Heat diffusion: Thermodynamic depth complexity of networks

Francisco Escolano[*]
University of Alicante, Alicante, Spain

Edwin R. Hancock[†]
University of York, York, United Kingdom

Miguel A. Lozano[‡]
University of Alicante, Alicante, Spain

In this paper we use the Birkhoff–von Neumann decomposition of the diffusion kernel to compute a polytopal measure of graph complexity. We decompose the diffusion kernel into a series of weighted Birkhoff combinations and compute the entropy associated with the weighting proportions (polytopal complexity). The maximum entropy Birkhoff combination can be expressed in terms of matrix permanents. This allows us to introduce a phase-transition principle that links our definition of polytopal complexity to the heat flowing through the network at a given diffusion time. The result is an efficiently computed complexity measure, which we refer to as flow complexity. Moreover, the flow complexity measure allows us to analyze graphs and networks in terms of the thermodynamic depth. We compare our method with three alternative methods described in the literature (Estrada's heterogeneity index, the Laplacian energy, and the von Neumann entropy). Our study is based on 217 protein-protein interaction (PPI) networks including histidine kinases from several species of bacteria. We find a correlation between structural complexity and phylogeny (more evolved species have statistically more complex PPIs). Although our methods outperform the alternatives, we find similarities with Estrada's heterogeneity index in terms of network size independence and predictive power.

PACS number(s): 89.75.Kd, 89.75.Hc, 89.75.Fb

I. INTRODUCTION

The quantification of the *intrinsic complexity* of undirected graphs and networks has attracted significant attention due to its fundamental practical importance, not only in network analysis [1] but also in other areas such as pattern recognition and control theory. Such a quantification not only allows the complexity of different graph or network structures to be compared, but also allows the complexity to be traded against goodness of fit to data when a structure is being learned in an information theoretic description length framework [2]. Some of the existing quantifications are easily computable, i.e., they have polynomial computational complexity, but others are not since they rely on NP-hard problems (where NP indicates solvable by a non deterministic Turing Machine in polynomial time) or noncomputable quantities. In this paper we focus on the polynomial class of methods, although we also briefly discus the most influential methods falling into the second class. Specifically, we focus in more detail on the dichotomy induced by distinguishing between *randomness complexity* and *statistical complexity*. Randomness complexity aims to quantify the degree of randomness or disorganization of a combinatorial structure; this approach aims to characterize an observed graph structure probabilistically and compute its associated Shannon entropy. Statistical complexity, on the other hand, aims to characterize a combinatorial structure using statistical features such as node degree statistics, edge

density, or the Laplacian spectrum. Viewed historically, most early work in this area falls into the randomness class, while recent work is statistically based.

Turning our attention first to randomness complexity, we note that one of the earliest contributions was Körner's graph entropy [3,4]. The method poses complexity characterization as an optimization problem. Assuming there is a probability distribution associated with the vertices of the graph, the complexity is the minimal cross entropy between the probability distribution and the vertex packing polytope of the graph. Unfortunately the approach is not applicable to more general unweighted graphs. Körner's approach relies on computing chromatic numbers, which is in general an NP-complete problem, but polynomial for perfect graphs. Recently, the von Neumann entropy (or quantum entropy) has been applied to the domain of graphs [5,6] through a mapping between discrete Laplacians and quantum states [7]. If the discrete Laplacian [8] is scaled by the inverse of the volume of the graph, we obtain a density matrix whose entropy can be computed using the spectrum of the discrete Laplacian. The measure distinguishes between different structures. For instance it is maximal for random graphs, minimal for complete ones, and takes on intermediate values for star graphs. In addition, when there is degree heterogeneity then the Shannon (classical) and von Neumann (quantal) entropies are correlated. Turning our attention to [9,10], graphs are decomposed into substructures by grouping vertices so that the discrete entropy is computed from the local vertex density. In [11], this approach is generalized by using j-spheres (subgraphs around a vertex having a maximum topological distance or minimal path length j) to construct the required probability distribution over vertices. Moreover, this approach allows a complexity trace to be defined and this can

———————
[*]sco@dccia.ua.es
[†]erh@cs.york.ac.uk
[‡]malozano@dccia.ua.es

be used to characterize phase changes in structure. In these two approaches, a complete graph must be as complex as any other graph with the same number of vertices. Off-diagonal complexity [12] relies on taking statistics from the number of joint occurrences of pairs of nodes having a given pair of cardinalities. After normalization the method allows a discrete entropy to be computed, which takes on the value zero for both regular lattices and complete graphs and has small values for random graphs and large values for apparently complex structures.

The main drawback of randomness complexity is that it does not capture properly the correlations between vertices [13]. Statistical complexity aims to overcome this problem by measuring regularities beyond randomness and does not necessarily grow monotonically with randomness. Both completely random systems and completely ordered ones must have a minimal statistical complexity. Logical depth [14] (the time required by a universal Turing machine to run the minimal program that reproduces it) is the statistical counterpart of Kolmogorov-Chaitin [15,16] randomness complexity because it is based on the notion of a process rather than a measure. In [17], graph complexity is characterized using the minimal dimension in which its corresponding intersection graph can be embedded. Jukna [18] uses similar ideas and defines the complexity of a graph to be the minimum number of unions and intersections (i.e., Boolean operations) needed to obtain its complete set of edges commencing from stars. Neal and Orrison's linear complexity of a graph [19] is given by the number of additions, subtractions, and multiplications needed to perform matrix multiplication of the adjacency matrix with an arbitrary vector. Linear complexity is independent of the permutations of the vertices. If we reduce the original adjacency matrix by removing redundant rows (those with identical connectivity patterns), rows of zeros (isolated vertices), and rows with only a single one (vertices of unit degree), the linear complexity of the reduced matrix does not change. Relating the linear complexity of a reduced matrix to that of its transpose, we can obtain recursively the linear complexity of the original matrix. Gell-Mann's effective complexity [20,21] (the length of a highly compressed description of its regularities) leads to quantification of a high degree of randomness in terms of high complexity and regularity in terms of low complexity. One interesting property of effective complexity is its ability to distinguish purely random structures from regular ones. This property is shared with off-diagonal complexity. The latter approach is formally linked with the minimum description length principle, and has been used in practice for learning tree structure [2]. Turning our attention to graph-spectral methods, we note that Song *et al.* [22] have recently explored the use of the Laplacian energy [23], i.e., the sum of absolute differences between the eigenvalues and the average vertex degree, as a complexity measure for graphs. In a regular graph, the Laplacian energy is equal to the energy of the graph (the sum of the absolute values of the eigenvalues of its adjacency matrix). The Laplacian energy is also low for graphs associated with polygons.

An important recent addition to the graph-spectral literature is Estrada's network heterogeneity index [24]. The index gauges differences in degree for all pairs of connected vertices. This index can be expressed as a quadratic form of the Laplacian and it is equal to zero for regular and random graphs and equal to 1 for the star graph.

