, \qquad (36)$$

$$|Q|_{\min} = J_{\min}|X|_{\min} = 0. \qquad (37)$$

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}

$$P_{K,P}(Q_{ij} = q; \sigma) = P_{K,P}(X_{ij} = q/J_{ij}; \sigma)$$
$$= \frac{1}{\sqrt{2\pi P}} \exp\left\{ -\frac{q^2(4^\sigma - 1)}{2P[4^{\sigma(1-d_{ij})} - 4^{-K\sigma}]} \right\}, \quad (38)$$

where, exploiting the central limit theorem, we replaced the binomial distribution with a Gaussian distribution [31].

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 34) and we start the investigation by looking at how the distribution of weights $n(Q)$ is affected by the modulation induced by $\mathbf{X}$. Results for several choices of the parameters σ and P are shown in Fig. 9 (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 $\mathbf{J}$ only, here the set of possible values for weights Q_{ij} is $2P + 1$ or $2P$ times larger,

according to the parity of P:

$$J_{ij} \in \{J_1, J_2, \ldots, J_K\} \to Q_{ij} \in \{0, \pm 2J_1, \pm 4J_1, \ldots, \pm PJ_1,$$
$$\pm 2J_2, \ldots, \pm PJ_K\} \text{ if } P \text{ even}$$
$$J_{ij} \in \{J_1, J_2, \ldots, J_K\} \to Q_{ij} \in \{\pm J_1, \pm 3J_1, \ldots, \pm PJ_1,$$
$$\pm J_2, \ldots, \pm PJ_K\} \text{ if } P \text{ odd.}$$

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 $P - 1$ values whose occurrence follows a binomial distribution.

Notice that a large P implies a broader distribution; similarly, a small σ implies a larger support. Therefore, we expect that the pattern of $\mathbf{Q}$ 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. 10).
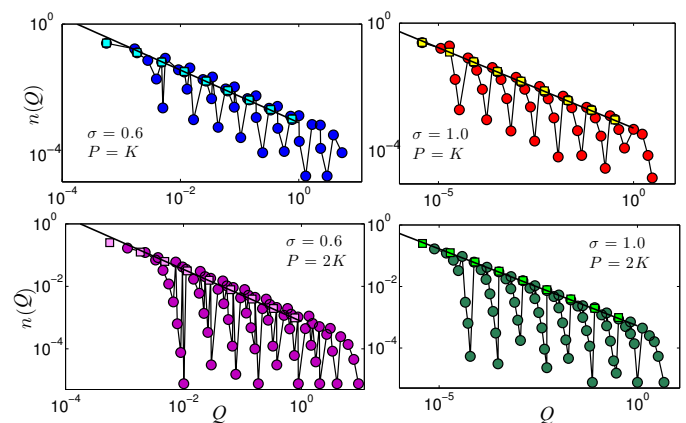

FIG. 9: (Color online) Distribution $n_K(|Q|)$ for $K = 9$ and different choice of the parameters σ and P as specified. Bullets represent data points for the graph generated by $\mathbf{Q}$ (see Eq. 33), squares represent the distribution one would obtain from the ultrametric contribution $\mathbf{J}$ only (see Eq. 38), and straight lines correspond to $y = J^{-1/(2\sigma)}/(2N)$ (see Eq. 13).

