

Tree networks of real-world data: Analysis of efficiency and spatiotemporal scales

Davide Cipollini¹* and Lambert Schomaker²

*Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, [University of Groningen](#), Groningen, Netherlands
and Cognigron - Groningen Cognitive Systems and Materials Center, [University of Groningen](#), Groningen, Netherlands*



(Received 4 February 2025; accepted 8 June 2025; published 1 July 2025)

Hierarchical tree structures are common in many real-world systems, from tree roots and branches to neuronal dendrites and biologically inspired artificial neural networks, as well as in technological networks for organizing and searching complex datasets of high-dimensional patterns. Within the class of hierarchical self-organized systems, we investigate the interplay of structure and function associated with the emergence of complex tree structures in disordered environments. Using an algorithm that creates and searches trees of real-world patterns, our work stands at the intersection of statistical physics, machine learning, and network theory. We resolve the network properties over multiple phase transitions and across a continuity of scales, using the von Neumann entropy, its generalized susceptibility, and the recent definition of thermodynamic-like quantities, such as work, heat, and efficiency. We show that scaleinvariance, i.e., power-law Laplacian spectral density, is a key feature for constructing trees capable of combining fast information flow and sufficiently rich internal representation of information, enabling the system to achieve its functional task efficiently. Moreover, the complexity of the environmental conditions the system has adapted to is encoded in the value of the exponent of the power-law spectral density, which is inherently related to the network spectral dimension and directly influences the traits of those maximally functionally efficient networks. Thereby, we provide a novel metric to estimate the complexity of high-dimensional datasets.

DOI: [10.1103/ztmk-glc4](https://doi.org/10.1103/ztmk-glc4)

I. INTRODUCTION

Why complex structures emerge in a real-world environment is a fascinating question that has not yet found a fully definitive answer. In a recent article [1], the emergence of complex structures is devised as the macroscopic thermodynamic result of, on the one side, the optimal ability to store information about the environment, on the other side, the ability to access and transmit information across the structure [2] (see Methods, Sec. II A 2). For instance, in living systems, information about the environment flows into the system's genome through natural selection. The genome encodes such information into the organism's structural traits, which in turn reflect the functional capabilities of the system. Thus, the system's very own structure may be regarded as the internal representation of the environment [3–7], and its structural traits could themselves enclose the complexity of the environmental conditions the system needs to adapt to.

In this light, the emergence of functional topological configurations, that is, structures enabling distinctive performance in a task, may be seen as the macroscopic outcome of the maximization of the outlined thermodynamic functional, encoding

the trade-off between information content and its accessibility. This maximization may explain the emergence of a set of system configurations that have adapted to, and perform their function within, a specific complex environment. Crucially, the degree of function, defined as the system's quantitative ability to perform its function with minimal resources, is *context dependent*, as a configuration that is functional in one environment may not be so in another.

Moreover, identifying the topological features that maximize the system's adaptation to the environment and its functional performance remains an open question. The hypothesis of optimal functioning at the border of order and disorder suggests critical-like features as the potential candidates to obtain functional advantages [8–14] and maximize the fitness to the outer world, or, when in the context of learning machines, to obtain a better performance in a statistical generalization problem [10]. Among the various attributes exhibited by truly critical systems and systems with critical-like properties, especially relevant to our work is the property of scale invariance, exhibited through the power-law distribution of some meaningful quantity of the system under consideration.

Hierarchical trees enable us to investigate, within a relevant class of systems, the interplay between structure and function associated with the emergence of complex structures in disordered environments. Trees are common to many real-world systems, from tree roots and branches, to neuronal dendrites [15–19] and biologically inspired artificial neural networks [20–25], as well as in technological networks for organizing and searching collections of high-dimensional pat-

*Contact author: d.cipollini@rug.nl

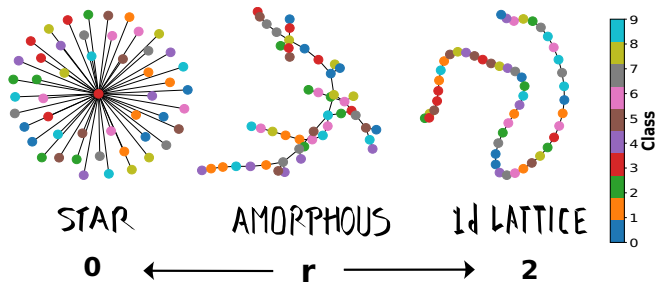


FIG. 1. Tree structures generated by the Portegys algorithm. The algorithm creates star-like networks up to a 1D lattice when $r \rightarrow 2$. For intermediate values of r , networks resembling an arborescent tree are produced. The process is reminiscent of the transition *vapor/liquid* $\rightarrow$ *amorphous solid* $\rightarrow$ *1D lattice*. Depicted networks are generated from a subset of 50 samples randomly extracted from the MNIST dataset with $r = 0.05, 0.95, 1.50$, respectively. Classes are indicated by the 10 different colors in the color bar.

terns [26–28]. In this paper, the term *pattern* refers to a vector in $\mathbb{R}^p$, where p is the number of pixels.

We use the Portegys algorithm [27,28] to produce hierarchical trees that self-organize given a collection of high-dimensional patterns. The algorithm recursively organizes (*structural aspect*) high-dimensional pattern vectors randomly drawn from a dataset into trees (see in Methods, Sec. II B 1). Using the metaphor introduced by Jacob for natural selection [29], the Portegys tree-generation process acts as a *tinkerer* that creates a structure using what is available at the time, adding one node after another during the tree formation. Each node stores a copy of the original pattern vector along with its class label, and its placement in the growing tree is determined by a parameterized, distance-based rule.

The algorithm includes a search phase, endowing the system with a *functional aspect*: It navigates the tree structure to perform pattern matching, identifying the stored pattern most similar to a novel input in order to infer its class.

The algorithm is simple and explainable; it produces trees with a controllable number of nodes, and it is subject to a single control parameter r governing both the number of hierarchical levels and the branching. It is capable of generating a large variety of tree structures (Fig. 1) and multiple phase transitions (Sec. III C).

Other models can generate hierarchies, such as the model in Ref. [30], where a single parameter controls the branching and the hierarchical depth, producing structures in much analogy to the ones in Fig. 1, or the Galton–Watson stochastic process, originally proposed to model the extinction of family names [31,32].

Nevertheless, the Portegys algorithm incorporates two key attributes central to our analysis: the *system’s functional aspect*, defined by the pattern-matching task, and the *external disorder* introduced by the dataset. Thanks to these features, we can work at the intersection of three disciplines: statistical physics, machine learning, and network science. We can show that information-theoretical tools based on network density matrices [33–35] (see Methods, Sec. II), while monitoring the structural properties of the system, reflect the functional capabilities of the network.

We show that emerging efficient structures exhibit a scale-invariant spectrum [36], i.e., power-law Laplacian spectral density. We reproduce this result over multiple datasets, ranked based on their complexity. Our findings suggest that a large range of availability in the power-law spectral density provides the advantage of a large repertoire of network “frequencies”. These can be exploited to achieve the functional purpose of the system efficiently, akin to critical-like systems exhibiting scale invariance in the distribution of a relevant quantity [11,13,14,37,38]. Such behavior is observed consistently, whether we take efficiency as gauged by the *task-agnostic* network thermodynamic efficiency [1], estimating the trade-off between the diversity of information pathways and fast information propagation (see Sec. II A 2 in Methods), or the *task-related* pattern-matching efficiency (Sec. II B 2 in Methods), quantifying the trade-off between classification accuracy and the number of visited nodes throughout the pattern-matching task.

Furthermore, the exponent of the power-law spectral density, related to the network generalized dimension [39–42], results in being driven by the complexity of the environment the system adapted to, directly influencing the traits of those functionally efficient networks.

Finally, from a complementary perspective, using a hierarchical topology to map the distribution of high-dimensional patterns of a dataset provides insights into the structured disorder of the data [43]. In particular, we measure the dataset network’s spectral dimension and show that it estimates the dataset’s complexity. Thus, among the other results, our work provides an alternative way to gauge the dataset’s complexity [43–48].

We outline the structure of this paper to facilitate navigation.

In Sec. II, we provide a detailed discussion of the methods:

In Secs. II A 1 and II A 2, we present the theoretical framework underpinning our analyses of network formation. We introduce, among the others, the thermodynamic-like quantities and a variational principle explaining a *network formation process from a macroscopic point of view* [1].

In Sec. II B, we describe the *microscopic rules* of the tree-formation process dictated by the Portegys’ algorithm [27]. We also detail the search phase, which implements the functional aspect of the algorithmic system through the pattern-matching task.

In Sec. II C, we describe the datasets and the preprocessing steps used before our evaluations. These datasets serve as the disordered environments, each exhibiting a distinct level of *complexity*, to which the algorithmic system adapts during the tree formation process to later actuate its function.

In Sec. III, the results are presented:

In Sec. III A, the datasets are ranked based on their complexity using a variety of methods for cross validation.

In Sec. III B, we quantify the *degree of function* of the system by defining the task-related pattern-matching efficiency, which measures the trade-off between the performance achieved and the resources used to accomplish the task.

In Sec. III C, we characterize the *phase portrait of the algorithm*: We identify the phase transitions occurring, and we resolve the structural features over a continuity of scales using the information-theoretic tools outlined in Sec. II A 1.

In Secs. III D to III F, we evaluate the system's performance across different levels of environmental complexity. We narrow down the r -region in phase space associated with higher functional (pattern-matching) efficiency by maximizing the thermodynamic-like efficiency. From a complementary perspective, we use a *degree-of-function-based variational principle* to infer the spatial scales, that is, the r -scales, that resolve the intrinsic geometry of each dataset.

In Sec. III G, we conclude our narrative of the results by showing that the complexity of the dataset is encoded in the structural traits of the maximally efficient networks. Consequently, we estimate each dataset's complexity by measuring the spectral dimension of the networks that maximize the efficiency.

In Sec. IV, we discuss our results and highlight the possible broader implications across various scientific domains.

The code and the datasets used in this study are open source and available at Refs. [49,50].

II. METHODS

A. Theoretical framework

In this section, we outline the theoretical framework underlying our analyses and results. Specifically, we provide a comprehensive discussion of recently developed theoretical tools [1,33–36,51], grounded in the definition of the network density matrix [Eq. (1)]. Based on this definition, the von Neumann entropy [Eq. (2)] and its generalized susceptibility are introduced. These will be used in Sec. III to characterize the algorithm's phase portrait. We then introduce the thermodynamic-like efficiency [Eq. (6)], which frames the formation process as the macroscopic outcome of a variational principle that maximizes efficiency. Furthermore, we discuss the spectral dimension of a disordered network and relate it to the generalized entropy susceptibility. The spectral dimension will enable us to decode the dataset's complexity from the network's topological traits in Sec. III.

1. Statistical physics of networks

Information spreading over networked structures is governed by the Laplacian matrix [52], and its spectrum of eigenvalues encodes many relevant topological properties of the graph [53,54]. Let us consider an undirected and unweighted network $G(V, E)$ of order $|V| = N$ with the number of links equal to $|E| = m$. The network Laplacian is defined as $\hat{L} = \hat{D} - \hat{A}$, where $\hat{D}$ and $\hat{A}$ are the diagonal matrix of the node degrees and the adjacency matrix, respectively. The Laplacian operator can be used to define a density operator [33], where the Laplacian takes the role of the Hamiltonian of statistical physics [37] as in Eq. (1).

Given an initial state of the network encoding the concentration of a quantity, e.g., information, in one or more nodes, $\underline{s}(0)$, the state at time τ is given by $\underline{s}(\tau) = e^{-\tau\hat{L}}\underline{s}(0)$. Where $\hat{K} = e^{-\tau\hat{L}}$ is the network propagator and is the discrete counterpart of the path integral formulation for general diffusion processes [55,56].

As $\hat{K}$ is a positive semidefinite Hermitian matrix, it admits the spectral decomposition in terms of the sum of projectors $\hat{K} = \sum_i^N e^{-\tau\lambda_i} |k_i\rangle\langle k_i|$. Thus, for any state at τ

it holds $|\phi(\tau)\rangle = \sum_i^N e^{-\tau\lambda_i} |k_i\rangle\langle k_i| |\phi(0)\rangle$. This relation emphasizes that the eigenvectors of the Laplacian are akin to the wave vectors in the Fourier space defined by the spectrum of the continuum Laplacian. At the same time, τ is coupled by the diffusion equation to the resolved network “wavelengths”, thus it defines an operational resolution scale for observations [35]. Moreover, the ensemble of accessible diffusion modes at time τ can be constructed by defining a probability measure over the accessible modes decomposed by the spectral theorem. Such a canonical ensemble [33,34,57] is encoded into the density matrix,

$$\hat{\rho}(\tau) = \frac{e^{-\tau\hat{L}}}{Z}. \quad (1)$$

The elements ρ_{ij} represent the normalized amount of information transferred in a diffusion process between nodes i and j at time τ . Remarkably, ρ_{ij} takes into account all possible topological pathways between the two nodes, and assigns smaller weights to longer ones tantamount to path integrals [34,55]. The partition function $Z = \text{Tr}[\hat{K}] = \sum_{i=1}^N e^{-\tau\lambda_i}$ is a function of the eigenvalues, λ_i , of the Laplacian and is proportional to the average return probability of a continuous time edge-centric random walker to be back in its initial location after time τ [57]. Note that the object $\hat{\rho}$ is a positive semi-definite and Hermitian matrix whose trace sums to unity, allowing its eigenvalues to be interpreted as probabilities.