The thermodynamic depth [25] is an alternative measure which takes on low values for both random and ordered systems. The thermodynamic depth (TD) shares with logical depth the idea of using physical depth within a network. However, the TD relies on the causality of heat flow, rather than the number of computational steps. Parallel complexity fills the gap between logical depth and TD. Here depth is defined as the number of computational steps needed by a parallel computer to construct a given graph [26]. Within this framework complex structures require long times to emerge [27]. The length of the time span depends on the logical depth (the number of steps taken by a Turing machine with only local communication), and this differs from parallel depth [the number of steps taken by parallel random access machines (PRAMs), which are also endowed with global communication]. In addition, logical depth refers to individual system states, whereas parallel depth refers to ensembles. Parallel depth may be viewed as the running time of a Monte Carlo algorithm that generates a typical state of the system by sampling with a PRAM endowed with polynomial bounded hardware. Interactions between the components of the system play a fundamental role and can potentially lead to deep states at critical points (e.g., simulating the Ising model between highly ordered and highly disordered states). In this paper we borrow some ideas from the parallel depth (or physical depth) including that of phase transitions. However, we will focus on the application of the TD to measure the complexity of graphs. The main problem in extending the TD to graphs is to define the required macroscopic states [28]. When dealing with graphs, the TD aims to quantify the difficulty in constructing a given graph (the macroscopic state) from scratch (microscopic states). If there is little uncertainty about the process and all the possible causal trajectories have narrow variability, then the graph is narrow (simple). Otherwise, when historical uncertainty arises and many causal trajectories have been excluded for reaching a given structural design, the graph is deep (complex).

In this paper we link the ideas of information flow and thermodynamic depth to construct a different characterization of graph complexity. To commence, we establish a link between heat kernels and Birkhoff polytopes. This choice is motivated by the notion of Körner's entropy. The use of Birkhoff polytopes is motivated by the fact that the heat kernel of a graph can be decomposed into permutation matrices. We draw on the work of Birkhoff and von Neumann to characterize structural complexity in terms of the entropy of the polytopal decomposition coefficients. As the heat kernel is parametrized by time (diffusion extent), we obtain a complexity trace as time evolves. We find that, in practice, these traces are endowed with a phase change. Unfortunately, the mathematical framework for polytopes makes it difficult to characterize the phase change. We show the relationship between heat kernels and the permanent of a graph. The permanent of a graph is a necessary prerequisite to computing the maximum entropy decomposition of the kernel. As the latter problem is known to be #P (the set of the counting problems associated with the decision problems in the set NP), we derive a fast complexity measure (*flow complexity*) which has a similar qualitative behavior. We establish that the existence of a phase-transition point in the flow complexity trace implies that the entropy of the corresponding Birkhoff decomposition is maximal. As a

result, we can assimilate phase transitions (maximum flow) into maximum entropy. Once the flow complexity trace is defined, we proceed to specify (a) the microstates, i.e., the nodes of the graphs, and (b) the causal histories, i.e., sequences of subgraphs rooted at each node, where each subgraph is generated from the previous one by reaching nonvisited nodes through one edge. According to this picture, each node generates a history which is a path to the macroscopic state. Each history is characterized by a flow complexity trace. We quantify the historical variability with the minimal radius of the Bregman ball that encloses all traces. We also consider the projection of the flow complexity trace corresponding to the van der Waerden matrix to exploit the notion of entropy as the divergence from the uniform distribution. The proposed measure of depth combines both historical variability and divergence from the uniform. The experimental evaluation of the method focuses on the analysis of protein-protein interaction (PPI) networks. Given a key protein implicated in signal transduction, our approach finds a significant correlation between the heat kernel–thermodynamic depth complexity and the evolution of PPIs that include the protein in question but in samples from different species of bacteria. We compare the predictive power of our proposed method with three alternative graph-spectral complexity measures, namely, (a) Estrada's heterogeneity index, (b) the von Neumann entropy, and (c) the Laplacian energy (all embedded in the framework of the thermodynamic depth).

As an example of how our approach works, in Fig. 1 we show the heat kernel–thermodynamic depth complexities for several graphs with the same numbers of nodes. As regularity increases then so complexity decreases. In this regard, an eight-neighbor grid graph constrained to a rectangular grid is more regular than a four-neighbor grid graph constrained to the same boundary because it is closer to the complete graph. In our approach any complete graph has zero complexity.
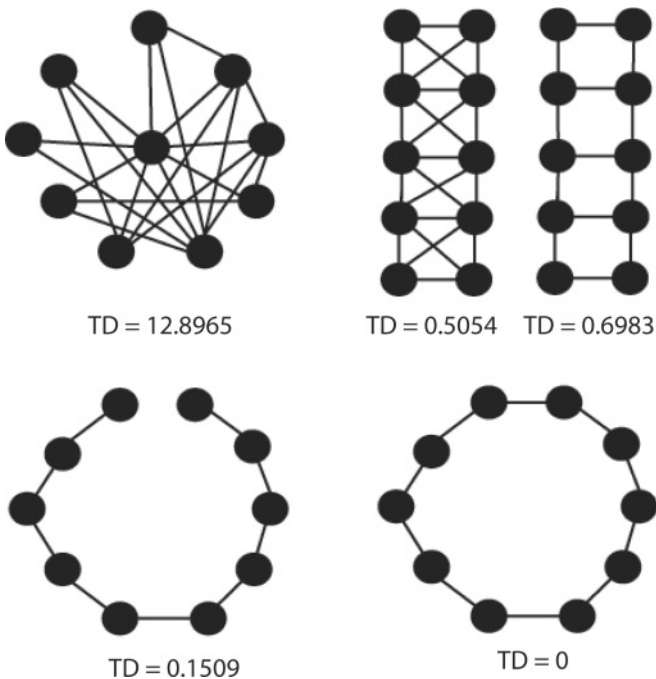

FIG. 1. Heat kernel–thermodynamic depth complexities for graphs of $|V| = 10$. From top to bottom and left to right, Gauss10, Grid8N10, Grid4N10, Line10, and Circle10 graphs.

A complete graph is characterized by having the maximal regularity degree. In this example, it is particularly interesting to note the difference between the loop (zero complexity) and the linear graph. From a TD perspective, each node in the loop has the same history before reaching the macroscopic state. However, a nonzero TD is derived from the fact that each node has a *different view* of the process leading to the emergence of the graph. Each of the possible alternatives is built into the flow complexity traces which register the amount of information flowing through the network at each particular diffusion time. This proceeds from a zero time, where each node retains a unit of heat, until the equilibrium is reached where we have the same amount of heat both in the nodes and in the edges. For any graph the equilibrium state of the diffusion process is always the van der Waerden matrix, and there is no loss of heat (heat conservation). Consequently, part of the heat encodes indirect or transitive relationships between the vertices of the graph, which are not encoded by the graph itself. The different histories leading to the macroscopic state arise from the high variability in the set of complexity traces emanating from different vertices. In the case of complete graphs, the variability is zero, and this is a paradigmatic example of building a graph without imposing any constraint in the design (topology) of the microstate. Without constraints, information flows freely through the network. This situation is similar to that of the loop, also with zero complexity, because in this latter case information flows isotropically or symmetrically. However, when we break a link of the loop, the result is that the rate of heat loss at the extreme vertices is smaller than that in the interior. This is true for all diffusion times, but quantitively different for different histories, and thus leads to a small but nonzero complexity.

II. PRELIMINARIES

In this section we review the prerequisites to the development of our complexity characterization. We commence by defining the Birkhoff polytope, then explain links to the heat kernel, and finally show how to define complexity using the Shannon entropy divided by the expansion coefficients of the Birkhoff–von Neumann decomposition of the heat kernel.