We now calculate the weighted degree of node i defined as

$$w_i = \sum_{j=1}^{N} Q_{ij} = \sum_{j=1}^{N} J_{ij}X_{ij}. \qquad (39)$$

Differently from the hierarchical ferromagnet 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 $w = \sum_{i=1}^{N} J_{ij}$) is normally distributed with variance scaling with P. This is indeed checked numerically as shown in Fig. 11.
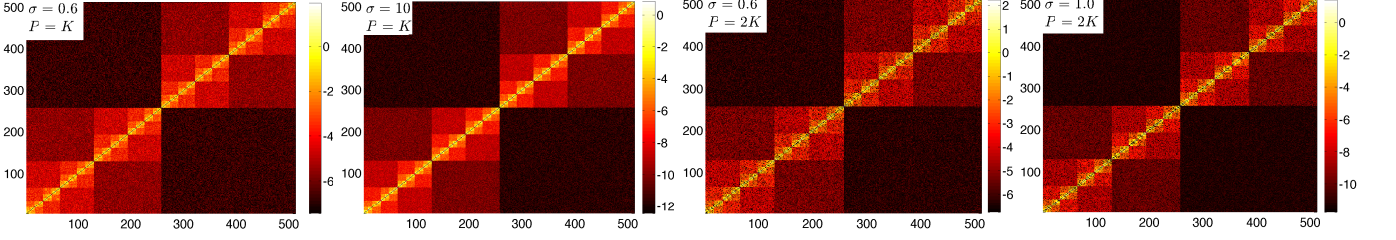
FIG. 10: (Color online) Patterns of weights Q_{ij} for $K = 9$ and different choices of σ and P as reported; each pattern refers to a different colorbar. Notice that for a better visibility we plotted $\log(|Q_{ij}|)$, and that, for a given σ, by increasing P the pattern gets more noisy; moreover, small values of σ imply a larger support. This figure mirrors Fig. 9.
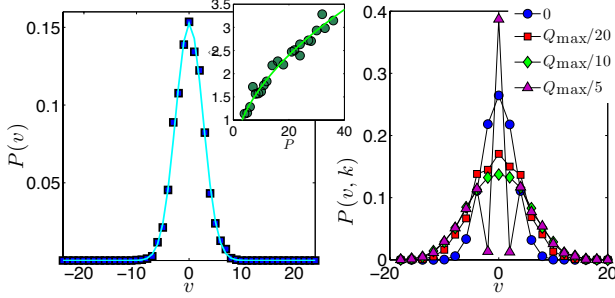

FIG. 11: (Color online) Left panel: in the main plot we show the distribution of weighted degrees $P(v)$ for a system with $K = 12$, $P = 2K$ and $\sigma = 1$. The distributions pertaining to different choices of parameters K and P were fitted with a central normal distribution with standard deviation $\hat{\mu}$. The best-fit values for σ are shown versus P in the inset where the solid line represents the function $y \sim \sqrt{x}$. Right panel: distribution $P(v, k)$ of generalized degree q in a network with $K = 12$, $P = 2K$, $\sigma = 1$ and different level of noise k. In particular, the set of noise is such that links corresponding to weights Q_{ij} smaller than a given threshold are cut; the threshold considered are represented with different symbols as explained by the legend.

In this case, since w_i is a random variable, we are interested in its mean and variance with respect to the random variables ξ. Recalling that

$$w_i = \sum_{i \neq j} J_{ij} X_{ij}, \text{ with } J_{ij} X_{ij} = J(d_{ij}) \sum_{\mu=1}^{P} \xi_i^\mu \xi_j^\mu, \quad (40)$$

one can see that the expected (according to the distribution in Eq. 2) value of w_i is $\mathbb{E}(w_i) = 0$ and, computing the variance of w_i, we obtain

$$\text{Var}(w_i) = \sum_{i \neq j} J(d_{ij}) \text{Var} \left[\sum_{\mu=1}^{P} \xi_i^\mu \xi_j^\mu \right] = \sum_{i \neq j} P \times J(d_{ij}) =$$

$$= \frac{P}{2(4^\sigma - 1)} \left[\frac{(2N)^{2\sigma} - N(3 \times 2^{2\sigma - 1} - 1) + 2^{2\sigma} - 1}{N^{2\sigma}(2^{2\sigma - 1} - 1)} \right],$$

where P is the variance of X_{ij}.