Finally, the von Neumann spectral entropy can be introduced [33] as a multiscale descriptor of the network structure,

$$S(\hat{\rho}(\tau)) = - \sum_{i=1}^N \mu_i(\tau) \log \mu_i(\tau), \quad (2)$$

where $\{\mu_i(\tau)\}_{i=1}^N$ is the set of $\hat{\rho}(\tau)$ eigenvalues that are related to the set of Laplacian eigenvalues through $\mu_i(\tau) = e^{-\tau\lambda_i}/Z$. The entropy $S(\hat{\rho}(\tau))$ is a function of the normalized time τ and is bounded between $[\log C_{\text{conn}}, \log N]$ where C_{conn} is the number of connected components and N is the number of nodes. Thus, the von Neumann entropy S as a function of τ reflects the *entropic phase transition* of information propagation (or loss) over the network [34]. Specifically for a connected network, at $\tau \rightarrow 0$, $S(\tau) = \log N$ reflects the *segregated heterogeneous phase* where information diffuses from the single nodes to the very local neighborhood; at $\tau \rightarrow \infty$, $S(\tau) = 0$ and the diffusion is governed by the smallest nonzero eigenvalue of the Laplacian, associated with the so-called Fiedler eigenvector, reflecting the *integrated homogeneous phase* where the information has propagated all over the network.

The network entropy S thus counts the (logarithmic) number of informationally unconnected nodes in the heterogeneous network. It is possible to identify the characteristic resolution scales of the network [34,35] by inspection of the entropy S , and through the definition of its generalized susceptibility (or specific heat),

$$C = - \frac{dS}{d \log \tau}. \quad (3)$$

The specific heat C identifies transition points in the information diffusion process reflecting the intrinsic diffusion

τ -scales of the network. For instance, a single large peak of the specific heat C identifies a single scale at which there is a fast and unique information diffusion (or information loss) across the network. The latter is the typical situation of a star graph, depicted in Fig. 1, or an Erdős-Rényi network. Before the peak, the informationally connected components are the single nodes, while after the peak, the informationally connected components are the whole network, thus revealing a single trivial scale.

On the contrary, a constant specific heat reflects the *scale invariance* of the network in a resolution- τ region where information is lost at a constant rate. This situation is typical of homogeneous D -dimensional lattices and of fractal heterogeneous networks whose spectral density $P(\lambda)$ exhibits asymptotic power-law behavior in the infrared limit, $P(\lambda) \sim \lambda^\gamma$ as $\lambda \rightarrow 0$. Scale invariance may be defined through a more severe condition of power-law behavior over the whole spectrum support in the infinite network-size limit. Still, the outlined asymptotic equivalence provides a more readily detectable condition in both finite-size synthetic and real networks [36].

The specific heat plateau height C_0 is also related to the spectral dimension $d_s = 2(\gamma + 1)$ [39,41], according to the relation $d_s = 2C_0$ [35,36], which implies $C_0 = \gamma + 1$. The spectral dimension generalizes the concept of the Euclidean dimension to disordered structures and coincides with the Euclidean dimension for lattices. For instance, we expect $C_0 = 1/2$ for 1D lattice; $C_0 = 1$ for the regular grid with periodic boundary conditions; and $C_0 = 4/6$ for random trees with finite-degree variance $\langle k^2 \rangle$ showcasing fractal dimension $d_s = 4/3$ [36].

From these considerations, a Laplacian-renormalization group theory has been recently developed, allowing for a coarse graining of the network structures into super-nodes and for the unraveling of the mesoscale structures [34,35]. The theory relies on information diffusion to identify *informationally equivalent blocks* akin to the ones proposed by Kadanoff [58], and moves to the k -space to explain the procedure akin to Wilson's formulation [59]. In this context, the first peak of the specific heat C or a marked change in its slope, identifies the most convenient resolution scales where to perform the coarse-graining over the informationally connected nodes, as it allows the iteration of the procedure without an excessive network reduction, and, when the network is scale invariant, maintaining the dilation symmetry across the renormalization group procedure [37].

To conclude, intuitively, we can use the diffusion parameter τ to resolve the network's topological structures at different resolution scales. Thus, the necessity of analyzing complex networked systems at all scales is solved by resorting to the conjugate Fourier space, the k -space, defined by the Laplacian spectrum and the "wavelengths" of its modes.

To use an analogy from an experimental setup and provide an intuitive understanding of the role of τ , let us suppose we want to probe the features of a material with light. If we aim to resolve fine-grained structures, we would use high-frequency (short-wavelength) light. Conversely, to examine large-scale features, we would use low-frequency (long-wavelength) light. The parameter τ plays an analogous role: it effectively sets the wavelength of the probe used to

analyze the network structure, where large τ will probe large-scale features (e.g., number of connected components in a network); short τ will probe short-scale features (e.g., node's degrees and very local neighbors); while intermediate τ will probe mesoscale features on a continuous scale (e.g., clusters, network's compartmentalization into modules). We refer to Fig. 1 in Ref. [60] for a visual illustration.

2. Thermodynamics of networks

The statistical physics theory for networks introduced above allows us to view network phenomena from a new perspective and possibly to bridge the gap to network macroscopic observables. In this regard, network thermodynamic quantities such as work W and heat Q have been defined in mathematical analogy to their thermodynamic counterparts [1]. Since a network of isolated nodes is the only network with maximum entropy $S_{\text{iso}} = \log N$ for $\tau > 0$, progressively adding edges diminishes the network's entropy. Lower entropy reflects the diminished ability of the network to express information about a stochastic perturbation, or in other terms, the loss in the internal representation capabilities encoded in the network's topological traits. Thus, the entropic cost is given by

$$Q = \frac{S - S_{\text{iso}}}{\tau} = \frac{S - \log N}{\tau} \leq 0, \quad (4)$$

which mathematically resembles the heat dissipation in a thermodynamic process.

On the other hand, adding edges increases the gain in information flow quantified by the (Helmholtz) free energy $F = -\log Z/\tau$ [1,57]. Less obvious is how, for a fixed number of edges, the degree distribution influences the flow of information. As the Laplacian governs the continuous-time *edge-centric* random walk, the presence of large hubs implies an enhancement of the process with faster propagation of the diffusing quantity [52,61]. As the free energy is inversely related to the partition function, $Z = \sum_{i=1}^N e^{-\tau \lambda_i}$, the bounds of the former are given by the bounds of the latter. The partition function is bounded as $1 \leq Z \leq N$, where the lower bound corresponds to the connected network at large $\tau \gg 1$, and the upper bound is given by either a network of isolated nodes $\forall \tau$, or by a connected network at small scales ($\tau = 0$). Thus, the $-\log N/\tau \leq F \leq 0$ and the gain in information flow is reflected by the variation of the free energy from the case of isolated nodes,

$$W = F - F_{\text{iso}} = F + \frac{\log N}{\tau} \geq 0, \quad (5)$$

which is tantamount to the work in the canonical ensemble.

Moreover, $U = W + Q \geq 0$, which implies $W \geq |Q|$. A formation process can be viewed as the trade-off between these two necessities of the networked complex system: the need to maximize the information flow, while at the same time retaining an amount of structural entropy such that the network can encode and express sufficient diversity through its traits.

Thus, to complete the analogy to thermodynamics, the trade-off between the diversity of information and information

flow is captured by the efficiency [1],

$$\eta = \frac{W + Q}{W} = 1 - \frac{|Q|}{W} = \frac{U}{W}, \quad (6)$$

where $0 \leq \eta \leq 1$ and $U \equiv -\partial_\tau \log Z \equiv \text{Tr}[\hat{L}\hat{\rho}]$. For scale-invariant networks, for some τ -scales where the specific heat shows a plateau, $U = \langle \lambda \rangle_\tau = C_0/\tau$ [35,36]. Hence, in this regime, the efficiency can be expressed in terms of the spectral dimension, $\eta = \frac{1}{\tau} \frac{C_0}{W} = \frac{1}{\tau} \frac{d_c}{2W}$. The variational principle here exposed has been devised in Ref. [1] to macroscopically explain the sparsity and the empirical scaling law relating the number of nodes and the number of edges, as well as the emergence of topological features such as modularity, small-worldness, and, with partially less clear results, heterogeneity.

In our work, we will show that the investigated tree-network traits reflect the information about the environment (i.e., a real-world dataset of patterns), which the system has extracted to perform its function. We will partly use the outlined thermodynamic variational principle to explain the structural properties of the network subclass composed of heterogeneous hierarchical trees, and the specific heat will be used to identify the network τ -resolution scales at which such a variational principle is effective for our subclass of networks.

B. Portegys algorithm

The Portegys algorithm [27] was designed to efficiently identify the stored patterns that are nearest neighbors to a novel test pattern.

The algorithm has two phases: the tree formation process and the search phase. The latter performs the pattern-matching function. During the tree formation, labeled patterns from the training set are copied into pattern-bearing nodes, which are dynamically organized into a growing tree. The final tree will contain as many nodes as there are patterns in the training set, with each node storing a pattern along with its class label.

By leveraging the hierarchical tree-based organization of the stored patterns and its search scheme based on relative distance, the algorithm (i.e., our system) efficiently balances high pattern-matching accuracy with a low number of visited nodes during the search phase.

In the following sections, we describe in detail the Portegys algorithm [27] to ensure the self-containment of this paper. We outline the microscopic rules governing the tree-formation protocol in Sec. II B 1 and the search phase in Sec. II B 2.

However, we emphasize that the primary focus of this paper is not on the microscopic rules of the algorithm itself. We are rather focused on the dataset-dependent emergence of functional topological configurations, i.e., those enabling efficient pattern matching, as macroscopic outcomes of maximizing the efficiency functional.

1. Network growth process

Given the value of the single control parameter r and a collection of high-dimensional patterns drawn from the dataset (i.e., the training data), the system creates a topological representation that captures the proximity relationships between patterns in the original high-dimensional space.

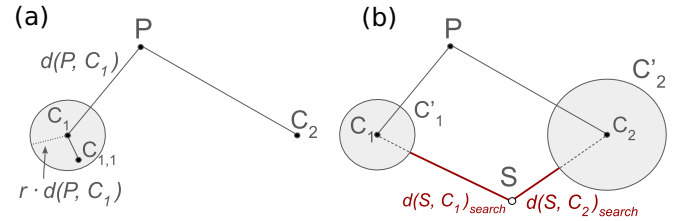


FIG. 2. Tree formation and search method. In (a), the insertion of a node into the tree is illustrated. At each tree-growth step, a new pattern is copied into a new node. According to the insertion protocol, this node is either linked to the current parent node P , or passed to one of its children C_i for link evaluation. The parameter r modulates the region of attraction of the child's subtree, indicated by the gray circle around C_1 , relative to that of the parent node. In (b), the search method based on relative distance is illustrated. When a test pattern arrives at node P , it must decide whether to search the subtree of C_1 or that of C_2 . The choice is based on minimizing the distance to the expected patterns contained within the subtree of C_i . In the illustrated case, patterns closer to the search pattern S are expected to be stored in the subtree of C_2 . Based on [62].

In this sense, the algorithm maps information content about the spatial correlations in the sample space into a network using the nearest-neighbors approach based on a relative distance parametrized by r . The emerging networks are loopless and sparse, i.e., $|E| = N - 1$, where $|E|$ is the number of edges, and N is the number of nodes corresponding to the number of patterns in the training set.

The tree's growth process begins by randomly sampling without replacement the first pattern from the dataset. The pattern is copied to the first node, which is labeled as the root node of the tree and designated as the *current* parent node P .

Subsequently, at each tree-growth step, a sample is randomly picked from the training set and copied into a new node. Thus, a pattern-bearing node is inserted into the tree through a recursive insertion process that starts at the root node and tries to locate each new pattern-bearing node as deep as possible in the forming hierarchical structure.

Insertion process. At each step of the insertion process, a decision is made: Either the new node is linked directly to the current parent node P , or it is passed to the parent's first child that satisfies a proximity condition. Specifically, the new pattern-bearing node is passed (though not necessarily linked yet) to an already existing child of parent P only if its pattern lies within a radius centered on the child's pattern and equal to the Euclidean distance between the parent's pattern and the child's pattern multiplied by r . For instance, in Fig. 2(a), the new node $C_{1,1}$ is passed to the child C_1 of the current parent node P because it satisfies the proximity criterion,

$$d(C_1, C_{1,1}) < r \cdot d(P, C_{1,1}), \quad (7)$$

where $d(\cdot)$ is the Euclidean distance. If the pattern-bearing node is passed to the child, then that child becomes the new parent node, and the algorithm will repeat the decision process, evaluating the children of the current parent, trying to locate the node as deep as possible. On the contrary, if the new node is linked to the current parent node, then the insertion stops, and the growth process starts again: picking a new pattern from the training set, copying it into a new node,

and restarting the insertion process from the root node to link the new node to one of the nodes in the growing tree.

We highlight that the order in which patterns are presented influences the formation process, as each new node is always inserted into the first branch that satisfies the proximity condition outlined above, even if multiple branches meet the criterion. Therefore, in Sec. III, all results are averaged over multiple disorder realizations, each generated by randomly permuting the order in which patterns from the dataset are presented to the growing tree.

To avoid the excessive accumulation of children on a single node, the tree undergoes a reorganization each time a new pattern-bearing node is linked to the tree. When a new child creates a bond with a parent node, all preexisting children of that parent are reevaluated to assess whether they should be linked to the newly added child instead, in accordance with the pattern each one is storing. If this condition is met, the entire subtree rooted at the preexisting child is detached, and its pattern-bearing nodes are reinserted, starting the insertion process from the current parent. Remarkably, those nodes are not necessarily linked to the new child, as they may not satisfy the proximity condition relative to it.