A. The Birkhoff polytope

The nth Birkhoff polytope $\mathcal{B}_n$ is the set of doubly stochastic matrices of dimension $n \times n$:

$$\mathcal{B}_n = \left\{ B = [b_{ij}]_{n \times n} \; : \; \begin{array}{l} \sum_i b_{ij} = 1, \, \forall \, j \\ \sum_j b_{ij} = 1, \, \forall \, i \\ b_{ij} \geqslant 0, \, \forall i \, \forall \, j \end{array} \right\}. \quad (1)$$

Such a polytope is convex, has dimension $(n-1)^2$, and its extreme points are the set of $n!$ permutation matrices $P = [p_{ij}]_{n \times n}$ (entries are 0 except for a single 1 in each row and column).

According to the *Birkhoff–von Neumann (BvN) theorem*, Every doubly stochastic matrix (DSM) can be expressed as a convex combination of permutation matrices (PMs) [29]:

$$B = \sum_\alpha p_\alpha P_\alpha, \quad \forall \, B \in \mathcal{B}_n, \quad \text{and} \quad \sum_\alpha p_\alpha = 1,$$

$$p_\alpha \geqslant 0 \, \forall \, \alpha, \quad (2)$$

where α indexes the coefficients of the convex combination. Thus $\mathcal{B}_n$ is the convex hull of the set of the $n \times n$ permutation matrices. However, the representation of a DSM in terms of many PMs is not unique because $\mathcal{B}_n$ is not a simplex. The barycenter of $\mathcal{B}_n$ is the so-called van der Waerden constant matrix B_* with all entries equal to $1/n$.

B. Birkhoff–von Neumann decomposition of heat kernels

Consider an undirected graph $G = (V, E)$ where $|V| = n$ with $n \times n$ adjacency matrix A. The Laplacian of the graph is the $n \times n$ matrix $\mathcal{L} = D - A$ where $D = \operatorname{diag}(d_i = \sum_{j \in V} A_{ij}, i \in V)$ is the degree matrix. The $n \times n$ *heat or diffusion kernel* $K_\beta(G)$ of the graph is the solution to the heat or diffusion equation: $\frac{\partial K_\beta}{\partial \beta} = -\mathcal{L} K_\beta$, and is given by the matrix exponentiation $K_\beta(G) = \exp(-\beta \mathcal{L})$ (with $\beta \geqslant 0$). The matrix K_β is a doubly stochastic matrix, where $K_{\beta_{ij}}$ encodes the probability of reaching vertex j from vertex i, and vice versa, through lazy random walks [30]. In fact the state vector of the random walk at time β is $q_\beta = K_\beta(G) q_0$, where q_0 is the initial state vector of the walk. For a given value of β, the decomposition $K_\beta(G) = \sum_\alpha p_\alpha P_\alpha$ will not only map a graph (for a given β) into the Birkhoff polytope $\mathcal{B}_n$, but will also associate with the mapping a set of reals (the expansion coefficients) satisfying the axioms of probability. In other words it associates with the mapping a probability distribution over the convex combination of polytopes P_α. This defines the solution to a heat transportation problem for a given β. An interesting example is the complete graph of n vertices C_n whose diffusion kernel is the barycenter $K_{\beta>0}(C_n) = B_*$, that is, the van der Waerden matrix. In addition, the probability distribution, with exactly n coefficients, associated with the BvN decomposition is the uniform one $\mathcal{U}_n$ where $p_\alpha = 1/n, \forall \alpha$.

III. POLYTOPAL COMPLEXITY

Having established the link between the heat kernel and the Birkhoff–von Neumann decomposition, in this section we exploit this link to define a complexity measure for a graph using the Shannon entropy divided by the expansion coefficients.

A. The BvN constructive decomposition

We compute the BvN decomposition recursively. The procedure is as follows: (i) initially express the doubly stochastic matrix $K(G)$ as a convex combination of a single permutation matrix and residual doubly stochastic matrix $K(G) \equiv K_1 = \lambda_1 P_1 + (1 - \lambda_1) K_2$; and (ii) iteratively repeat this procedure using the decomposition $K_r = \lambda_r P_r + (1 - \lambda_r) K_{r+1}$ until the final residual DSM is itself a permutation. Let γ be the number of permutations needed to encode K, satisfying $2 < \gamma < n^2$, and let r be the step number of the recursion process. With these ingredients, the decomposition is

$$K(G) = \lambda_1 P_1 + (1 - \lambda_1) \lambda_2 P_2 + (1 - \lambda_1)(1 - \lambda_2) \lambda_3 P_3 + \cdots$$

$$= \sum_{\alpha=1}^{\gamma} \underbrace{\left\{ \prod_{r=1}^{\alpha-1} (1 - \lambda_r) \right\} \lambda_\alpha}_{p_\alpha} P_\alpha = \sum_{\alpha=1}^{\gamma} p_\alpha P_\alpha. \quad (3)$$

B. BvN graph complexity

Given an undirected and unweighted graph $G = (V, E)$ with diffusion kernel $K_\beta(G)$, and BvN decomposition $K_\beta(G) = \sum_{\alpha=1}^{\gamma} p_\alpha P_\alpha$, we define the *polytopal complexity of G at β* as the normalized entropy

$$\mathcal{B}_\beta(G) = \frac{H(\mathcal{P})}{\log_2 n}, \quad (4)$$

where $\mathcal{P} = \{p_1, \ldots, p_\gamma\}$ is the set of Birkhoff–von Neumann expansion coefficients and $H(\mathcal{P}) = -\sum_{\alpha=1}^{\gamma} p_\alpha \log_2 p_\alpha$ is the Shannon entropy. The definition depends on the value of β and also takes into account both the size of the graph and the number of components of the decomposition. For a complete graph C_n, where the set of expansion coefficients is $\mathcal{U}_n$, our definition of graph complexity reduces to the entropy ratio $H(\mathcal{P})/H(\mathcal{U}_n)$. Moreover, when β approaches zero or infinity, we have the following limiting values:

$$\lim_{\beta \to 0} \mathcal{B}_\beta(G) = 0 \quad \text{and} \quad \lim_{\beta \to \infty} \mathcal{B}_\beta(G) = 1, \forall G. \quad (5)$$

The polytopal complexity is determined by the structure of the heat kernel and the value of β. To analyze the polytopal complexity in more detail, we perform the Taylor expansion

$$K_\beta(G) = e^{-\beta \mathcal{L}} = I_n - \beta \mathcal{L} + \frac{\beta^2}{2!} \mathcal{L}^2 - \frac{\beta^3}{3!} \mathcal{L}^3 + \cdots, \quad (6)$$

where I_n is the $n \times n$ identity matrix, as $\lim_{\beta \to 0} K_\beta(G) \approx I_n - \beta \mathcal{L}$. That is, for low values of β the kernel is dominated by the Laplacian. Under these conditions, the role of I_n is to preserve the double stochasticity of the kernel. In terms of the diffusion process, having $K_\beta(G) = I_n$ at $\beta = 0$ means that every node still retains all its heat (the unit) and therefore the total heat in the system is n. As I_n is itself a member of the set of permutation matrices, its decomposition is trivial and so $\mathcal{B}_{\beta=0}(G) = 0$.