The next step is to evaluate the degree of modularity. Differently from the hierarchical ferromagnet, where

there is a perfect homogeneity in the weight of nodes, here we expect the regular ultrametric structure to be perturbed by the stochastic factor $\mathbf{X}$. The generalised overlap matrix $\mathbf{O}$ (see Eq. 20 for its extended formula) is computed and shown in Fig. 12, along with a dendrogram plot capturing the dissimilarity (see Eq. 21) between nodes. For the realization considered the highest values of overlap are still obtained for dimers, yet the resulting structure is not fully regular as previously found for the hierarchical ferromagnet 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 ξ_i^μ, $\mu = 1, ..., P$ peaked at zero implies that it is no longer 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.

V. GRAPH GENERATION IN THE HIERARCHICAL SPIN GLASS

This section is devoted to the study of the generation of the weighted graph $\mathcal{G}$ in the case of hierarchical spin glass. As introduced in Sec. II, in this case the couplings among spins are defined as quenched variables drawn from a standard centered Gaussian distribution $\mathcal{N}[0, 1]$. This means that we can write

$$Q_{ij} = \chi_{ij} J_{ij} = \chi_{ij} \frac{4^{\sigma(1-d_{ij})} - 4^{-K\sigma}}{4^\sigma - 1}, \quad (41)$$

where χ_{ij} are independent, centered Gaussian variables. Due to the contribution of χ_{ij} in the definition (41), the weight of nodes is site dependent. More precisely, we have that the expected value of w_i is $\mathbb{E}_{\chi_{ij}}[w_i] = 0$ and its variance reads as

$$\text{Var}_{\chi_{ij}}[w_i] = \text{Var}_{\chi_{ij}} \left[\sum_{i \neq j} \chi_{ij} J_{ij} \right] =$$

$$= \frac{(N-1)}{2(4^\sigma - 1)} \left[\frac{(2N)^{2\sigma} - N(3 \times 2^{2\sigma - 1} - 1) + 2^{2\sigma} - 1}{N^{2\sigma}(2^{2\sigma - 1} - 1)} \right],$$
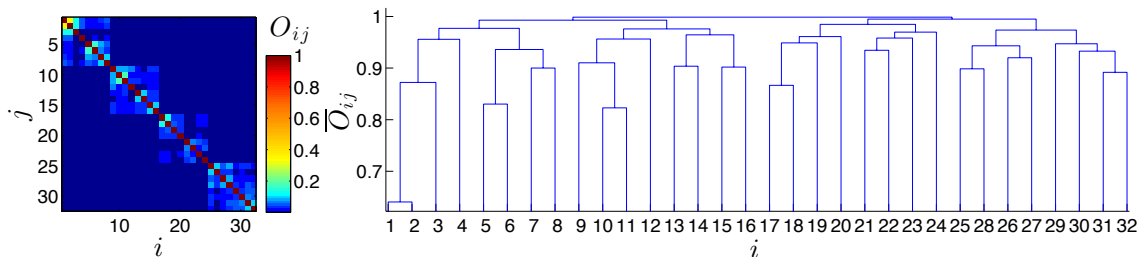
FIG. 12: (Color online) Representation of modularity between nodes using the topological overlap matrix $\mathbf{O}$ for fixed $K = 5$, and $\sigma = 0.9$. Left panel: Matricial representation of overlaps obtained via Eq. 20. Different colours represent different values of overlap, as explained by the colorbar on the right. Due to the presence of random variables in the construction the coupling matrix $\mathbf{Q}$, we can see partial loss of regularity in the structure of the matrix. Right panel: dendrogram showing the dissimilarity between nodes in the graph: nodes at distance $d_{ij} = 1$ (e.g., node $i = 3$ and $j = 4$) typically display high overlap (hence have low dissimilarity) with respect to those at distance $d_{ij} = 2$ (e.g., $i = 13$ and $j = 15$), up to the maximum distance $d_{ij} = 5$ (e.g., $i = 1$ and $j = 24$), underlying the ultrametric structure of the network. However, differently from the case of hierarchical ferromagnet, this case is irregular and overlaps, especially between nodes at close distance, are broadly distributed.