The growth process is expected to produce subtrees of a given node that have, on average, decreasing distances between their patterns for decreasing hierarchical levels.

The parameter r . As it can be intuitively understood from the observation of Figs. 1 and 5(b), the control parameter r governs both the branching of the tree and the number of hierarchical levels, i.e., the elongation of the network.

Nevertheless, we highlight here a key point that will be discussed extensively throughout the paper. The parameter r acts as an effective distance cutoff in assembling the pattern-bearing nodes into a tree. It captures correlations between patterns up to a certain scale in the original sample space and thereby scans the geometric organization of the data.

We emphasize that, while the diffusion time τ resolves the geometry of the network (as discussed in Sec. II A 1), the parameter r resolves the geometry of the dataset. In other words, it translates correlations between patterns in the original sample space into topological correlations among the nodes of the network. At the end of this paper, by maximizing the efficiency with respect to r , we will infer the spatial (r -)scales that best resolve the intrinsic geometry of each dataset. We will also show that the network spectral features of these maximally efficient structures reflect the underlying complexity of the dataset.

2. Search phase

After the tree formation described in Sec. II B 1, the Portegys algorithm enters the search phase, during which it attempts to match an unseen test pattern to one of those previously stored in the tree's nodes to infer the class of the test pattern. This is performed using a depth-first search strategy: at each step, the algorithm explores the subtree expected to be closest to the search pattern.

We outline the process. The search phase starts at the root node, which is labeled as the current parent pattern P , similarly to what was done in the formation process. If the parent has multiple children, the algorithm must select the

most promising child's subtree, denoted C'_i , to explore next. To make the choice, it computes the *search distance* between the test pattern S , and all patterns stored on each child C_i , relative to the parent pattern P ,

$$d(S, C_i)_{\text{search}} = d(S, C_i) - r \cdot d(P, C_i), \quad (8)$$

where $d(\cdot)$ is the Euclidean distance. If $d(S, C_i)_{\text{search}} < 0$, it is set to zero. Note that with a little abuse of notation, we used the same symbol C_i to indicate the child node and the pattern it stores. The same is true for the parent P .

The search distance $d(S, C_i)_{\text{search}}$ quantifies the distance between the S -pattern and the patterns expected in the subtree of the i -child of P . These expected patterns are collectively represented by the gray circle C'_i , in Fig. 2(b). The red-bold segments indicate the search distance between the test pattern S and the pattern stored in child C_i of parent P .

The child whose subtree is estimated to store the patterns most similar to the test pattern, i.e., the node with the least $d(S, C_i)_{\text{search}}$, is then selected. This child becomes the new current parent node, and the search continues recursively within its subtree until the lowest hierarchical level of the tree is reached.

At each hierarchical level, while evaluating the children of a parent, a list of the class labels of the top $k(=5)$ best-matching node's patterns is updated [28]. Thus, throughout the search phase, the k best matches are maintained in memory, and upon completion of the tree traversal, a majority vote among these candidates is used to determine the final matching class.

Conversely, if the parent node has only a single child, that child automatically becomes the next parent node without any distance computation required. In this case, the list of k -nearest neighbors is not updated. This scenario typically arises when the formation protocol generates elongated structures, as illustrated in Fig. 1. When one-dimensional chains are produced, no classification accuracy is computed.

C. Datasets

The datasets provide the disordered environment, i.e., a disordered collection of samples, to which the algorithmic system adapts during the tree-formation process.

Unless otherwise stated, in all our experiments a subset of $N = 2000$ patterns is randomly collected from the training set and converted into a network of N nodes and $m = N - 1$ edges. A subset of $M = 2000$ patterns from the test set is used to evaluate the functional efficiency of the algorithm, based on its ability to perform classification.

Our results are evaluated on four different datasets: the NIST [63], the MNIST [64], the FashionMNIST [65], and the CIFAR10 [66]. Each dataset presents a distinct level of complexity (see the ranking in Sec. III A), influencing both the tree-formation process and the functional degree of the trees relative to the pattern-matching task.

The MNIST database contains 24×24 gray-scale images of handwritten digits [0–9] commonly used to benchmark machine- and deep-learning algorithms. The NIST dataset is the ancestor of the MNIST, it contains 16×16 black-and-white images of the handwritten digits [0–9]. The FashionMNIST is a large dataset of 28×28 gray-scale im-

ages of fashion products from 10 categories. The CIFAR-10 dataset contains 32×32 color images belonging to 10 classes. It should be noted that in this type of data, a pattern class can generally not be represented by a single centroid vector, but will contain multiple densities (“sub-style patterns”) such that the sample point cloud can be ordered naturally in a network structure.

Samples from the FashionMNIST and MNIST are always binarized. The samples from the CIFAR10 are normalized between 0 and 1, dividing each pixel value by 255, and values are rounded to the eighth decimal.

In all cases, images are unrolled into vectors of size equal to the number of pixels, and the 10 classes are equally represented in both the training and test sets. In the Supplemental Material (SM) [67], we report results for larger networks generated using $N = 4000$ and tested with $M = 4000$ samples from the MNIST and FashionMNIST datasets.

III. RESULTS

We adopt the following strategy: In Sec. III A, we quantify the complexity of the environmental conditions faced by the system. Next, in Sec. III B, we quantify the system’s degree of function, defining a pattern-matching efficiency. In Sec. III C, we characterize the algorithm’s phase plane, revealing the presence of phase transitions and the topological transformations that occur when they are crossed. In Sec. III D, we examine the system’s thermodynamic and pattern-matching efficiency under the conditions defined by each complexity level. In Sec. III E, we then discuss efficiency in relation to scale-invariant networks. Finally, in Secs. III F and III G we show that the complexity of the environment conditioning the tree formation process can be inferred from the topological traits of the networks that maximize functional efficiency.

A. Dataset complexity

Before studying the tree-formation process of our system and assessing its functional capabilities in relation to environmental complexity, we first need to quantify and rank the complexity of the environment.

To this end, we estimate the complexity of each dataset using the same subset of $N = 2000$ training samples used throughout the paper. We employ multiple methods with the purpose of cross validating the ranking, which is shown in Fig. 3.

The first heuristic approach is based on the test accuracy of a small multilayer perceptron (MLP) (2 layers with 128 units each). In pattern recognition, the complexity of a dataset is determined by the complexity of the curvilinear hyperplanes that would be needed to optimally separate patterns from different classes. A simple but effective parameter to gauge this facet of complexity is the accuracy that a classifier (such as the MLP) can achieve in determining the true class label for a sample. Therefore, a more complex dataset will typically result in lower accuracy for neural networks of fixed size, compared to a less complex dataset. For this method, we also used the $M = 2000$ test samples from each test set.

The second method we use is the computation of the Lempel-Ziv complexity [45]. Lempel-Ziv complexity is de-

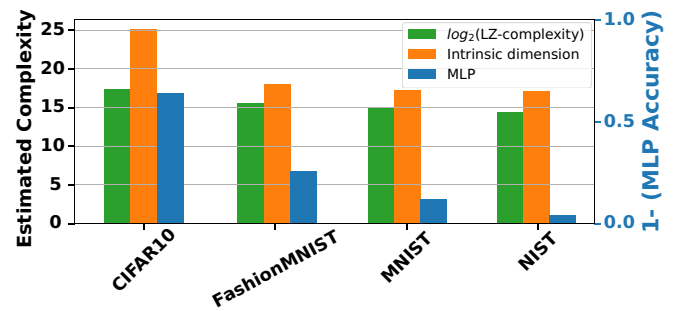


FIG. 3. Dataset complexity is estimated with multiple methods for each training dataset of $N = 2000$ samples encountered in this work. In green, an estimate is given computing the Lempel-Ziv complexity [45]. In orange, the intrinsic dimension is estimated with the maximum likelihood estimation method [46–48]. In blue, the accuracy on the test set for a multilayer perceptron (MLP) with two hidden layers of 128 units each is computed using a test set of $M = 2000$ samples. The left axis refers to the quantities in green and orange, while the right axis refers to the MLP accuracy. The CIFAR10 dataset is estimated as the most complex, while the NIST dataset is the least complex among the four.

fined as the number of different binary substrings encountered when viewing a data vector from beginning to end. Thus, for example, a dataset where some substrings are frequently encountered results in a lower complexity compared to one where substrings are more diverse. To measure this quantity for each dataset, we binarized and concatenated the whole set of training samples before feeding it to the algorithm computing the Lempel-Ziv complexity.

The third method concerns the measure of the dataset’s intrinsic dimension [43,48]. Generally speaking, a given dataset of gray-scale images *a priori* lies on a high-dimensional space $\mathbb{R}^p$, where p is the number of pixels. Nevertheless, because it might contain a specific class of objects, e.g., handwritten digits, it reasonably spans a lower-dimensional manifold than the one defined through the number of pixels [43,48]. Although the definition of such a manifold remains an open problem, its dimension can be estimated based on the distances between the n -nearest neighbors. It is expected that datasets with higher complexity require a higher-dimensional manifold to represent them adequately.

We use the maximum likelihood estimation (MLE) method introduced in Ref. [46] with the correction proposed in Ref. [47]. The method is derived by applying the maximum likelihood principle to distances between samples in a close neighborhood, and it is based on the assumption that their distribution can be locally approximated with a homogeneous Poisson process.

A detailed description of the MLE method is out of the scope of this paper, but the interested reader is referred to the original works [46,47] or to Sec. S9 within the SM [67]. Here, we provide the essence of the MLE intrinsic dimension estimate. The MLE estimate measures, for each sample x in the dataset, the average (logarithmic) ratio $\log(d(x, x^{(j)})/R)$ over all $n - 1$ samples $x^{(j)}$ lying within the hyperdimensional sphere of radius R centered on x . The radius of the sphere is defined by the distance to the n th neighbor from x . $d(\cdot)$

is the Euclidean distance. This provides an estimate of the *local* dimension per sample. The *global* intrinsic dimension is estimated by repeating the process for each sample in the dataset and taking the overall average. Indeed, the choice of n affects the estimate, as the same authors note [46]; in fact, it is possible for a dataset to have different intrinsic dimensions at different scales and, for example, an average over multiple values of n could be taken. Nevertheless, the authors prescribe a small radius of the sphere, so we set $n = 3$. We restrict ourselves to this value as we are not interested in the actual value of the intrinsic dimension, but rather in the ranking of the datasets.

As expected, the three methods yield the same ranking, as shown in Fig. 3, thereby validating the robustness of the results.

B. Pattern-matching efficiency

We define a measure to quantify the degree of function of the system by defining the task-related pattern-matching efficiency.

Mapping a high-dimensional collection of patterns into a hierarchical tree structure limits the number of comparisons needed to match a novel pattern to a prototypical example. Namely, the algorithm introduced by Portegys [27,28] is equipped with a pattern-searching method illustrated in Fig. 2 and described in Sec. II B 2, representing the functional aspect of the system. The class match accuracy A is defined as the number of correctly classified test patterns over the total number of test patterns presented during the search phase (Sec. II B 2). We define a pattern-matching efficiency, θ , in terms of the class match accuracy A and the number of distances computed N_{dist} to perform the task,

$$\theta = \frac{A}{\log N_{\text{dist}}}, \quad (9)$$

the logarithm is taken since the range of N_{dist} can virtually span the interval $N_{\text{dist}} \in [0, N \times M]$, where N is the number of nodes in the tree, which is equivalent to the number of samples in the training set, and M is the number of samples in the test set. In the case of a star graph, corresponding to $r \rightarrow 0$, the number of computed distances is maximum, $N_{\text{dist}} = N \times M$. As r increases, the number of computed distances decreases, potentially reducing the class-match accuracy. Nevertheless, as shown in Fig. 4(d), the CIFAR10 dataset exhibits a global accuracy peak at an intermediate value of r . Thus, for specific datasets (see also Sec. S7 with the SM [67] the case of the FashionMNIST for $N = 2000, 4000$), a certain level of semantically meaningful hierarchical organization is more effective at encoding the dataset's geometrical information than directly comparing the test patterns to all stored ones.

When defining the pattern-matching efficiency, we are making the simplified assumption that the pattern-matching algorithm runs on a physical substrate such that accessing any node has a fixed cost in terms of energy or in terms of general resources. Hence, the efficiency θ quantifies the trade-off between the cost of used resources and the gain in performance, as depicted in Figs. 4(c) and 4(f) for the MNIST and CIFAR10, respectively.

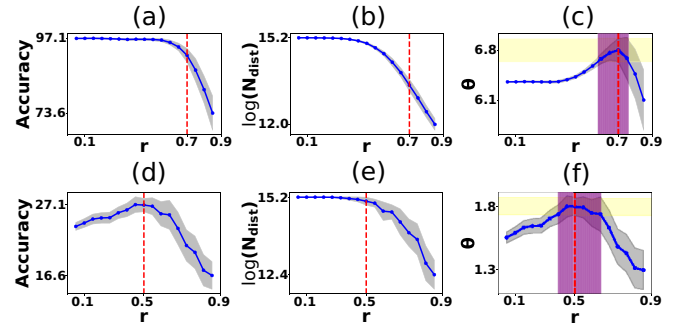


FIG. 4. Pattern-matching efficiency in (c) and (f) measured over the NIST and CIFAR10, respectively. The efficiency θ is defined as the trade-off between the classification accuracy [(a),(d)], and the logarithm of the number of distances computed to find the best match [(b), (e)]. The dashed-red vertical line indicates the optimal r value in terms of efficiency θ . The shaded gray area designates the standard deviation. The purple-shaded area in (c) and (f) indicate a tolerance r -domain of optimal efficiency. The tolerance domain is defined as the r -values such that the corresponding θ values are included in one standard deviation computed at the maximum θ , whose y-axis projection is indicated by the yellow-shaded area. Curves are computed over 50 independent simulations with a subset of 2000 samples extracted from the NIST and CIFAR10 datasets for both the training and testing. The 10-digit classes are equally represented both in the train and test sets.