To understanding the behavior of the polytopal complexity at large values of β we use the spectral decomposition of the heat kernel:

$$K_\beta(G) = e^{-\beta \lambda_1} \phi_1 \phi_1^T + e^{-\beta \lambda_2} \phi_2 \phi_2^T + \cdots + e^{-\beta \lambda_n} \phi_n \phi_n^T, \quad (7)$$

where $\lambda_1 = 0 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_n$ are the ordered eigenvalues of $\mathcal{L}$ and $\phi_1, \phi_2, \ldots, \phi_n$ the corresponding eigenvectors. As a result, for large β the spectrum of K_β is dominated by the smallest eigenvalues of $\mathcal{L}$. Since $\lambda_1 = 0$, and we are dealing with graphs having only a single connected component (i.e., the multiplicity of the smallest eigenvalue is 1), we have that $\lim_{\beta \to \infty} K_\beta(G) \approx \phi_1 \phi_1^T + e^{-\beta \lambda_2} \phi_2 \phi_2^T$. Since $\phi_i = \frac{1}{\sqrt{n}} e$, where e is the all-1 's vector of length n, then $\phi_1 \phi_1^T = B_*$, the van der Waerden matrix. As a result when β becomes large enough, say $\beta_{\max}$, the diffusion process reaches the equilibrium state, that is, we have $\mathcal{B}_{\beta_{\max}}(G) \equiv K_{\beta>0}(C_n) = 1$. One consequence of this result is that a complete graph will have unit polytopal complexity.

The above analysis suggests that the graph complexity trace $\mathcal{B}_\beta(G)$ is a signature of the interaction between the heat diffusion process and the structure or topology of the graph as β (and thus the range of interactions between vertices) changes. It can also be interpreted as a trajectory in $\mathcal{B}_n$ between the extreme point given by the identity permutation $P_I = I_n$ and the barycenter $B_* = K_{\beta>0}(C_n)$. The typical

signature is heavy tailed, monotonically increasing from 0 to $\beta^+ = \arg\max\{\mathcal{B}_\beta(G)\}$, and either monotonically decreasing or being stable from β^+ to $\beta_{\max}$ (the smallest $\beta \in R^+$ satisfying the equilibrium condition). Thus, β^+ represents the most significant *topological phase transition* regarding the impact of the diffusion process in the topology of the input graph. For instance, when comparing star graphs with linear graphs of the same size n, we have that in the star graph G_{star} the central (salient) node diffuses its heat faster than the remainder of its $n-1$ neighbors. This allows that $\mathcal{B}_{\beta^+}(G_{\text{line}}) \geqslant \mathcal{B}_{\beta^+}(G_{\text{star}})$, since all the edges connect one of the $n-1$ peripheral nodes to the salient one, and heat transfer along the edges is isotropic. This isotropy has two consequences. First, it produces a small amount of heat in the network before equilibrium is reached. Second, it results in fast convergence toward equilibrium (the heat at the salient node reduces toward $1/n$ faster than the decrease of the peripheral ones). However, for G_{line} we have two salient nodes (the extreme ones) with one edge, and $n-2$ internal nodes with two connecting edges. Heat flows slowly from the extreme nodes and rapidly from the internal nodes. As a result heat transfer is anisotropic, and the closer to one of the extreme nodes the slower the heat transfer. This implies a higher amount of heat in the network than in the star case for all β before the equilibrium is reached. Beyond particular examples, a more formal analysis of the complexity trace is clearly required. In previous work [31,32] we show experimental results supporting our hypothesis about the polytopal complexity trace. However, no characterization principle has been enunciated so far in order to validate the latter assumption.

IV. POLYTOPAL VS HEAT FLOW COMPLEXITY

A. Links with matrix permanents

The analysis of the behavior of the diffusion process in the interval $[0, \beta^+)$ is key to understanding the phase transition point. In this interval, there is a decreasing number of $K_{\beta_{ii}} \approx 1$ of $K_\beta(G)$, and equivalently an increasing number of elements for which $K_{\beta_{ij}} = 0$, $i \neq j$, as β increases. This motivates the analysis of polytopal complexity in terms of *the rate of loss of perfect matchings over this interval*. Let A be the $n \times n$ adjacency matrix associated with a given heat kernel $K_{\beta_{ij}}$ such that $A_{ij} = 1$ if $K_{\beta_{ij}} > 0$, for a given value of β. The *permanent* of A is

$$\text{per}(A) = \sum_{\pi \in \mathcal{S}_n} \prod_{i=1}^{n} A_{i\pi(i)}, \qquad (8)$$

where $\mathcal{S}_n$ is the set of permutations of $\{1, 2, \ldots, n\}$ and $\pi(i)$ is the ith element of the π permutation [33]. As each product can be 1 only when a perfect matching exists in the bipartite graph induced by A, the overall sum counts the number of perfect matchings in such a bipartite graph. As $K_{\beta=0}(G) = I_n$ the minimum number of perfect matchings in the interval $[0, \beta^+]$ is 1.

For the sake of polytopal complexity it is useful to note that $\forall B \in \mathcal{B}_n : 0 < \text{per}(B) \leqslant 1$ and $\text{per}(B) = 1 \Leftrightarrow B$ is a permutation matrix. Thus, an alternative definition of graph complexity relies on $\mathcal{B}'_\beta(G) = 1 - \text{per}[K_\beta(G)]$ defining the $\mathcal{B}'_\beta(G)$ profile in $[0, \beta_{\max}]$. However, the latter profile is

informative only in the interval $[0, \beta^+)$ and it is typically flat and equal to 1 for $\beta \geqslant \beta^+$.

The van der Waerden conjecture, proposed in 1926, states that $\forall B \in \mathcal{B}_n : \text{per}(B) \geqslant \text{per}(B_*) \equiv n!/n^n$ and $\text{per}(B) = \text{per}(B_*) \Leftrightarrow B = B_*$. The proofs underpinning the latter conjecture (which embodies a minimization problem and establishes a minimal value for the permanent) reveal that the permanent is a *mixed discriminant* of diagonal quadratic forms [34]. Mixed discriminants are considered generalizations of the permanent and rely on determinants. Valiant [35] proved that computing the permanent is a #P problem. He also conjectured that the permanent cannot be computed by circuits of polylogarithmic depth (using a PRAM model with a polynomial number of processors). This means that the problem is not in NC: the set of decision problems decidable in polylogarithmic time on a parallel computer with a polynomial number of processors. The latter conjecture is key to understanding not only the hardness of computation of the permanent, but also to understanding the relation between NC and #P classes. Mulmuley has explored the determinant vs permanent problem in the so-called geometric complexity theory. He has recently shown that #P is not included in NC [36]. The latter proof relies on showing that the permanent of an $n \times n$ matrix cannot be represented linearly as the determinant of another matrix of dimension $m \times m$ where $m = 2^{\ln^a n}$, $a > 0$. This in turn means that the problem of computing the permanent of a matrix is highly sequential. In terms of *parallel complexity* it is a so-called *deep problem*.

In practice, one of the best algorithms [37] (the Markov chain Monte Carlo fully polynomial time approximate scheme) for computing an approximation of the permanent takes $O(n^{23})$ on average only for generating a sample of the Markov chain on which it it relies. However, when small graphs are available we may use Ryser's exact algorithm, which requires $\Theta(n2^n)$ operations.

The problem of finding $\text{per}(B)$ is closely related to that of finding the maximum entropy (and unique) BvN decomposition. As pointed out by Slater [38], the problem of finding such a decomposition is #P hard, since it involves the computation of matrix permanents. Therefore, finding a unique BvN decomposition is too hard to do in practice unless the graph is small or can be simplified (for instance with the method proposed in [39]). Graph simplification is required for $n \geqslant 70$ even when we exploit the constructive proof of the Birkhoff–von Neumann theorem to find one of the possible decompositions: as the number of iterations of the BvN decomposition is $O(n^2)$ and a Kuhn-Munkres algorithm [$O(n^3)$] is executed at each iteration, we have an $O(n^5)$ complexity per β value. This complexity precludes the use of the descriptor for practical analysis of complex networks. In addition the analysis of phase change is very cumbersome in the polytopal framework. Thus, a different descriptor, qualitatively similar but more efficient than the current one, and also providing a simpler analytical framework, is needed.