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 $\mathbb{E}_{\chi_{ij}}[\overline{w}] = \mathbb{E}_{\chi_{ij}}[\frac{\sum_{i=1}^{N} w_i}{N}]$ are shown in Fig. 13.
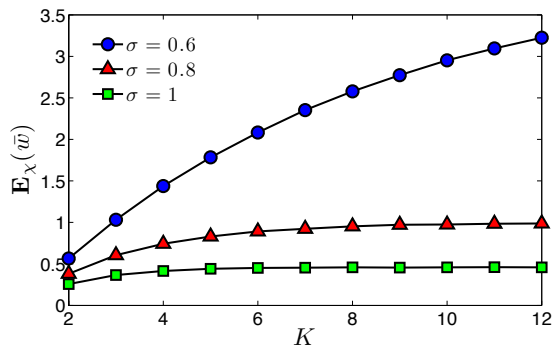

FIG. 13: (Color online) Plot of $\mathbb{E}_{\chi_{ij}}[\overline{w}] = \mathbb{E}_{\chi_{ij}}[\frac{\sum_{i=1}^{N} w_i}{N}]$ as a function of the system size ($N = 2^K$) and of σ (as explained in the legend). To realize it, 200 realization of χ_{ij} were produced, and, for each, we obtained $w_i, \forall i \in [1, N]$. Then, the algebraic mean value over the realizations was computed, with fixed K. As expected, the largest values of w are obtained for the highest values of K and for the lowest values of σ.

We also checked the modularity of this networks by exploiting again the generalised topological overlap matrix [19] given by Eq. 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 definition of the model: the presence of the random variables χ_{ij}, that contribute to construct the coupling matrix, introduces a random component that affects the overlap between dimers, squares, octagons ect., as depicted in Fig 14.

The remarks highlighted for the neural network model

(see the conclusion of the previous section) remain valid for the hierarchical spin-glass model as well: the presence of weights on links depending on random variables leads to a loss of symmetry in the structure of the network: the links favoured by the ultrametric distance are not necessarily the same favoured by the random coupling χ_{ij}.

VI. CONCLUSIONS AND OUTLOOKS

In the last decade hierarchical networks have been found to play a crucial and widespread role in natural phenomena [21, 22], particularly in biological systems [23, 24]. Furthermore, these structures turn out to be also quasi-tractable in statistical mechanics [14], even when glassiness is present [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 couplings among nodes: Herarchical Ferromagnet (HFM), Hierarchical Neural Network (HNN) and Hierarchical Spin Glass (HSG). In particular, we show that the subtle metastabilities exhibited by HFM (see e.g. [1, 2]) can also be evidenced in terms of ergodictiy breakdown for Markov processes defined on the hierarchical weighted graph embedding the spin system.

More precisely, 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.