While in Sec. III C we discuss the full range of r , in the rest of the paper, we limit the discussion to $r \leq 0.85$ as the accuracy, and hence the pattern-matching efficiency, quickly drops for larger r , moreover, the algorithm shows instability as it does not necessarily converge for $r \geq 0.9$ and $N, M \geq 2000$.

C. Phase portrait of the system

The Portegys algorithm produces three prototypical structures as those depicted in Figs. 1 and 5(b), depending on the value of the control parameter r (see in Methods, Sec. II B).

For $r \rightarrow 0$, every node is linked to the root node and a star graph results (see in Fig. 1 the central node, which is the first node selected in the growth process). For some intermediate value of r , a ragged tree structure with progressively smaller distances between the patterns is produced; while for $r \rightarrow 2$, a 1D lattice is created.

Two structural phase transitions involving two different network resolution scales can be identified [Fig. 5(a)]. The transition at the largest resolution scale can be gauged using as an order parameter a quantity that describes the large-scale network properties: $\hat{d} = d/m$, where d is the diameter of the tree and m is the number of edges equal to $m = N - 1$. For hierarchical trees, another choice would be the normalized number of hierarchical levels $\hat{h} = h/(N - 1)$, where h is the number of hierarchical levels ranging from $h \in [0, N - 1]$ [see illustration in Fig. 6(b)]. Both quantities describe the network “elongation” controlled by the parameter r . We chose the normalized diameter $\hat{d}$, and we identify this phase transition at $r = 1.05$ as indicated by the peak of the variance (dotted lines) in Fig. 6(b) for the two network sizes $N = 500, 1000$. This phase transition border distinguishes

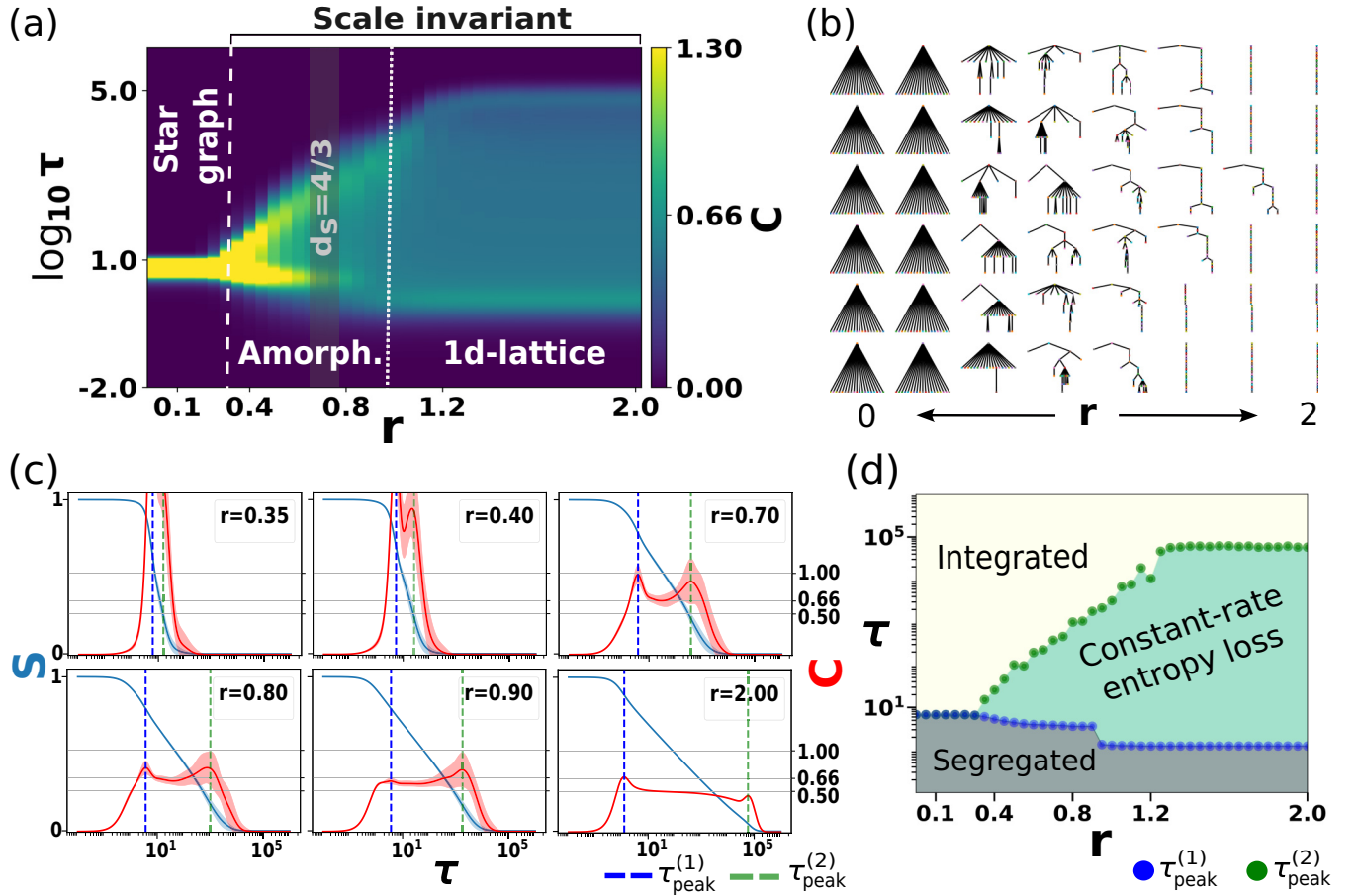


FIG. 5. Multiple phase transitions in hierarchical tree networks. In (a) the specific heat C across the range $r \in [0, 2]$. At $r = 0.35$, indicated by the dashed white line, the hierarchical network transitions from the star graph to the amorphous configuration with the emergence of scale-invariant traits evidenced by a plateau in the specific heat. The scale-invariant region shows a second phase transition from amorphous to regular 1D lattice at $r = 1.05$ indicated by the dotted-white line. For $r > 0.35$, the network is always scale invariant with decreasing spectral dimension up to the transition point, above which the spectral dimension saturates to $d_s = 1 = 2 \times C_0$, where C_0 is the plateau height in (c). Note that at roughly $r = 0.7$ the spectral dimension is $d_s = 4/3$ as expected for random trees with finite variance in the node degree. In (c) the entropy (blue) and the specific heat (red) for $r = \{0.35, 0.40, 0.70, 0.80, 0.90, 2.00\}$. Vertical-dashed blue and green lines indicate the τ -scale of the first and second peaks $\tau_{\text{peak}}^{(1,2)}$ of the specific heat C , respectively (see legend). In (d), the τ -scales of the two specific heat peaks are illustrated with blue and green dots respectively delimiting the three dynamical phases: the segregated phase (gray), the integrated phase (light yellow) and the green region where *self-similar structures* are integrated at a constant rate across the network resolution scales. In (b) exemplary hierarchical trees of $N = 20$ for growing r from left to right. All curves in this figure are realized for networks of $N = 500$ randomly sampled from the FashionMNIST dataset and averaged over 50 independent realizations.

between networks in the amorphous regime from those with a regular 1D lattice structure. As it is expected for a 1D lattice, in Figs. 5(a) and 5(c), when $r > 1$, the spectral dimension is $d_s = 1$. On the left-hand side of the transition point, $d_s = 4/3 = 2C_0$ at roughly $r = 0.7$ [Figs. 5(a) and 5(c)], typical of random trees with finite variance in the node degree's distribution [68]. C_0 is the specific heat plateau height. See Sec. S1 within the SM [67] for the power-law density of the Laplacian spectrum whose exponent γ is related to $d_s = 2(\gamma + 1)$.

The second change of phase is the transition from the star graph to the amorphous configuration at $r = 0.35$. Here, the emergence of scale-invariant structural properties is reflected by the progressively larger specific heat plateau in between the two C peaks, as shown in Figs. 5(a) and 5(c).

Before the transition, a single peak in the specific heat identifies a single trivial τ -resolution scale below which the informationally connected components are the single nodes,

and after which the informationally connected component is the whole graph. After the phase transition, a second peak of the specific heat emerges at larger τ scales. This second peak of C in Fig. 5(c) identifies the scale at which the information is integrated between the distinct modules. In other words, the second peak reflects the “degree” of the large-scale modules, in much analogy to the first peak, which reflects the presence of hubs at the local connectivity scale of the nodes.

The star graph and amorphous network phase space regions are those investigated in this work because the peaks in the thermodynamic and pattern-matching efficiencies always occur in the amorphous and scale-invariant region of the phase plane. Conversely to the amorphous to regular-lattice transition discussed above, the phase transition separating the star graph and amorphous network phase space regions cannot be identified by $\hat{d}$, which is a quantity resolving the large-scale properties of the network. Indeed, we observe in Fig. 6(b)

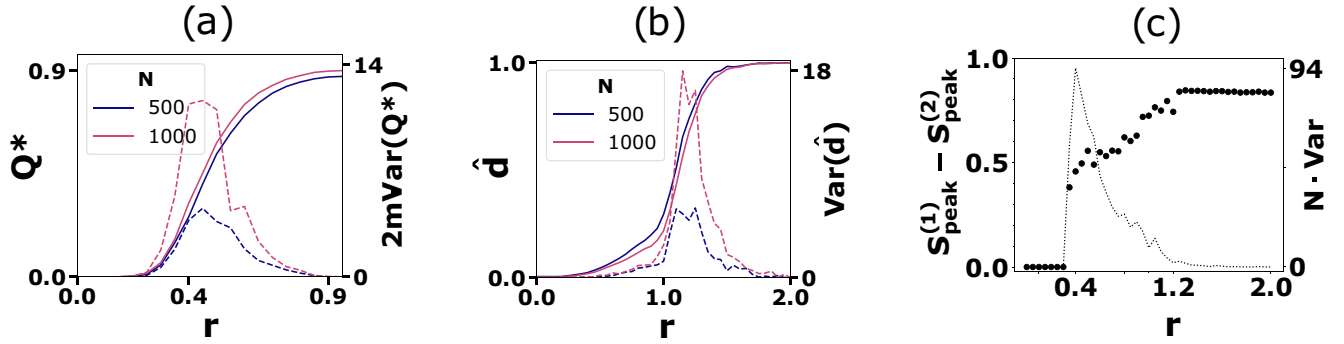


FIG. 6. Order parameters. In (a), the modularity Q^* resolves the transition involving the mesoscale structure. In (b), the normalized diameter, $\hat{d} = d/m$, resolves the large-scale transition. Their respective variances (dashed lines) identify the transition points. In (c), the difference of entropy at the two $\tau_{\text{peak}}^{(1,2)}$ peaks quantifies the normalized (log) number of informationally connected nodes across the scale-invariant τ -domain, i.e. the number of nodes included in the scale-invariant region. Note the discontinuity at the transition from star to amorphous. Curves in (a) and (b) are realized for networks of $N = 500, 1000$ randomly sampled from the FashionMNIST dataset and averaged over 50 independent realizations. Curves in (c) are for networks of $N = 500$.

the absence of a peak in the variance of $\hat{d}$ for $r < 1$, which would suggest no indication of a *bona fide* phase transition in this r -range, if one were to rely solely on $\hat{d}$. Nevertheless, a phase transition in that r -range exists, and an appropriate order parameter needs to be employed to identify it.

This phase transition occurs at smaller resolution τ -scales compared to the first transition we previously discussed, that is, the transition from the amorphous to the regular 1D lattice phase. Otherwise, it is reasonable to expect that it would have already been detected via $\hat{d}$. Specifically, the transition involves the mesoscale properties changing from the star graph to the scale-invariant regime [see the bifurcation of the specific heat in Figs. 5(a) and 5(c)]. We can identify this transition using as an order parameter the modularity [69] Q^* of the tree, which we show in Fig. 6(a). Note that we use the symbol $*$ to distinguish it from the notation used for the network heat. Modularity is defined as

$$Q^* = \frac{1}{2m} \sum_{i,j} \left(A_{ij} - \frac{k_i k_j}{2m} \delta(c_i, c_j) \right),$$

where m is the total number of edges, A_{ij} is the element of the adjacency matrix, k_i and k_j are the degrees of nodes i and j , respectively, c_i and c_j are the community assignments of nodes i and j , and $\delta(x, y)$ is the Kronecker delta.

We are taking advantage of the fact that the modularity for trees is bounded between $[0, 1]$ and that in the case of a 1D lattice, the modularity is equal to 1 if we partition the network into N communities, which is the case if we use the community detection algorithm introduced in Ref. [70]. We stress that modularity is a mesoscale descriptor, while the diameter of the network is a large-scale descriptor. As the large-scale descriptor [Fig. 6(b)] fails in detecting the phase transition from star graph to scale-invariant, we decide to monitor the mesoscale resolution to detect the structural changes occurring from the star graph phase to the amorphous scale-invariant phase. This motivates the choice of the τ -mesoscale employed in Sec. III.

As a second possible order parameter, for the mesoscale phase-transition, we show in Fig. 6(c) the difference of entropy at the two $\tau_{\text{peak}}^{(1,2)}$ specific heat peaks. This order

parameter counts the normalized (log) number of informationally connected nodes across the τ -scales of the specific heat plateau. In simple terms, we count the fraction of nodes included in the scale-invariant region. We observe a discontinuity at the border of the scale-invariant transition that is not observed at the border of the amorphous-to-regular-lattice transition.