B. Heat flow complexity

The connection of a phase change at β^+ with the gain of weighted perfect matchings, and consequently with the permanents of doubly stochastic matrices, is intriguing. Although it does not yield a practical method for simplifying

the computation of polytopal complexity, it has inspired the dynamic analysis of the *diffusion flow over the structure* as β changes. As we known, the spectral decomposition of the diffusion kernel is $K_\beta(G) = \exp(-\beta\mathcal{L}) \equiv \Phi\Lambda\Phi^T$, where $\Lambda = \text{diag}(e^{-\beta\lambda_1}, e^{-\beta\lambda_2}, \ldots, e^{-\beta\lambda_n})$ and $\Phi = [\phi_1, \phi_2, \ldots, \phi_n]$. Thus

$$K_{\beta_{ij}} = \sum_{k=1}^{n} \phi_k(i)\phi_k(j)e^{-\lambda_k\beta}, \tag{9}$$

and $K_{\beta_{ij}} \in [0,1]$ is the (i,j) entry of a doubly stochastic matrix. Doubly stochasticity for all β implies *heat conservation* in the system as a whole, that is, not only in the nodes and edges of the graph but in the transitivity links eventually established between nonadjacent nodes: if i is not adjacent to j, eventually an entry $K_{\beta_{ij}} > 0$ will appear for a β high enough. The overall amount of heat is always $n = |V|$. The way this heat is diffused for $\beta \in [0, \beta_{\max}]$ is highly dependent on the structure of the graph. One of the features of our approach is that we assimilate the lack of *design* constraints into simplicity. For instance, a linear graph of n nodes is more complex than a complete graph since the number of diffusion constraints is higher, despite its high two-regularity. A higher number of diffusion constraints implies the creation of more transitivity links as β grows. We illustrate this process in Fig. 2 where we show the heat kernel for a linear graph of $n = 100$ nodes.

The latter example suggests a simple way of defining a complexity trace with similar qualitative behavior than the polytopal one. Such a new trace accounts for the total heat flowing through the graph at a given β *(instantaneous flow)*, which is given by

$$F_\beta(G) = \sum_{i=1}^{n}\sum_{j\neq i}^{n} \delta_{ij}\left(\sum_{k=1}^{n} \phi_k(i)\phi_k(j)e^{-\lambda_k\beta}\right), \tag{10}$$

where $\delta_{ij} = 1$ if $A_{ij} = 1$ and $\delta_{ij} = 0$ otherwise, that is, if $\delta_{ij} = 1$ if $(i,j) \in E$. A more compact definition of the flow is $F_\beta(G) = A : K_\beta$, where $X : Z = \sum_{ij} X_{ij}Z_{ij} = \text{trace}(XZ^T)$ is the Frobenius inner product. Instantaneous flow accounts only for the heat flowing through the edges of the graph, not for the heat remaining in the nodes nor for that encoding transitivity links. The limiting cases are $F_0 = 0$ and $F_{\beta_{\max}} = \frac{2|E|}{n}$.

Therefore the *heat flow complexity* $\mathcal{F}_\beta(G)$ is a β-dependent function simply defined as the instantaneous flow $F_\beta(G)$ although its edge-normalized version $\mathcal{F}_\beta(G) = n/(2|E|)F_\beta(G)$ can be used for the sake of achieving

$$\lim_{\beta\to 0} \mathcal{F}_\beta(G) = 0 \quad \text{and} \quad \lim_{\beta\to\infty} \mathcal{F}_\beta(G) = 1, \forall\, G. \tag{11}$$

Although heat flow complexity, which takes $O(n^2)$ for each β, is not normalized in the same terms as the polytopal complexity, its complexity trace is qualitatively similar. See Fig. 3 (center) for typical flow complexity traces corresponding to subgraphs of a gridlike graph.

C. Characterization of polytopal and flow complexity

Linking formally the complexity traces of both polytopal and heat flow complexities allows us not only to better understand both the heat flow process and matrix permanents but to justify the use of a simple structural complexity measure (heat flow trace) in practice. Given an

undirected graph $G = (V, E)$ with node set V ($|V| = n$) and edge set E, the existence of a phase-transition point (PTP) at a $\beta^+ \geq 0$ can be inferred from the analysis of the difference between the sum of off-diagonal elements of the diffusion kernel $K_\beta(G)$ and its trace. For $K_{\beta^+}(G)$ we have $\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_{\beta^+_{ij}} < \text{trace}(K_{\beta^+})$ whereas for $\beta > \beta^+$ we always obtain $\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_{\beta_{ij}} \geq \text{trace}(K_\beta)$. This means that at a PTP the quantity $\Xi_\beta = \text{trace}(K_\beta) - \sum_{i=1}^{n}\sum_{j\neq i}^{n} K_{\beta_{ij}}$ is still positive. At $K_0 = I_n$ we have $\Xi_0 = n$ and at $K_\infty = \mathcal{B}_*$ we obtain $\Xi_\infty = 1 - (n-1) = 2 - n$. Actually $2 - n$ may be reached as soon as the kernel converges to $\mathcal{B}_*$ (reaches the *equilibrium point*). Local maxima of Ξ_β are precluded by the monotonic nature of the diffusion process and therefore Ξ_β is a monotonically decreasing function with a minimum at equilibrium

The existence of a unique PTP is key to relating heat flow and maximal entropy, but it is not enough to enunciate a characterization principle for the change of phase which involves a dependency between heat flow and entropy. Such a principle is the phase transition principle.

Phase transition principle. Let $\beta^+ > 0$ define a PTP. Then, the heat flow $F_{\beta^+}(G)$ is maximal among all choices of β. This implies that if $K_{\beta^+}(G) = \sum_{\alpha=1}^{\gamma} p_\alpha P_\alpha$ is the maximum entropy BvN decomposition for β^+ then its entropy, i.e., $H_{\beta^+}(\mathcal{P})$ with $\mathcal{P} = \{p_1, \ldots, p_\gamma\}$, is maximal with respect to all maximum entropy decompositions in the range $[0, \beta_{\max}]$.

In other words, flow maximality occurs at β^+ as well as entropy maximality with respect to all BvN maximum entropy decompositions. For both $\beta < \beta^+$ and $\beta \geq \beta^+$ *flow maximality* relies on the elements of the kernel matrix associated with edge links (the first elements in the system receiving heat from the nodes). When $\beta < \beta^+$, elements associated with edge links must transfer heat to create transitivity links and consequently the sum of their values decreases with respect to that at β^+. We can write $\sum_{ij} \delta_{ij}K_{ij} = A : K$, where A is the adjacency matrix of G. It follows that $A : K_\beta < A : K_{\beta^+}$, that is, $F_{\beta^+} > F_\beta$. When $\beta > \beta^+$ the individual values of both diagonal and off-diagonal elements tend to $1/n$ as β increases. Elements associated with transitivity links (paths in the graph) are always smaller than or equal to those associated with edge links. If the equilibrium is reached later than β, only off-diagonal transitivity links increase. However, edge links decrease, which implies $A : K_{\beta^+} > A : K_\beta$, that is, $F_{\beta^+} > F_\beta$. If equilibrium is reached at β^+ we have $F_{\beta^+} = \frac{2|E|}{n}$.