Further, these structures also exhibit high clustering (at least according to the definition (23)) and modularity which are two important properties well-evidenced in many real systems [8, 17].
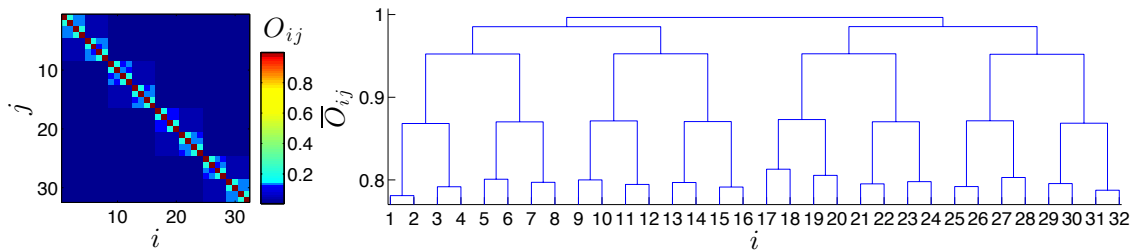
FIG. 14: (Color online) Representation of modularity between nodes using topological overlapping measure when $K = 5$, $\sigma = 0.9$. Left panel: Matricial representation of overlaps using obtained via Eq. 20. Different colors represent different values of overlap, as explained by the colourbar on the right. Due to the presence of Gaussian random variables in the construction the coupling matrix, we can see again its random behavior, but the network anyway results more regular than the hierarchical neural network. Right panel: dendrogram showing the dissimilarity between nodes in the graphs: nodes at distance $d_{ij} = 1$ (see for instance node $i = 1$ and $j = 2$, at the lower level of the dendrogram) have higher values of overlap (hence have lower values of dissimilarity) with respect to those at distance $d_{ij} = 2$ (second level from the bottom of the dendrogram), up to the maximum distance $d_{ij} = 5$ (first level on the top). As shown for the hierarchical neural network, also the hierarchical spin-glass partially loose the symmetric structure for the presence of random variables that contribute to tune the elements of adjacency matrix.

Analogous 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 loss 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 weighted degree of nodes, that is peaked at zero and with variance scaling with P. For the HSG, the Gaussian term leads to a weighted degree peaked at zero and with variance scaling with N. In general, the Hebbian contribute induces a smaller broadness but a larger noise.

These investigations deepen the strong connection between the thermodynamic behaviour of a statistical mechanics model and the topological properties of the underlying structure. The ultimate goal is to contribute to the description of non-mean-field systems.

Appendix A: Notes about modularity

In this section we deepen the analysis of the network modularity introduced in Sec. III for the hierarchical ferromagnet.

In order to figure out which is the most effective partition, we can apply the formula introduced in [17] (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 $\bar{m}$, i.e. $c_i = 1, ..., \bar{m}$. More precisely,

$$M = \frac{1}{m} \sum_{ij} \left[J_{ij} - \frac{w_i w_j}{m} \right] \delta(c_i, c_j), \qquad (A1)$$

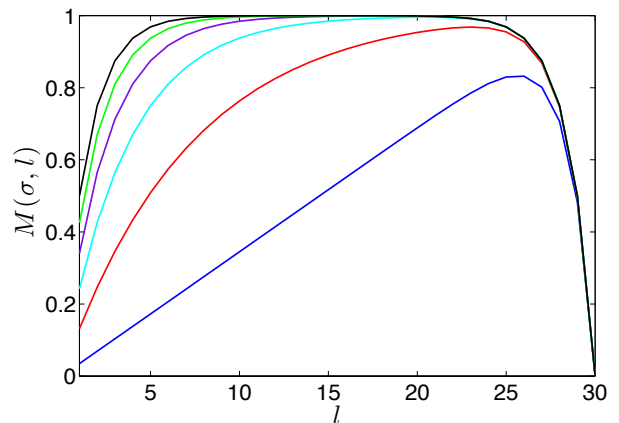where $m = \sum_i w_i$. In particular, exploiting the homogeneity of the hierarchical graph we can write Eq. (A1)

FIG. 15: (Color online) Degree of modularity $M(\sigma, l)$ for $K = 30$ and different values of σ. Curves in different colors correspond to different values of σ: as σ is varied from 0.5 to 1.0 the peak moves from right to left; see Eq. A3.

in a simpler form as

$$M(\sigma, l) = \frac{1}{Nw} \sum_{i<j} \left[J_{ij} - \frac{w}{N} \right] \delta(c_i, c_j). \qquad (A2)$$

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

$$
\begin{aligned}
M(\sigma, l) &= \frac{1}{Nw} \sum_{d=1}^{l} \left[J(d) - \frac{w}{N} \right] 2^{l+d-1} \qquad (A3) \\
&= \frac{2^l}{2^K} \left(\frac{w_l(\sigma)}{w_K(\sigma)} + \frac{2^l - 1}{2^K} \right) = \\
&= \frac{t^K}{t^l} \frac{\Gamma(t, l)}{\Gamma(t, K)} + \frac{2^l(2^l - 1)}{2^{2K}},
\end{aligned}
$$