As already discussed in Sec. II A 1, also along the τ -direction an informational phase transition exists [34,35] delimited by the peaks of C at $\tau_{\text{peak}}^{(1,2)}$. Using the specific heat peaks, we can distinguish three *dynamical* phases [Fig. 5(d)]. For $\tau < \tau_{\text{peak}}^{(1)}$, the network is in the informationally segregated phase, where information is locally segregated in small-scale structures (refer to Sec. II A 1); while for $\tau > \tau_{\text{peak}}^{(2)}$ is in the informationally integrated phase, where information has diffused all over the network. Only for $r > 0.35$ and $\tau_{\text{peak}}^{(1)} < \tau < \tau_{\text{peak}}^{(2)}$, we observe the *dynamical* regime of *constant-rate entropy loss* characteristic of scale-invariant structures with dilation symmetry [35,36].

D. Efficiency

In general, a complex system emerges from microscopic rules, such as those outlined in Sec. II B, which are blind to the macroscopic topological features that arise. However, the same process might also be described in macroscopic terms through a variational principle governing structural formation [1].

In this section, we evaluate the outcome of the system's formation process with η , and its resulting performance using θ , across different environments, each characterized by a distinct level of complexity. Moreover, leveraging the principle of maximal thermodynamic-like efficiency, we can narrow down the r -region of higher functional (pattern-matching) efficiency prior to the test phase.

In Fig. 7, panels (a)–(d) depict the thermodynamic-like efficiency η , and panels (e)–(h) depict the pattern-matching efficiency θ . See Eqs. (6) and (9), respectively.

The thermodynamic efficiency $\eta = (Q + W)/W$ quantifies the trade-off between two quantities. The former, $Q =$

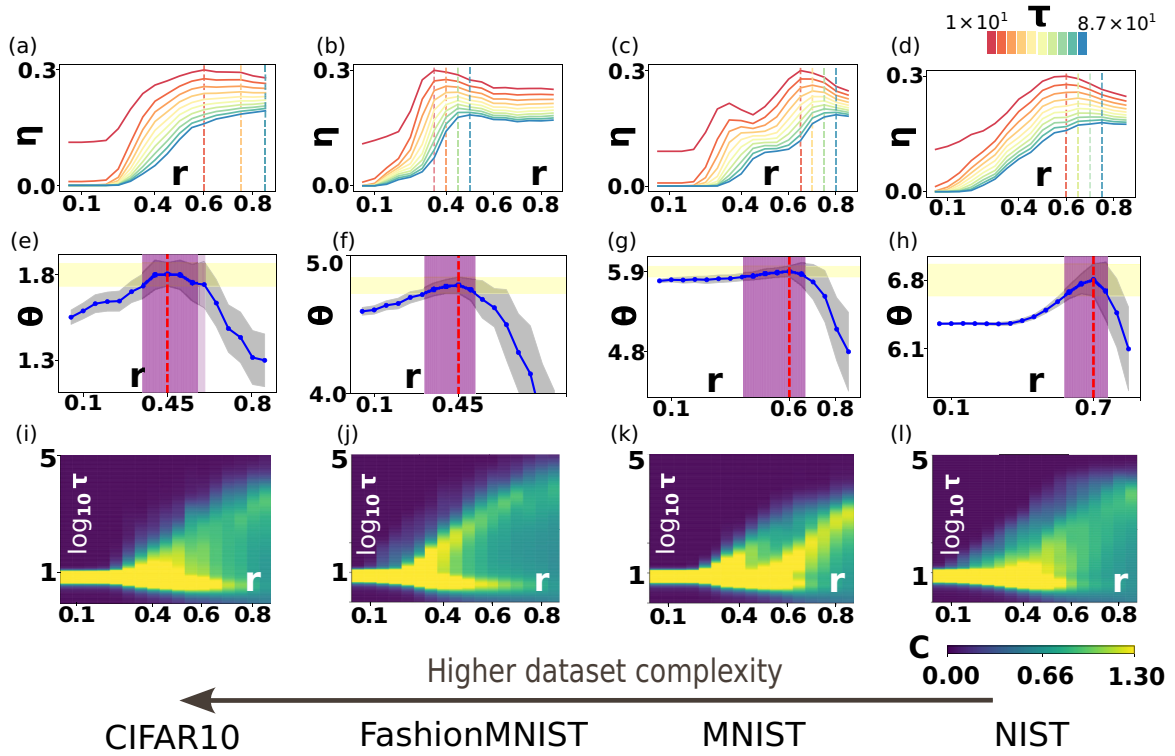


FIG. 7. Thermodynamic and pattern-matching efficiency. In (a)–(d), the thermodynamic efficiency η as a function of r . The thermodynamic efficiency is measured in the τ -range of one order of magnitude from the bifurcation of the specific heat C depicted in panels (i)–(l). The τ -resolution scale $\tau \sim 10$ identifies the relevant network scale resolving the transition from the single-peak to the emergence of a roughly plateau region, $C = C_0 > 0$, reflecting a roughly scale-invariant structure. The symbol $\sim$ is used to denote values of τ close to the reference value $\tau = 10$, specifically within the range $\tau \in [1, 9) \times 10^1$, and to exclude values outside this interval, as indicated by the colorbar at the top right. In (e)–(h), the pattern-matching efficiency θ is plotted as a function of r for the four investigated datasets of increasing complexity (see arrow at the bottom of the figure). The shaded purple area indicates the tolerance r -range for the optimal pattern-matching efficiency, which includes points in one standard deviation from the global maximum indicated by the vertical red-dashed line. In panels (b), (f) and (d), (h) a noticeable cohesion of the peaks of the two efficiencies is observed. Vertical-dashed lines indicate the global maxima of the efficiencies. The efficiency peaks, both η and θ , are always in the amorphous region of the phase plane. For datasets of higher complexity, they tend to be closer to the structural phase transition border separating the star graph from the amorphous region. Quantities are averaged over 50 independent simulations.

$\delta S/\tau$, is proportional to the variation of entropy content, $\delta S = S - S_{\text{iso}}$, from the case of isolated nodes (see Methods). It estimates the ability to express response diversity to a stochastic perturbation, or in other terms, the internal representation capabilities provided by the network topological traits. We refer to Fig. 1(c) in Ref. [1] for an illustrative visualization. In particular, note the case of a fully connected network (last column on the right): the system maps every perturbation to the same state, namely with all nodes turning red, resulting in a *many-to-one* mapping. Such a mapping hinders distinguishing the original perturbation based on the system's response, resulting in a network with reduced representational power.

The latter quantity involved in the equation of η is the variation of information flow with respect to a network of isolated nodes, $W = \delta F$, where $\delta F = F - F_{\text{iso}}$; so to speak, how fast information travels across the network (see Methods).

The pattern-matching efficiency θ , defined in Eq. (9), quantifies the trade-off between accuracy and number of visited nodes in the pattern-matching task, representing the functional aspect of the algorithm (see Methods).

The vertical-dashed lines indicate the peaks of the efficiencies, while the shaded purple regions in panels (e)–(h) indicate an r -domain tolerance for θ . The tolerance domain is defined as those r -values such that the corresponding θ values are included in one standard deviation computed at the best trade-off.

The two efficiencies θ and η show some consistency in the x -range of highly efficient structures, see panels (b) and (d) related to the NIST and the FashionMNIST. If we include points in the tolerance r -range spanned by the purple-shaded areas in panels (e)–(h), we observe the same consistency for all datasets when η is computed at $\tau = 10$ [red-dashed lines in panels (a)–(d)]. We discuss this matter in Sec. IV. For now, we note that η serves as a practical proxy for θ owing to the observed empirical overlap between the maxima of the two efficiencies, combined with the fact that η requires only the network structures resulting from the growth phase (Sec. II B 1) and does not involve executing the search phase (Sec. II B 2). Identifying regions of the phase space where η is high enables reducing the exploration of the r -range in which the maximum pattern-matching efficiency is expected to lie.

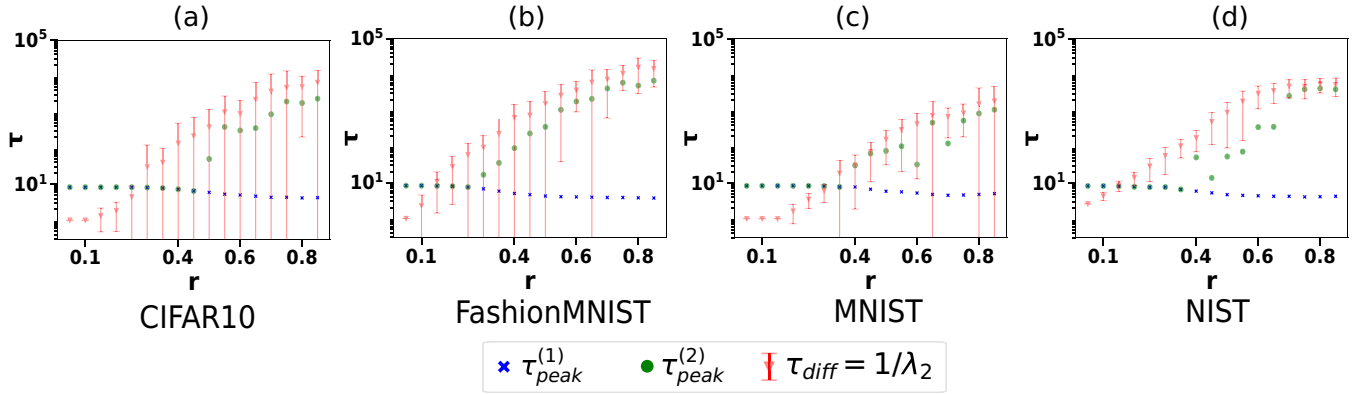


FIG. 8. Role of the ensemble average on the estimation of the τ -scale. Panels (a)–(d) depict the $\tau_{\text{peak}}^{(1,2)}$ and τ_{diff} as a function of r . $\tau_{\text{peak}}^{(1,2)}$ are the τ -location of the first and last peak of the specific heat in Fig. 7. τ_{diff} is the inverse of the Fiedler eigenvalue which provides a good estimation of the end of the diffusion process only for sufficiently large r . Note that the red curves retrace the profile of the upper branch of the specific heat for $\tau \geq 10$ in panels (i)–(l) of Fig. 7. At smaller r values, the equilibrium time is better reflected by $\tau_{\text{peak}}^{(2)}$. The equilibrium of the diffusion process is an increasing function of r . τ_{diff} curves are averaged over 50 independent simulations. The $\tau_{\text{peak}}^{(1,2)}$ curves are computed from the specific heat curves, which are averaged over 50 independent simulations.

A relevant aspect we still need to discuss is the network resolution scale at which we compute the η efficiency. In seminal work in Ref. [1], it is suggested to use as a reference the τ -scale equal to the inverse of the Fiedler eigenvalue, i.e., the smallest nonzero Laplacian eigenvalue $\lambda_2 = 1/\tau_{\text{diff}}$, and divide τ_{diff} by increasing powers of 10 to define the large propagation scales, $\tau_{\text{large}} \in (\tau_{\text{diff}}/10^1, \tau_{\text{diff}})$, the middle scales, $\tau_{\text{mid}} \in (\tau_{\text{diff}}/10^2, \tau_{\text{diff}}/10^1)$, and the small scales, $\tau_{\text{small}} \in (\tau_{\text{diff}}/10^3, \tau_{\text{diff}}/10^2)$, where to compute η . The Fiedler eigenvalue provides the largest network time scale where the diffusion process has reached equilibrium. Nevertheless, τ_{diff} seems to only approximate the end of the diffusion process, which is better reflected by $\tau_{\text{peak}}^{(2)}$, especially in the star-graph region for small r for highly degenerate diffusion states (see Sec. S1 within the SM [67]). Panels (a)–(d) of Fig. 8 depict the τ_{diff} curves in red, next to the $\tau_{\text{peak}}^{(1,2)}$ curves. The key difference is the thermal ensemble average: a weighted average that employs all the eigenvalues to compute the specific heat and identify its peaks.

Moreover, note in Fig. 8 the increasing “diffusion volume” of the network indicated by the $\tau_{\text{peak}}^{(2)}$ curve as a function of r as a consequence of the increasing diffusion times. We remark that $\tau_{\text{peak}}^{(2)}$ is a function of solely the parameter r . The latter shapes solely the degree distribution, as the number of nodes and edges is fixed across the phase plane and fixed in the regime of sparsity, $N/(m-1)$, where N and m are the number of nodes and edges respectively. This observation makes a connection between the role of the degree distribution and the “volume” of the network.

Concluding, differently from what was proposed in Ref. [1], we do not use τ_{diff} , or $\tau_{\text{peak}}^{(2)}$ from Fig. 8. In fact, that would mean adopting the largest possible resolution, which does not monitor the structural mesoscale changes involved in the transition from the single trivial scale of the star graph to the more complex topological traits of the amorphous scale-invariant networks.

The resolution scale $\tau \sim 10$ at which we compute η is the scale involved in the mesoscale structural transition from the star graph to amorphous scale-invariant, as we discussed in

Sec. III C, Fig. 6. The use of the symbol $\sim$ is specified in the caption of Fig. 7. Such critical resolution scale is identified by the bifurcation of the specific heat peaks in Figs. 7(i)–7(l) and in Figs. 8(a)–8(d).