Regarding the *implication between flow maximality and entropy maximality* at β^+, as entropy maximality is referred to maximum entropy decompositions in the range $[0, \beta_{\max}]$ we link the mechanisms driving a PTP with the dual formulation of finding the maximum entropy BvN decomposition. From [40] (Lemma 4) we have that the maximum entropy BvN decomposition of a DSM B is the solution to the left optimization problem below (primal), whose dual one is on the right:

$$\begin{array}{ll} \min & \sum_{\alpha\in\mathcal{S}_n} p_\alpha(\ln p_\alpha - 1) \qquad \max \quad B : Y - \sum_\alpha e^{Y:P_\alpha} \\[2ex] \text{such that} & \sum_\alpha p_\alpha P_\alpha \leq B \qquad\quad \text{such that} \quad \sum_\alpha e^{(Y:P_\alpha)}P_\alpha \leq B \\[2ex] & p_\alpha \geq 0 \qquad\qquad\qquad\qquad\quad 0 \geq Y \end{array} \tag{12}$$
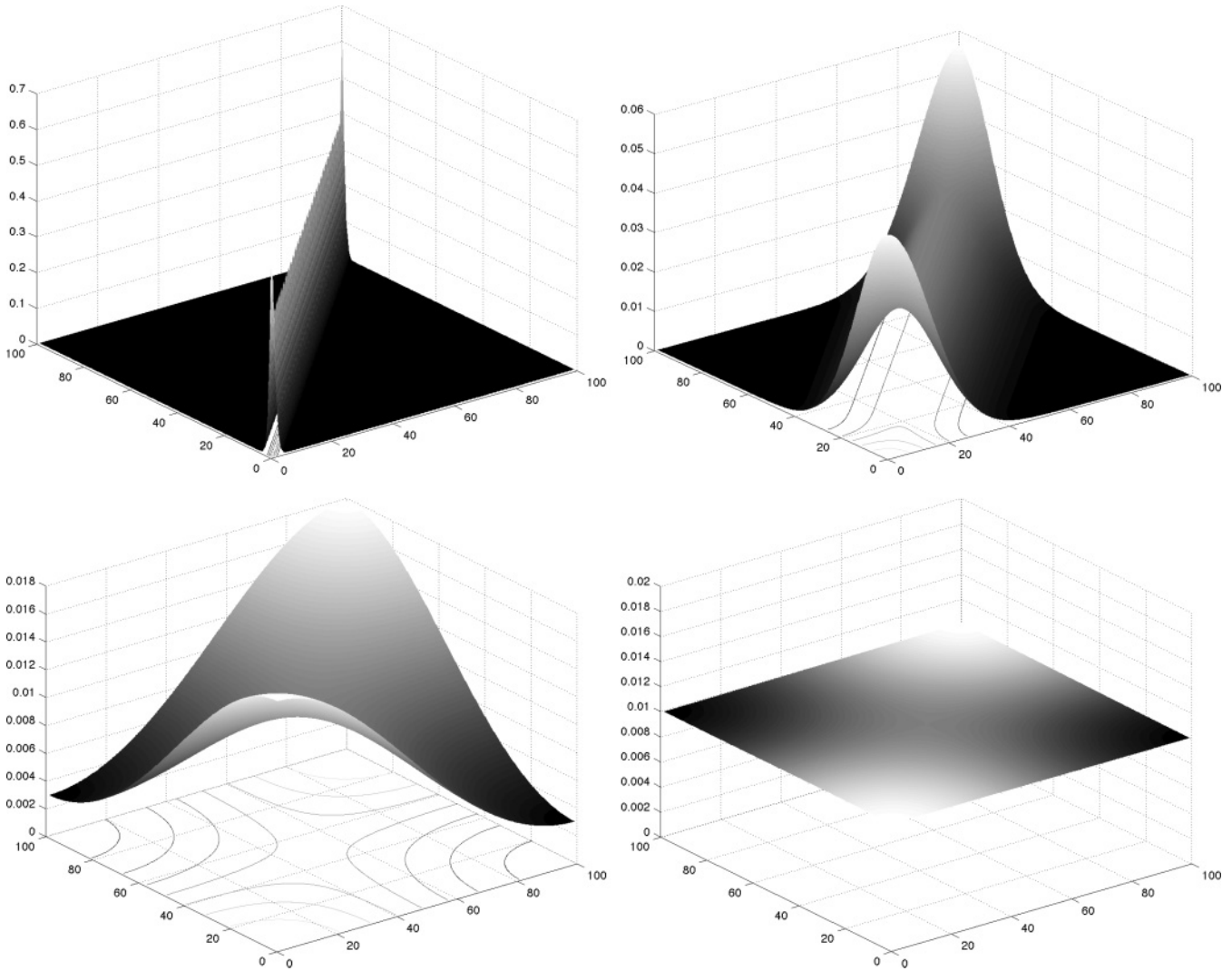
FIG. 2. Diffusion process in a linear graph with $n = 100$ nodes. The horizontal axes are assigned to nodes $1, \ldots, n$ (therefore the horizontal plane represents the adjacency matrix). Snapshots taken at $\beta = 0.64$, 100, 1000, and 6070 (top left to bottom right). The vertical axis represents the value of the kernel at a given point. In each case lighter gray means higher $K_{\beta_{ij}}$. In all cases the kernel overlays a contour plot. Extreme nodes (1 and 100) lose less heat than the rest of the nodes. As $\beta > 0.64$ increases, the closer the nodes are to the extremes the less heat they lose. The true value of $\beta_{\max}$ is 10 717, which gives an idea of the slow convergence of the diffusion process. In terms of flow we have $F_{\beta=0.64} = 43.274$, $F_{\beta=100} = 6.5185$, and $F_{\beta=1000} = 2.7483$. For $F_{\beta=6070} = 1.9849$ we are close to the equilibrium flow $[(2 \times 99)/100 = 1.9800]$.

where $\mathcal{S}_n$ is the set of permutations of $\{1, 2, \ldots, n\}$, and $Y \in R^{n \times n}$ is the matrix of Lagrange multipliers each corresponding to one constraint (component) in $B = \sum_\alpha p_\alpha P_\alpha$. Although the primal problem (left) is relaxed in the sense that equality in $\sum_\alpha p_\alpha P_\alpha \leqslant B$ is not required, the proof in Lemma 4 in [40] (see Appendix B.3) shows that it is impossible to have $\sum_\alpha p_\alpha P_\alpha < B$ for the optimal solution. Consequently, the relaxed primal problem shares its optima with the nonrelaxed one. Anyway both problems are #P hard since $\sum_\alpha e^{Y:P_\alpha} = \mathrm{per}(e^Y)$ (componentwise exponentiation). Therefore, the analysis of the second element of the phase transition principle (flow maximality–entropy maximality implication) relies on the analyis of how the Lagrange multipliers in Y interact with the the diffusion process encoded in $K_\beta = B$ at the optimal

solutions (maximum entropy decompositions). For instance, as in the dual problem we must maximize $K_\beta : Y - \sum_\alpha e^{Y:P_\alpha}$, we have that the optimal choice of the multipliers actually maximizes $K_\beta : Y - 1$. In addition, the relaxed formulation of the primal leads to bounding the optimal multipliers in the sense that $-n \ln n \leqslant K_\beta : Y \leqslant 0$ and $0 \geqslant Y_{ij} \geqslant \frac{-n \ln n}{k_{\min}}$ where $k_{\min} = \min_{ij} \{K_{\beta_{ij}}\}$. Therefore, we reduce our analysis to ensure that $-n \ln n \leqslant K_\beta : Y_\beta < K_{\beta^+} : Y_{\beta^+} \leqslant 0$ for $\beta \neq \beta^+$ in order to verify the flow maximality–entropy maximality implication. This rationale is driven by analyzing the possibility of assigning close-to-zero multipliers to the maximum number of kernel elements in order to maximize the Frobenius product. The number of available close-to-zero multipliers decreases as β increases, although this does not preclude having a maximal