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

$$w_l(\sigma) = \sum_{d=1}^{l} 2^{d-1} J(d);$$
$$\Gamma(t,j) = 2^j - 1 + t - 2^{j+1}t + 2^j t^{j+1}, j \in [1, K].$$

As shown in Fig. 15, the function $M(\sigma, l)$ 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. A1) is the one where the graph is divided in a relatively small number of clusters, but for large σ (namely where the hierarchy is less important, see Eq. 8) this number gets smaller.

Finally, we introduce an alternative formulation for extending the formula introduced in [19] and reported in Eq. 19. In fact, exploiting the discreteness of the entries of the coupling matrix $\mathbf{J}$ we can write

$$O'_{ij} = \frac{\sum_{l=1}^{K} |N_l(i) \cap N_l(j)| J(l)}{\min\{w_i, w_j\}}, \qquad (A4)$$

where $N_l(i)$ is the number of links with coupling $J(l)$ stemming from node i. In particular, the expression in Eq. A4 can be applied to the hierarchical ferromagnet, obtaining the following

$$O'_{ij} = \frac{(2N)^{2\sigma} 2^{d(1-2\sigma)} + 2N(1 - 2^{2\sigma}) + 2^d(2^{2\sigma} - 1)}{(2N)^{2\sigma} - 2N(2^{2\sigma} - 1) + 2^{2\sigma} - 2}, \tag{A5}$$

where to ligthen the notation we posed $d = d_{ij}$. In the thermodynamic limit $O'_{ij} \sim 2^{-d(2\sigma-1)}$, namely the similarity between two nodes decreases exponentially with their distance.

[1] E. Agliari, A. Barra, A. Galluzzi, F. Guerra, D. Tantari, F. Tavani *From Dyson to Hopfield: Processing on hierarchical networks*, Phys. Rev. Lett. **114**, 028103 (2015)
[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*, accepted for publication in Neural Networks, 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. Barthélemy, R. Pastor-Satorras, A. Vespignani, *T*he architecture of complex weighted networks Proc. Natl. Acad. Sc. 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. Parisi, *The Hierarchical Random Energy Model*, Phys. Rev. Lett. **104**, 127206, (2010).
[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*, J. Stat. 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 netwirks*, 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] B. Zhang, S. Horvath, *A General Framework for Weighted Gene Co-expression Network Analysis*, Statistical Applications in Genetics and Molecular Biology, **4**, 17 (2005).
[21] 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).
[22] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.U. Hwang, *Complex networks: Structure and dynamics*, Phys. Rep. **424**, 4-5, 175-308, (2006).
[23] P. Moretti, A.M. Munoz, *Griffiths phases and the stretching of criticality in brain networks*, Nature Comm. **4**, 2521, (2013).
[24] S. Maslov, K. Sneppen, *Specificity and Stability in Topology of Protein Networks*, Science **293**, 910-913, (2002).
[25] J. Saramäki, M. Kivelä, J.-P. Onnela, K. Kaski, J. Kertész, *G*eneralizations of the clustering coefficient to weighted complex networks, Phys. Rev. E **75**, 027105 (2007).

[26] J.-P. Onnela, J. Saramäki, J. Kertész, and K Kaski, *I*ntensity and coherence of motifs in weighted complex networks, Phys. Rev. E **71**, 065103(R) (2005).

[27] 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 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: $||x||_p = p^{-\mathrm{ord}_p(x)}$: where, if x is a natural number, then $ord_p(x)$ 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.

[28] This alternative definition of distance can also be useful for the numerical implementation of the hierarchical graph under study.

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

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

[31] 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 -trough this perspective- the topology of the hierarchical spin glass too, we allow ourselves in taking such an approximation.