The resolution scale, $\tau \sim 10$, is shared across all topologies in the phase plane and, remarkably, defines an “ultraviolet” cut-off in the characteristic frequency of the networks across the transformation from one phase to the other. This is analogous to integrating out the large eigenvalues responsible for the finer details, as discussed in the context of the Laplacian-renormalization group [35]. Note in Sec. S1 within the SM [67] that our choice of $\tau = 1/\lambda \sim 10$ is supported by the exclusion of the resolution region where the power-law behavior of the spectral density, $P(\lambda)$, is lost. In Sec. S3 within the SM [67], the efficiency η is computed in the range interval $\tau < 10$. Whether computed at those scales, η shows a change in concavity and no relation with the pattern-matching efficiency.

Moreover, we observe in Fig. 8 that, while the value of r where a structural transition appears is dependent on the dataset, i.e., $r = 0.40$ for the NIST, $r = 0.40$ for the MNIST, $r = 0.30$ for the FashionMNIST, and $r = 0.50$ for the CIFAR10, the $\tau \sim 10^1$ scale is the same for all the investigated datasets. Finally, we note that the bifurcation τ -scale is also unaffected by the number of nodes in the network (see Sec. S2 within the SM [67]).

E. Efficient performance at the emergence of scale-invariance

In this section, we discuss the connection between the system’s efficient performance and the emergence of scale-invariant properties.

All optimal networks are found in the amorphous region of the phase plane where dilation symmetry is observed in the spectral density and reflected in the specific heat plateau. To discuss this observation, we refer to the following argument [11,71,72]: A system needs tuning, or self-tuning, at the edge of a phase transition when the task demands a high level of

complexity. Specifically, this was investigated in the context of second-order phase transitions [11,71]. But what is the advantage of being close to a second-order phase transition? One potential advantage arises from the power-law behavior of the density function for certain quantities, such as the density of cluster sizes in percolation or the density of avalanche sizes and durations in systems with temporal dynamics. The power law offers the system a wide range of availability in the density of the quantity of interest, allowing the system to exploit this as an informational resource for tasks, whether adapting to a complex environment or pursuing other goal-oriented objectives.

In our case, the amorphous region of the phase plane, where the network traits display self-similar structures in a limited τ -resolution range, exhibits power-law spectral density governing the network frequencies for $\lambda \rightarrow 0$ (see Sec. S1 within SM [67]). Thus, better performances in terms of efficiency, particularly pattern-matching efficiency, can be explained in terms of network frequencies, which provide effective resources for the real-world disorder structuring task and the subsequent pattern-matching task.

F. Datasets' resolution scale

In this section and the next one, using a complementary perspective to the one outlined in the previous sections, we frame a variational principle based on efficiency in terms of a method to infer the spatial (r)-scales resolving the intrinsic geometry of each dataset.

The topological complexity of the networks increases at a critical r , from exhibiting the trivial traits of a star graph to the structure of amorphous networks with a power-law spectrum for $\lambda \rightarrow 0$. Thus, in the phase plane, the emergence of nontrivial structures is reflected by the rise of a second peak in C together with a roughly $C = C_0 > 0$ plateau between the first and last peak. This is illustrated in Figs. 7(i)–7(l) for each investigated dataset. Such transition is also depicted in Fig. 8 by the bifurcation of the blue and green curves for $\tau_{\text{peak}}^{(1)}$ and $\tau_{\text{peak}}^{(2)}$ indicating the lower and upper branches in Figs. 7(i)–7(l). See Sec. S4 within the SM [67] for the spectral entropy and specific heat curves of the four datasets, shown independently for each r -value.

Once topological complexity increases, maximal functional efficiency is achieved within a narrow region of the amorphous phase for each dataset. In most cases, the two efficiency measures, θ and η , identify the same optimal region of the phase plane.

The MNIST is an interesting case because the dataset exhibits a clear separation of r -scales in the highly dimensional sample space. We start by noting that the specific heat of the MNIST has a third large peak in between $\tau^{(1)}$ and $\tau^{(2)}$ [see Fig. 7(k)]. This feature is only mitigated for larger networks, as shown in Figs. S2 and S7 within the SM [67].

This intermediate specific heat peak is reflected in two maxima for the thermodynamic efficiency, η , and a two-step transition in the amorphous region, in Figs. 7(c) and 7(k). The local maximum is at $r \in [0.35 - 0.45]$, depending on the specific τ -scale. Such a range aligns with the transition at $r = 0.40$ indicated by the bifurcation of the two specific heat peaks in Fig. 8. The global maximum in Fig. 7(c) is close

to the second-step transition at $r = 0.6$. Although the local maximum in η is not detected by θ in Fig. 7(g), it is reflected in its standard deviation, allowing us to include the r -range indicated by the shaded purple area, used to compare η and θ peaks in terms of r . This shaded region corresponds to the η -peak-to-peak r -range in Fig. 7(c).

The two- η -peaks is an intrinsic geometrical characteristic of the MNIST due to the cluster of points in the high-dimensional space corresponding to the digit “1” (see Sec. S8 within the SM [67]) that tend to cluster closer between each other when compared to the samples belonging to other classes. This produces a subtree of “1”-bearing nodes to form at smaller r values when compared to other patterns. For this reason, we say that the MNIST has a clear separation of scales, as opposed to the case of the other datasets where such distinction is smoother and a single resolution scale is identified by the maximization of efficiency.

G. Spectral dimension estimates the dataset complexity

In this section, we show that the complexity of the dataset is encoded in the structural traits of the maximally functionally efficient networks, specifically in their spectral dimension.

First, we examine the location of the optimal θ efficiency in the amorphous region of the phase plane for each dataset. Then, for each dataset, we inspect the traits of the highly efficient structures using the spectral dimension d_s [39–41]. It generalizes the concept of Euclidean dimension to disordered structures and takes noninteger values between 1 and 3. In the case of D -dimensional lattices, it takes the same value as the standard Euclidean dimension. The spectral dimension gauges the effective large-scale connectedness of disordered structures, with high values of d_s indicating high topological connectedness. Moreover, it is related to the asymptotic behavior of diffusion [40], electrical transport, and thermal stability of folded proteins [42]

Depending on where the maximal θ efficiency is observed in the amorphous region of the phase plane in Fig. 7, we see networks changing the value of the specific heat plateau, reflecting their spectral dimension: $d_s = 2C_0$, as we discussed in Sec. II.

Moreover, we observe that the specific heat plateau is an approximately decreasing function of r in the scale-invariant region of the phase plane in Fig. 5(a). See also Sec. S4 within the SM [67]. For complex datasets such as CIFAR10 and FashionMNIST, the optimal θ values lie closer to the transition boundary separating the star graph from the amorphous phase. In contrast, for simpler datasets like NIST and MNIST, the optimal values are located farther from their respective transition points.

We measure the spectral dimension for each dataset ($N=2000$) by averaging over the exponents of the spectral density $P(\lambda) \sim \lambda^\gamma$ as $\lambda \rightarrow 0$ (see Sec. S1 within the SM [67]). The exponents are obtained by fitting the power law using the method introduced by Clauset *et al.* [73], as implemented in the Python library in Ref. [74]. Fits are computed over the range $\lambda \in [1 \times 10^{-3}, 2 \times 10^{-1}]$ for all those networks which better adapted to the dataset in terms of their degree of function, i.e., those included in the purple r -range of maximal

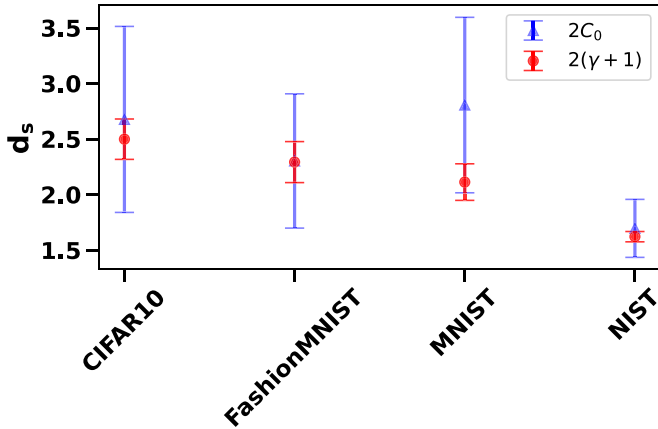


FIG. 9. Datasets of increasing complexity ranked based on spectral dimension. The amorphous networks’ spectral dimension gauges the dataset complexity (Fig. 3). The red points indicate the spectral dimension [41] d_s measured from the power-law exponent of the spectral density. The blue dots are obtained by monitoring the specific heat plateau [35,36], $d_s = 2C_0$, averaged over the resolution $\tau \in [1, 2] \times 10^1$. All points are averaged over the r -range of the phase plane of maximal θ -efficiency, see the shaded purple areas in Figs. 7(e)–7(h), and over 50 independent simulations.

θ efficiency in Figs. 7(e)–7(h). This aligns with the averaging over multiple n neighbors proposed in Ref. [46] when estimating the intrinsic dimension using the MLE method (Sec. III A). In our case, however, we average over the r -resolution scales in the sample space that enable a higher degree of function, i.e., maximal pattern-matching efficiency.

The spectral dimension, $d_s = 2(\gamma + 1)$, is plotted with red dots for each dataset in Fig. 9. We removed the value $\gamma(r = 0.35)$ when measuring the spectral dimension of the CIFAR10 because the spectral distribution was too close to the transition to obtain a good fit with a power law.

We also measure the spectral dimension by averaging the specific heat curves of those θ -efficient networks and then measuring the specific heat plateau C_0 over the resolution range $\tau \in [1, 2] \times 10^1$. The measurements, $d_s = 2C_0$, are shown with blue triangles in Fig. 9.

In a nutshell, once the efficiency is maximized, the datasets of increasing complexity (Fig. 3) are mapped to scale-invariant networks of increasing spectral dimension (Fig. 9). Thus, the network spectral dimension provides a way to gauge the complexity of the dataset. Moreover, we can assess that high-complexity datasets are arranged by the system in topological structures that are more compact than those of lower complexity.

IV. DISCUSSION

The ability of a system to survive is the capability of concentrating a flow of order from the environment—“drinking order from the environment” [75].

For the system analyzed here, we could say that the system is “drinking” the spatial correlations intrinsic to the dataset’s geometry. In other words, systems, whether living systems or machine(-learning) systems, need to learn an internal repre-

sentation of the external world (or the dataset) they are posed in contact with.

Ideally, to be competitive, such internal representation should be efficient in terms of the desired function and available resources. It can be encoded in terms of abstract probability distributions [10,11] or into the very network features of the system, as it happens in genetic regulatory networks playing a central role in morphogenesis [13]. Or in a mixed way, as in artificial neural networks where the engineer designs a complicated distribution through the definition of the architecture and the training phase fills in the parameters to obtain the desired function-approximation machine capable of achieving statistical generalization [76].

When confronted with the real-world environment, to survive, living systems not only need the capability of interpreting the environment, but they also need the capability to quickly react with a coherent response to the incoming stimuli, the latter being possibly drawn from the same distribution they adapted to [77,78]. Moreover, the same system needs to cope with the cost of sending information across its structure, which can lead to an enhanced exploitation of resources. In machine learning, the cost can be quantified by the number of computations required before the satisfying answer to a task, such as class matching, can be reported.

Here, we used an algorithm [27] that creates and navigates trees of real-world patterns to investigate the intertwined role of structure and function in hierarchical self-organized tree systems in a simple setting. The algorithm adapts to the diversity of the environment it is in contact with, represented here by each real-world dataset. It self-organizes, structuring into its network traits the disordered collection of high-dimensional patterns and the geometric information they contain.

The tree formation process undergoes two phase transitions observed at two distinct network scales: The former is a mesoscale transition with the emergence of an approximately scale-invariant structural behavior progressively involving a larger range of network scales; the latter is a large-scale transition from the irregular scale-invariant structure to the regular-scale invariant 1D lattice.

The recently developed theoretical tools of statistical physics for networks based on density matrices [1,33–35,51,57] permit quantifying the information encoded in the traits of the network with the von Neumann entropy and the transmission flow of information with the Helmholtz free energy. As the speed of information flow and a sufficiently rich representation of the external world are two competing capabilities, the formation process of a complex structure can be interpreted as the macroscopic thermodynamic outcome of the delicate balance between the two.

Thus, we first computationally applied the outlined variational principle, formulated in terms of thermodynamic efficiency [1], to explain why a particular network configuration emerges over another within a given environmental complexity.

However, given that the original purpose of the algorithm was to serve as an efficient pattern-searching method, we associated each configuration with a degree of function, that is, the system’s ability to perform a task, which we measured through pattern-matching efficiency.

Empirical evidence suggests that the thermodynamic and the pattern-matching efficiencies show some consistency, namely when we look at the global maxima of the two in the r -range. Thus, solely from the topology, the former can be used to narrow down the control parameter region of optimal function prior to any test phase.

Indeed, these two quantities involve two distinct processes that possess significant differences. On the one hand, the Portegys search phase is a directed process starting from the root node that navigates the structure under a homophily principle. On the other hand, the diffusion propagator describes an undirected ergodic process propagating from any possible node. Moreover, the thermodynamic efficiency is defined based on general statistical physics arguments, and thus, it is task agnostic. On the contrary, the pattern-matching efficiency is related to the specific task under consideration, thus, it is task-related.

Still, it is reasonable to expect similar behavior of the two efficiencies. Both quantities estimate a trade-off between the flow of a process and the multiscale relational information encoded in the system's topological traits.