$K_\beta : Y_\beta$ at β^+. The lower bound of $K_\beta : Y_\beta$ for a diffusion process occurs at $\beta = \beta_{\max}$, and it is $K_{\beta_{\max}} : Y_{\beta_{\max}} = -\ln n$.

V. HEAT FLOW–THERMODYNAMIC DEPTH COMPLEXITY

The application of the thermodynamic depth to characterize network complexity demands the formal specification of the microstates whose history leads to the macrostate (of the network). Here we define such microstates in terms of *expansion subgraphs*.

A. Node history, expansion subgraphs, and causal trajectories

Let $G = (V,E)$ with $|V| = n$. The first-order expansion subgraph of a given node $i \in V$ is given by i, its neighbors, and the edges connecting them. Then the *history of a node* $i \in V$ is $h_i(G) = \{e(i), e^2(i), \ldots, e^p(i)\}$ where $e(i) \subseteq G$ is the *first-order expansion subgraph* given by i and all $j \sim i$, $e^2(i) =$

$e(e(i)) \subseteq G$ is the *second-order expansion* consisting of $z \sim j : j \in V_{e(i)}$, $z \notin V_{e(i)}$, and so on until p cannot be increased. If G is connected $e^p(i) = G$, otherwise $e^p(i)$ is the connected component to which i belongs.

Every $h_i(G)$ defines a different causal trajectory leading to G itself, if it is connected, or to one of its connected components otherwise [see Fig. 3 (left) for a connected graph]. Thus, in terms of the TD the full graph G or the union of its connected components is the macrostate (macroscopic state). The *depth* of such a macrostate relies on the variability of the causal trajectories leading to it. The higher the variability, the more complex it is to explain how the macrostate is reached and the deeper is this state. Therefore, in order to characterize each trajectory we combine the heat flow complexities of its expansion subgraphs by means of defining *minimal enclosing Bregman balls* (MEBBs) [41]. Bregman divergences D_F define an asymmetric family of similarity measures, each one characterized by a strictly convex generator function $F : \mathcal{X} \to R^+$, where $\mathcal{X} \subseteq R^d$
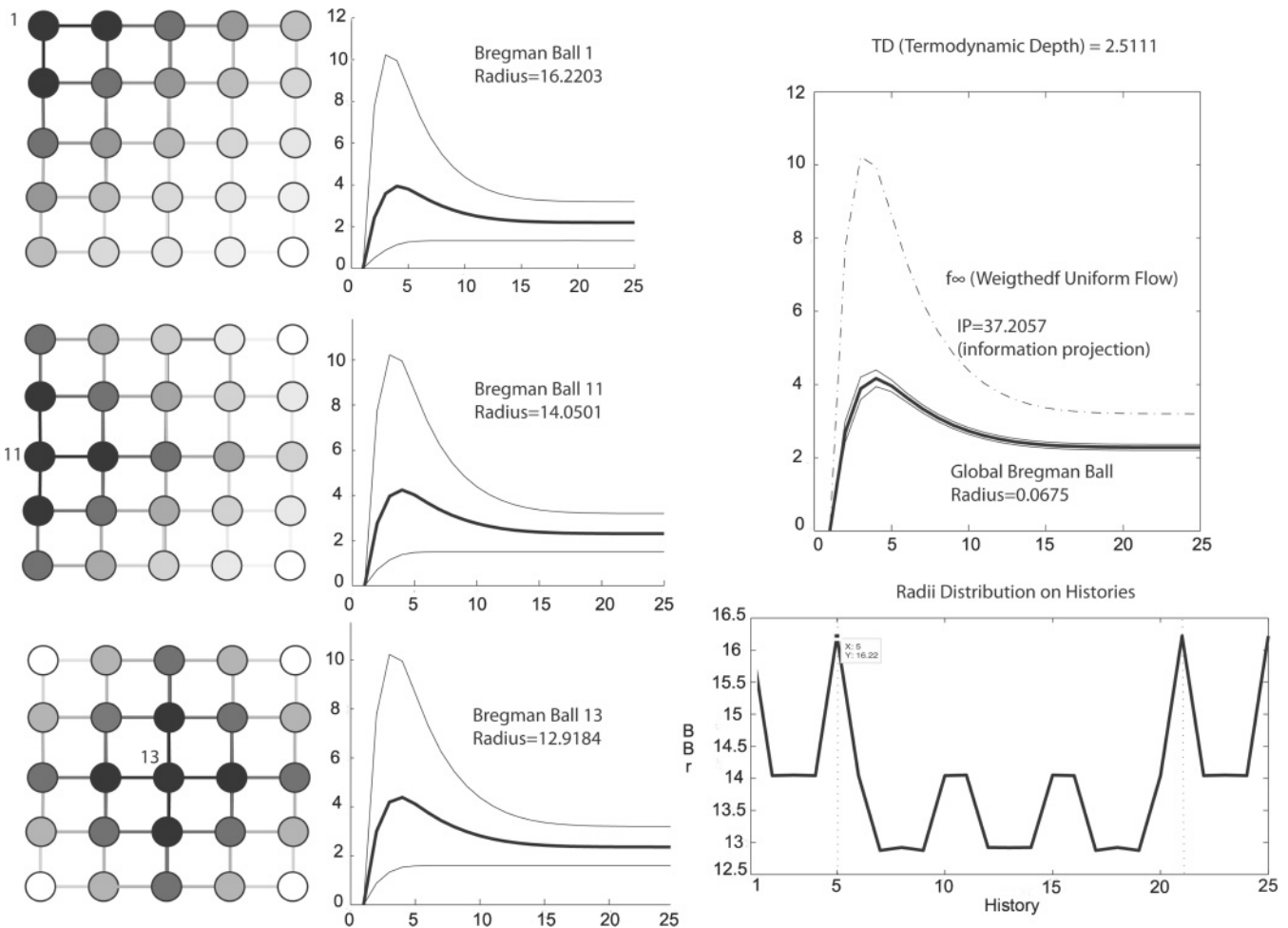

FIG. 3. Computing heat flow and thermodynamic depth for a four-connected grid graph of $n = 25$ nodes. Left column: expansion subgraphs and node histories at nodes 1 (top), 11 (middle), and 13 (bottom); in each case we show the first-order expansion subgraph in black, the next one in dark gray, and so on. Center column: Minimum enclosing Bregman balls for the histories of these nodes with their radii; in each case, the center of the ball (in bold) and the boundaries (given by the support vectors) are heat flow complexity traces inferred from the corresponding node history. Right column, top: projecting f^∞ (van der Waerden trace) on the Bregman ball given by the centers of all the $n = 25$ balls. Right column, bottom: plot of Bregman ball radii (BB r) corresponding to all node histories; symmetry is given by the structure and the maximal radii correspond to nodes 1, 5, 21, and 25 (the four corners).