When we look at the denominator of the two efficiencies, η and θ , we have two quantities measuring the information “flow”, and they both monotonically decrease for increasing r . In one case, the flow of information is quantified by the logarithm of the number of computed distances throughout the Portegys' search process, thus counting the number of nodes that have been visited. In the other case, using a diffusion process as an approximation for Portegys' search dynamics, at the denominator we have $W = \delta F = F + \frac{\log N}{\tau}$, where the free energy is $F = -\frac{\log Z}{\tau}$, with $Z \propto \mathcal{R}(\tau)$ where $\mathcal{R}(\tau)$ is the average return probability after time τ of a continuous time edge-centric random walker [57,79]. Thus, the partition function Z measures the tendency of information to be blocked somewhere in the network, while W , on the contrary, measures the tendency to favor the flow of information.

At the numerator of θ , the accuracy is not necessarily a monotonic decreasing function of the control parameter r . It can exhibit higher values at specific r -values as illustrated by the cases of the CIFAR10 (and the FashionMNIST in the SM [67]).

The central question is what makes a particular structure better suited to perform a function, such as the classification of patterns or a general complex input-output mapping. Thinking of artificial neural networks, performance can be improved either by increasing the number of parameters or by adjusting their values to implement more complex or refined functions, although neither necessarily leads to better generalization.

In the case of our network structures, the parameters are the network traits themselves, which, in Fourier space, become the network frequencies, i.e., the Laplacian eigenvalues λ_i . Unlike artificial neural networks, in our case, the number of parameters cannot be directly increased, as it is constrained by the number of training samples. Moreover, we cannot directly adjust their values. Still, we can use the control parameter r to govern the distribution of frequencies and maximize the extraction of geometric information flowing from the dataset into the self-organized structure. When we compute the numerator in Eq. (6), $U_\tau = \text{Tr}[\hat{\rho}\hat{L}] = \langle \lambda \rangle_\tau$, we measure a

weighted average of the parameters by computing the thermal average of the frequencies contained in the network, and those frequencies are there because of the geometric information collected and stored in the network about the clustering of the dataset's samples in the original sample space. Such information may be leveraged during the search phase for future pattern classification, although it can also include accumulated, spurious information that is not useful for the pattern-matching task. Thus, when we use the θ efficiency, we rely on the information included in the structures that have maximized the accuracy, whereas when we rely on η we also depend on information that is in the network but that might be useless to the task, some sort of useless structural complexity of the internal representation, in analogy to the theoretical works in Refs. [78,80].

Remarkably, in the scale-invariant regime, which includes the amorphous one, $U \propto C_0 \propto d_s$, where C_0 is the specific heat plateau and d_s is the spectral dimension [35,41], which governs the spectral density power-law decay. Note that all highly efficient structures were found in the amorphous regime, regardless of the dataset's complexity, partly supporting the hypothesis of functional benefits from presenting critical-like features [8–14], such as a power-law density related to some quantity of interest.

Nevertheless, for higher-complexity datasets, optimal efficiency in pattern-matching was achieved near the transition boundary separating the star graph and the scale-invariant regime. In contrast, for lower-complexity datasets, highly efficient structures were found on the opposite side of the amorphous phase plane, preceding the transition from amorphous to regular scale-invariant networks.

Then, what determines the location of a network that embeds the geometric and class label information about a distribution of high-dimensional patterns within the amorphous region of the phase plane, when it is queried about its function? By measuring the optimally pattern-matching efficient networks' spectral dimension [39,41] d_s , we related the dimensionality of the network to the complexity of the dataset it models. We observed higher spectral dimension, reflecting less branching and higher connectedness, for networks mapping high-complexity datasets. On the contrary, lower-complexity datasets were mapped into more elongated, complex branching structures with a lower spectral dimension. Thus, importantly, we provided an alternative way to measure the complexity of a dataset: by mapping it into a tree network that maximizes the degree of function and then measuring its spectral dimension.

Returning to the starting question of this paper, that is, why, within an environment, a certain structure emerges in place of another, the answer suggested by our findings is: to manage the complexity of the environment it is adapting to, converting the incoming correlations from the external world into internal topological correlations, producing an internal model of the external world. Moreover, the system exhibits a high degree of function if the dynamics of the functional process can effectively exploit the underlying topological correlations. In such cases, the system leverages the accumulated information for prediction, interpreting novel environmental inputs through the learned structure.

Conversely, we can infer the level of complexity the system has been exposed to by examining its topological traits.

Finally, from a complementary perspective, the principle of *maximal functional efficiency* serves us to identify the informationally meaningful spatial resolution scales of the dataset geometry.

Whether these considerations generalize to systems beyond the Portegys algorithm remains speculative and beyond the scope of this paper. Nevertheless, the methods presented here could, in principle, serve as diagnostic tools for other tree-based classification approaches, such as hierarchical clustering.

Suppose we are given multiple pretrained tree networks, each mapping an unknown dataset. In that case, we can estimate and rank the complexity of the training datasets, as well as assess the ability of each tree structure to cope with that complexity, without requiring a test phase. If a trained network “struggles” to produce a model of the dataset through its traits, it is expected to have a high spectral dimension and exhibit greater connectedness. Conversely, the same network system trained on a low-complexity dataset should display a more branched structure and a lower spectral dimension.

In the reverse direction, knowing the complexity of a dataset, we could initialize the search for the optimal control parameter r closer to values typically presenting one of the two transitions exhibited by the algorithm, reducing the number of simulations required to find the most efficient network structure.

Interestingly, developments of our analysis and methods for hierarchical tree networks could be applied to other topics in artificial intelligence, neuroscience, and neuromorphic computing.

We find it particularly suited to the context of biologically inspired [15–17] dendritic learning [20–23], where input units are connected to output units via loopless paths, to guide the design of tree-like dendritic structures for parameter-efficient learning in artificial neural network [20–23].

Neurons can be modeled as the amorphous tree structures discussed in this work. It would be interesting to experimentally relate the dendritic structure, specifically its spectral dimension, to the capability of the neuron to deal with environmental stimuli and complex input-output mapping of signals [15–19], or to the neuron’s ability to produce signals with richer diversity [77,81,82]. For instance, a neuron capable of dealing with the complexity of the stimulus it is subject to would be expected to exhibit traits with low spectral dimensionality and complex branching.

Moreover, potential applications include neuromorphic computing [44,83,84] currently looking for a possible solution to the ever-increasing computational power demand of recent years. Neuromorphic computing processors are designed to colocalize computation and memory on the physical substrate and emulate the brain’s learning capabilities *in-materia*. Complete neural networks with dendritic branching have been fabricated with memristive devices and have shown significant power efficiency [24]. The full potential of biological neural systems is likely achieved because of the interplay between their complex topological structures and the functional dynamics of the many elements evolving on it [85]. As in bi-

ological neural systems, structurally efficient routing may be crucial for *in-materia* computing processors built from self-assembled materials [86,87]. Emergent tree-like structures are currently physically fabricated in emerging materials representing potential candidates for neuromorphic computing. We mention tree structures produced at the critical point of a percolation transition for randomly dispersed nanoparticles [25,88], the electrochemically formed dendrite morphology at the origin of the resistive-switching of some memristive devices [25,89], and fractal trees produced during the dielectric breakdown [90]. Thus, their information processing capabilities could be investigated and enhanced based on this work and future developments.

V. CONCLUSIONS

In conclusion, we used a variational principle describing the trade-off between performance in a task and consumed resources to show that a complex (tree-)system holds within its spectral dimension the complexity of the environment in which it formed. Specifically, our approach constitutes a dataset-specific and task-related variational principle, in contrast to the one outlined in Ref. [1], which employs the functional $\eta[P(\lambda; \delta)]$, that does not account for the influence of the disordered environment or the specific task the system is required to perform. With δ we indicate a generic parameter governing the network’s structural traits.

In this work, we defined the pattern-matching efficiency functional $\theta[P(\lambda; r, \mathcal{D})]$, which depends on the spectral density $P(\lambda; r, \mathcal{D})$ of the tree network, is modulated by the parameter r , and is conditioned on the dataset $\mathcal{D}$.

We varied r to maximize the functional $\theta[P(\lambda; r, \mathcal{D})]$. Once the optimal value of r was identified, we analyzed the corresponding spectral properties of the network to infer the dataset’s complexity. In this way, we probed the dataset’s intrinsic geometric structure by identifying the tree topology that achieves the best trade-off between classification performance and computational cost.

From a complementary perspective, the variational principle based on θ allowed us to identify the informative r -resolution scales within the structured disorder of the environment, represented as a dataset of real-world patterns. Such r -resolution scale is the one capable of translating useful information about the dataset, i.e., correlations between patterns in the original sample space, into topological correlations between the network nodes. Namely, when using the term “useful”, it is meant in terms of the degree of function achieved by the system, specifically, its pattern-matching efficiency, which quantifies the system’s ability to perform its task, i.e., classifying unseen samples, using minimal resources.

Summarizing the contributions included in this work:

(1) We proposed the Portegys algorithm [27,28] as an alternative process to produce hierarchies for a statistical-physics analysis, incorporating the functional aspect of the accuracy of a machine-learning algorithm;

(2) Multiple phase transitions are identified while such an algorithm searches for an efficient internal representation of the dataset;

(3) Scale-invariance, i.e., power-law Laplacian spectral density [36], is a key feature in constructing and identifying efficient tree networks capable of combining fast information flow with a rich internal representation of information. Our analysis involved task-agnostic thermodynamic-like efficiency [1], based on general statistical-physics arguments, and task-related pattern-matching efficiency, quantifying the trade-off between accuracy and resources used to achieve the task;

(4) Studying efficient tree structures mapping the structured disorder of a dataset, we showed that the complexity of the environmental conditions directly influences the emergence of complex trees. Namely, the spectral dimension [39–41], proportional to the exponent of the power-law spectral density, directly encodes the complexity of the disordered environment;

(5) Therefore, we propose two operational measures to estimate dataset complexity: one by calculating the spectral dimension of the maximally efficient networks mapping the dataset, and the other by measuring their specific heat plateau [34,35]. Moreover, maximizing the functional efficiency of the network traits allows for the inference of the dataset's informational resolution scales.

ACKNOWLEDGMENTS

The authors are grateful to Jos van Goor for providing some of the scripts that enabled a comprehensive grid search across parameters and to Federico Balducci for the useful discussions. This work was funded by EUs Horizon 2020, from the MSCA-ITN-2019 Innovative Training Networks program “Materials for Neuromorphic Circuits” (MANIC) under the Grant Agreement No. 861153, as well as by the financial support of the CogniGron research center and the Ubbo Emmius Funds (Univ. of Groningen).

D.C. Theoretical framework, methodology, simulations, data curation, software, writing of original draft, reviewing, and editing. L.S.: Supervision, funding acquisition, writing, reviewing, and editing.

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

DATA AVAILABILITY

The data that support the findings of this article are openly available [49,50].

-
- [1] A. Ghavasieh and M. De Domenico, Diversity of information pathways drives sparsity in real-world networks, *Nat. Phys.* **20**, 512 (2024).
 - [2] E. Szathmáry and J. M. Smith, The major evolutionary transitions, *Nature (London)* **374**, 227 (1995).
 - [3] J. J. Hopfield, Neural networks and physical systems with emergent collective computational abilities., *Proc. Natl. Acad. Sci. USA* **79**, 2554 (1982).
 - [4] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski, A learning algorithm for Boltzmann machines, *Cognit. Sci.* **9**, 147 (1985).
 - [5] T. Kachman, J. A. Owen, and J. L. England, Self-organized resonance during search of a diverse chemical space, *Phys. Rev. Lett.* **119**, 038001 (2017).
 - [6] N. Perunov, R. A. Marsland, and J. L. England, Statistical physics of adaptation, *Phys. Rev. X* **6**, 021036 (2016).
 - [7] M. De Domenico, *Exploring evolution (in life) through information*, accessed: 2024-04-16.
 - [8] C. G. Langton, Computation at the edge of chaos: Phase transitions and emergent computation, *Physica D* **42**, 12 (1990).
 - [9] N. Bertschinger and T. Natschláger, Real-time computation at the edge of chaos in recurrent neural networks, *Neural Comput.* **16**, 1413 (2004).
 - [10] I. Mastromatteo and M. Marsili, On the criticality of inferred models, *J. Stat. Mech.* (2011) P10012.
 - [11] J. Hidalgo, J. Grilli, S. Suweis, M. A. Muñoz, J. R. Banavar, and A. Maritan, Information-based fitness and the emergence of criticality in living systems, *Proc. Natl. Acad. Sci. USA* **111**, 10095 (2014).
 - [12] B. C. Daniels, H. Kim, D. Moore, S. Zhou, H. B. Smith, B. Karas, S. A. Kauffman, and S. I. Walker, Criticality distinguishes the ensemble of biological regulatory networks, *Phys. Rev. Lett.* **121**, 138102 (2018).
 - [13] D. Krotov, J. O. Dubuis, T. Gregor, and W. Bialek, Morphogenesis at criticality, *Proc. Natl. Acad. Sci. USA* **111**, 3683 (2014).
 - [14] J. M. Beggs and D. Plenz, Neuronal avalanches are diverse and precise activity patterns that are stable for many hours in cortical slice cultures, *J. Neurosci.* **24**, 5216 (2004).
 - [15] G. M. Shepherd and R. K. Brayton, Logic operations are properties of computer-simulated interactions between excitable dendritic spines, *Neuroscience* **21**, 151 (1987).
 - [16] P. Poirazi and A. Papoutsi, Illuminating dendritic function with computational models, *Nat. Rev. Neurosci.* **21**, 303 (2020).
 - [17] S. Hodassman, R. Vardi, Y. Tugendhaft, A. Goldental, and I. Kanter, Efficient dendritic learning as an alternative to synaptic plasticity hypothesis, *Sci. Rep.* **12**, 6571 (2022).
 - [18] G. Kastellakis, A. J. Silva, and P. Poirazi, Linking memories across time via neuronal and dendritic overlaps in model neurons with active dendrites, *Cell Rep.* **17**, 1491 (2016).
 - [19] R. Makarov, M. Pagkalos, and P. Poirazi, Dendrites and efficiency: Optimizing performance and resource utilization, *Curr. Opin. Neurobiol.* **83**, 102812 (2023).
 - [20] S. Chavlis and P. Poirazi, Dendrites endow artificial neural networks with accurate, robust and parameter-efficient learning, *Nat. Commun.* **16**, 943 (2025).
 - [21] J. Guerguiev, T. P. Lillicrap, and B. A. Richards, Towards deep learning with segregated dendrites, *eLife* **6**, e22901 (2017).
 - [22] A. Payeur, J. Guerguiev, F. Zenke, B. A. Richards, and R. Naud, Burst-dependent synaptic plasticity can coordinate learning in hierarchical circuits, *Nat. Neurosci.* **24**, 1010 (2021).
 - [23] J. Sacramento, R. P. Costa, Y. Bengio, and W. Senn, Dendritic cortical microcircuits approximate the backpropagation algorithm, *Advances in Neural Information Processing Systems*, edited by S. Bengio, H. Wallach, H. Larochelle, K. Grauman,

- N. Cesa-Bianchi, and R. Garnett, Vol. 31 (Curran Associates, Inc. Newry, 2018).
- [24] X. Li, J. Tang, Q. Zhang, B. Gao, J. J. Yang, S. Song, W. Wu, W. Zhang, P. Yao, N. Deng *et al.*, Power-efficient neural network with artificial dendrites, *Nat. Nanotechnol.* **15**, 776 (2020).
- [25] J. S. Lee, S. Lee, and T. W. Noh, Resistive switching phenomena: A review of statistical physics approaches, *Appl. Phys. Rev.* **2**, 031303 (2015).
- [26] F. Nielsen, Hierarchical clustering, *Introduction to HPC with MPI for Data Science* (Springer International Publishing, Cham, Switzerland, 2016), pp. 195–211.
- [27] T. Portegys, A search technique for pattern recognition using relative distances, *IEEE Trans. Pattern Anal. Mach. Intell.* **17**, 910 (1995).
- [28] M. D. Accélération de l’algorithme des k plus proches voisins par réorganisation arborescente de la base de données, Ph.D. thesis, Ecole Polytechnique, Paris, 1999.
- [29] F. Jacob, Evolution and tinkering, *Science* **196**, 1161 (1977).
- [30] A. Benatti and L. d. F. Costa, Retrieving hierarchies, [arXiv:2204.07530](https://arxiv.org/abs/2204.07530).
- [31] H. W. Watson and F. Galton, On the probability of the extinction of families., *J. Anthropol. Inst. G.B. Irel.* **4**, 138 (1875).
- [32] T. E. Harris, *The Theory of Branching Processes*, Grundlehren der mathematischen Wissenschaften (Springer, Berlin, 2012).
- [33] M. De Domenico and J. Biamonte, Spectral entropies as information-theoretic tools for complex network comparison, *Phys. Rev. X* **6**, 041062 (2016).
- [34] P. Villegas, A. Gabrielli, F. Santucci, G. Caldarelli, and T. Gili, Laplacian paths in complex networks: Information core emerges from entropic transitions, *Phys. Rev. Res.* **4**, 033196 (2022).
- [35] P. Villegas, T. Gili, G. Caldarelli, and A. Gabrielli, Laplacian renormalization group for heterogeneous networks, *Nat. Phys.* **19**, 445 (2023).
- [36] A. Poggialini, P. Villegas, M. A. Muñoz, and A. Gabrielli, Networks with many structural scales: A renormalization group perspective, *Phys. Rev. Lett.* **134**, 057401 (2025).
- [37] M. Kardar, *Statistical Physics of Fields* (Cambridge University Press, Cambridge, 2007).
- [38] S. di Santo, P. Villegas, R. Burioni, and M. A. Muñoz, Landau–Ginzburg theory of cortex dynamics: Scale-free avalanches emerge at the edge of synchronization, *Proc. Natl. Acad. Sci. USA* **115**, E1356 (2018).
- [39] S. Alexander and R. Orbach, Density of states on fractals: «fractons», *J. Phys. Lett.* **43**, 625 (1982).
- [40] K. Hattori, T. Hattori, and H. Watanabe, Gaussian field theories on general networks and the spectral dimensions, *Prog. Theor. Phys. Suppl.* **92**, 108 (1987).
- [41] R. Burioni and D. Cassi, Universal properties of spectral dimension, *Phys. Rev. Lett.* **76**, 1091 (1996).
- [42] R. Burioni, D. Cassi, F. Cecconi, and A. Vulpiani, Topological thermal instability and length of proteins, *Proteins: Struct. Funct. Bioinf.* **55**, 529 (2004).
- [43] M. Mézard, Spin glass theory and its new challenge: Structured disorder, *Indian J. Phys.* **98**, 3757 (2024).
- [44] C. Mead, How we created neuromorphic engineering, *Nat. Electron.* **3**, 434 (2020).
- [45] A. Lempel and J. Ziv, On the complexity of finite sequences, *IEEE Trans. Inf. Theory* **22**, 75 (1976).
- [46] E. Levina and P. Bickel, Maximum likelihood estimation of intrinsic dimension, in *Advances in Neural Information Processing Systems*, edited by L. Saul, Y. Weiss, and L. Bottou, Vol. 17 (MIT Press, Cambridge, MA, 2004).
- [47] D. J. MacKay and Z. Ghahramani, Comments on “Maximum Likelihood Estimation of Intrinsic Dimension” by E. Levina and P. Bickel — inference.org.uk, <http://www.inference.org.uk/mackay/dimension/> [Accessed 10-07-2024].
- [48] P. Pope, C. Zhu, A. Abdelkader, M. Goldblum, and T. Goldstein, The intrinsic dimension of images and its impact on learning, *International Conference on Learning Representations (ICLR)*, 2021).
- [49] D. Cipollini, Dataset for “Tree networks of real-world data: analysis of efficiency and spatiotemporal scales”, Version V1, Zenodo (2024), <https://zenodo.org/records/13883206>.
- [50] D. Cipollini, GitHub Repository: CipolliniDavide/Laplacian Trees (2024), https://github.com/CipolliniDavide/Laplacian_Trees.
- [51] A. Ghavasieh and M. De Domenico, Generalized network density matrices for analysis of multiscale functional diversity, *Phys. Rev. E* **107**, 044304 (2023).
- [52] N. Masuda, M. A. Porter, and R. Lambiotte, Random walks and diffusion on networks, *Phys. Rep.* **716–717**, 1 (2017).
- [53] W. N. Anderson and T. D. Morley, Eigenvalues of the laplacian of a graph, *Linear Multilinear Algebra* **18**, 141 (1985).
- [54] E. Estrada, *The Structure of Complex Networks: Theory and Applications* (Oxford University Press, Oxford, 2011).
- [55] R. P. Feynman, *Statistical Mechanics: A Set of Lectures* (CRC Press, Boca Raton, FL, 1998).
- [56] P. Moretti and M. Zaiser, Network analysis predicts failure of materials and structures, *Proc. Natl. Acad. Sci. USA* **116**, 16666 (2019).
- [57] A. Ghavasieh, C. Nicolini, and M. De Domenico, Statistical physics of complex information dynamics, *Phys. Rev. E* **102**, 052304 (2020).
- [58] L. P. Kadanoff, Scaling laws for ising models near T_c , *Phys. Phys. Fiz.* **2**, 263 (1966).
- [59] K. Wilson, The renormalization group and the ϵ expansion, *Phys. Rep.* **12**, 75 (1974).
- [60] D. Cipollini, A. Swierstra, and L. Schomaker, Modeling a domain wall network in BiFeO₃ with stochastic geometry and entropy-based similarity measure, *Front. Mater.* **11**, 1323153 (2024).
- [61] C. Hens, U. Harush, S. Haber, R. Cohen, and B. Barzel, Spatiotemporal signal propagation in complex networks, *Nat. Phys.* **15**, 403 (2019).
- [62] L. Schomaker, D. Mangalau, L. Vuurpijl, and M. Weinfeld, *Two Tree-Formation Methods and Fast Pattern Search using Nearest Neighbor and Nearest-Centroid Matching* (International Unipen Foundation, 2000) p. 261–270.
- [63] P. Grother, Nist special database 19: Handprinted forms and characters database, 1995, Natl. Inst. Stand. Technol. available at <https://api.semanticscholar.org/CorpusID:59785963>.
- [64] L. Deng, The mnist database of handwritten digit images for machine learning research [best of the web], *IEEE Signal Process. Mag.* **29**, 141 (2012).
- [65] H. Xiao, K. Rasul, and R. Vollgraf, Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms, [arXiv:1708.07747](https://arxiv.org/abs/1708.07747).

- [66] A. Krizhevsky, G. Hinton, *et al.*, Learning multiple layers of features from tiny images, available at <http://www.cs.utoronto.ca/~kriz/learning-features-2009-TR.pdf>.
- [67] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/ztmk-gl4> for figures referenced in the main text but omitted for clarity.
- [68] C. Destri and L. Donetti, The spectral dimension of random trees, *J. Phys. A: Math. Gen.* **35**, 9499 (2002).
- [69] M. E. J. Newman, Modularity and community structure in networks, *Proc. Natl. Acad. Sci. USA* **103**, 8577 (2006).
- [70] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, Fast unfolding of communities in large networks, *J. Stat. Mech.* (2008) P10008.
- [71] B. Cramer, D. Stöckel, M. Kreft, M. Wibral, J. Schemmel, K. Meier, and V. Priesemann, Control of criticality and computation in spiking neuromorphic networks with plasticity, *Nat. Commun.* **11**, 2853 (2020).
- [72] A. Goudarzi, C. Teuscher, N. Gulbahce, and T. Rohlfs, Emergent criticality through adaptive information processing in boolean networks, *Phys. Rev. Lett.* **108**, 128702 (2012).
- [73] A. Clauset, C. R. Shalizi, and M. E. J. Newman, Power-law distributions in empirical data, *SIAM Rev.* **51**, 661 (2009).
- [74] J. Alstott, E. Bullmore, and D. Plenz, powerlaw: A python package for analysis of heavy-tailed distributions, *PLoS One* **9**, e85777 (2014).
- [75] E. Schrödinger and R. Penrose, *What is Life? With Mind and Matter and Autobiographical Sketches* (Cambridge University Press, Cambridge, 2012).
- [76] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, Cambridge, MA, 2016).
- [77] R. Pryluk, Y. Kfir, H. Gelbard-Sagiv, I. Fried, and R. Paz, A tradeoff in the neural code across regions and species, *Cell* **176**, 597 (2019).
- [78] K. Friston, The free-energy principle: A unified brain theory? *Nat. Rev. Neurosci.* **11**, 127 (2010).
- [79] A. Ghavasieh and M. De Domenico, Enhancing transport properties in interconnected systems without altering their structure, *Phys. Rev. Res.* **2**, 013155 (2020).
- [80] S. Still, D. A. Sivak, A. J. Bell, and G. E. Crooks, Thermodynamics of prediction, *Phys. Rev. Lett.* **109**, 120604 (2012).
- [81] Z. F. Mainen and T. J. Sejnowski, Influence of dendritic structure on firing pattern in model neocortical neurons, *Nature (London)* **382**, 363 (1996).
- [82] R. A. J. van Elburg and A. van Ooyen, Impact of dendritic size and dendritic topology on burst firing in pyramidal cells, *PLoS Comput. Biol.* **6**, e1000781 (2010).
- [83] G. Indiveri, B. Linares-Barranco, R. Legenstein, G. Deligeorgis, and T. Prodromakis, Integration of nanoscale memristor synapses in neuromorphic computing architectures, *Nanotechnology* **24**, 384010 (2013).
- [84] D. V. Christensen, R. Dittmann, B. Linares-Barranco, A. Sebastian, M. L. Gallo, A. Redaelli, S. Slesazeck, T. Mikolajick, S. Spiga, S. Menzel *et al.*, 2022 roadmap on neuromorphic computing and engineering, *Neuromorph. Comput. Eng.* **2**, 022501 (2022).
- [85] L. E. Suárez, B. A. Richards, G. Lajoie, and B. Misic, Learning function from structure in neuromorphic networks, *Nat. Mach. Intell.* **3**, 771 (2021).
- [86] A. Vahl, G. Milano, Z. Kuncic, S. A. Brown, and P. Milani, Brain-inspired computing with self-assembled networks of nano-objects, *J. Phys. D* **57**, 503001 (2024).
- [87] D. Cipollini, H. Grotorex, M. Mastella, E. Chicca, and L. Schomaker, Fused-MemBrain: a spiking processor combining CMOS and self-assembled memristive networks, *Neuro. Comput. Eng.* **5**, 024002 (2025).
- [88] J. B. Mallinson, S. Shirai, S. K. Acharya, S. K. Bose, E. Galli, and S. A. Brown, Avalanches and criticality in self-organized nanoscale networks, *Sci. Adv.* **5**, eaaw8438 (2019).
- [89] X. Guo, C. Schindler, S. Menzel, and R. Waser, Understanding the switching-off mechanism in Ag⁺ migration based resistively switching model systems, *Appl. Phys. Lett.* **91**, 2793686 (2007).
- [90] L. Niemeyer, L. Pietronero, and H. J. Wiesmann, Fractal dimension of dielectric breakdown, *Phys. Rev. Lett.* **52**, 1033 (1984).