is a convex domain and d the data dimension (in this case the number of discretized β diffusion times). Given two patterns (discretized functions in this case) $\vec{f}$ and $\vec{g}$, $D_F(\vec{f}||\vec{g}) = F(\vec{f}) - F(\vec{g}) - (\vec{f} - \vec{f})^T \vec{\nabla} F(\vec{f})$. Here, we use the I–Kullback-Leibler (I-KL) divergence $D_F(\vec{f}||\vec{g}) = \sum_{i=1}^{d} f_i \ln \frac{f_i}{g_i} - \sum_{i=1}^{d} f_i + \sum_{i=1}^{d} g_i$ with $F(\vec{f}) = \sum_{i=1}^{d} (f_i \ln f_i - f_i)$ (unnormalized Shannon entropy) which yields better results (more representative centroids of heat flow complexities) than other divergences or distortions like that of Itakura and Saito. When using the I-KL divergence in R^d, we have that $\vec{\nabla} F(f_i) = \ln f_i$ and also that $\vec{\nabla}^{-1} F(f_i) = e^{f_i}$. Using these formal ingredients we define the *causal trajectory* in terms of MEBBs.

Given $h_i(G)$, the heat flow complexity $\vec{f}_t = f(e^t(i))$ for the tth expansion of i, a generator F, and a Bregman divergence D_F, the *causal trajectory* leading to G (or one of its connected components) from i is characterized by the center $\vec{c}_i \in R^d$ and radius $r_i \in R$ of the MEBB $\mathcal{B}^{\vec{c}_i, r_i} = \{\vec{f}_t \in \mathcal{X} : D_F(\vec{c}_i||\vec{f}_t) \leqslant r_i\}$. Solving for the center and radius implies finding the $\vec{c}^*$ and r^* that minimize r subject to $D_F(\vec{c}_i||\vec{f}_t) \leqslant r \ \forall \ t \in \mathcal{X}$ with $|\mathcal{X}| = T$. Considering the Lagrange multipliers α_t, we have that $\vec{c}^* = \vec{\nabla}^{-1} F[\sum_{t=1}^{T} \alpha_t \vec{f}_t \vec{\nabla} F(\vec{f}_t)]$. The efficient algorithm in [41] estimates both the center and multipliers. This idea is closely related to core vector machines [42], and it is interesting to focus on the nonzero multipliers (and their support vectors) used to compute the optimal radius. More precisely, the multipliers define a convex combination and we have $\alpha_t \propto D_F(\vec{c}^*||\vec{f}_t)$, and the radius is simply chosen as $r^* = \max_{\alpha_t > 0} D_F(\vec{c}^*||\vec{f}_t)$. In Fig. 3 (center) we show some Bregman balls corresponding to different nodes of a four-connected grid graph.

B. Thermodynamic depth of a network

Given $G = (V, E)$, with $|V| = n$ and all the n pairs $(\vec{c}_i, r_i)$, the *heat flow–thermodynamic depth complexity* of G is characterized by the MEBB $\mathcal{B}^{\vec{c}, r} = \{\vec{c}_t \in \mathcal{X}_i : D_F(\vec{c}||\vec{c}_i) \leqslant r\}$ and $D_{\min} = \min_{f \in \mathcal{B}^{\vec{c}, r}} D_F(f^\infty||f)$, where $f^\infty = f(B_*) \in R^d$ is the van der Waerden complexity trace. As a result, the *TD depth of the network* is given by $\mathcal{D}(G) = r \times D_{\min}$.

The above definitions of complexity and depth are highly consistent with the summarizing of node histories to find a global causal trajectory which is as tightly bounded as possible. Here, r quantifies the *historical uncertainty*: the smaller r the simpler (shallower) is G. However, this is not sufficient for structures because many networks with quite different complexities may have the same value of r. Therefore, we define the depth of the network complementing randomness as suggested in the thermodynamic depth approach. In our case, the projection of f^∞ on the MEBB preserves the definition of entropy in terms of the distance to the uniform distribution [see Fig. 3 (left)]. The combinations or hierarchies of MEBBs have proved to be more effective than ball trees for nearest-neighbor retrieval [43]. In the computation of depths, the Legrendre duality (convex conjugate) is key because it establishes a one-to-one correspondence between the gradients $\vec{\nabla} F$ and $\vec{\nabla} F^{-1}$ due to the convexity of F. Therefore, the Bregman projection f of f^∞ on the the border of $\mathcal{B}^{\vec{c}, r}$ lies on the curve $f_\theta^{-1} = \theta \vec{\nabla} F(\vec{c}) + (1 - \theta) \vec{\nabla} F(f^\infty)$ with $\theta \in [0, 1]$

and $f_\theta = \vec{\nabla}^{-1} F(f_\theta^{-1})$. The projection f be easily found (approximately) through bisection search on θ.

Although we use the heat flow complexity trace for building the Bregman balls, the thermodynamic depth approach can be applied to any structural complexity measure (we will compare some of them in the experimental section) provided that we redefine f^∞ properly. One key ingredient of our proposal, which is especially useful when the basic complexity measure (e.g., heat flow trace) is based on spectral-graph theory, is the way we build node histories. As node histories rely on extending subgraphs and then computing the basic measure, if such a measure is spectral based this precludes many problems derived from isospectrality. In addition, as only the variability of the centers is considered, the thermodynamic depth approach introduces partial graph size independence (graphs with more nodes than others do not have necessarily higher complexity).

VI. COMPLEXITY OF PROTEIN-PROTEIN INTERACTION NETWORKS

A. Experimental setup

We have designed an experiment using PPIs extracted from STRING 8.2 [44]. This tool allows the user to select a protein and then a species (typically bacteria). In our experiments we are interested in both histidine kinase, a key protein in the development of signal transduction, and bacteria. Given the histidine kinase and the bacteria species we can select one of the proteins of the same family (query protein) and then STRING composes a PPI network centered in that protein. We consider only links based on physical and functional interactions. Other settings are the selection of high-confidence interactions ($\geqslant 0.7$), a maximum of 50 interactions per node, and a network depth of 5. The default parameters are (a) medium confidence ($\geqslant 0.4$), (b) a limit of 10 interactions (degree), and (c) network depth 1. These parameters yield networks that are too small for significant complexity comparisons, and this problem is exacerbated when we increase the confidence parameter. Increasing network depth, on the other hand, allows us to obtain more neighbors with a confidence above the threshold. Thus, in our experimental setting, we operate with the confidence close to its maximum value ($\geqslant 0.7$ vs $\geqslant 0.9$) in order to capture useful internode relations. We also set the maximum number of interactions to be 50 and the maximum depth to be 5.

In a first test, we consider PPIs related to histidine kinase corresponding to ten species belonging to ten phyla of bacteria. We select subjectively three PPIs (simple, complex, and more complex) from each species and compute their TDs. In 70% of the cases, the TD matches intuition. When comparing with Estrada's early spectral homogeneity descriptor [45], we also find that the ratio between intraclass and interclass variability is slightly better (smaller) for the TD (0.6840 vs 0.7383).

Our main experiment consists of analyzing 217 PPIs, related to histidine kinase, from six different groups (all the PPIs in the same group correspond to the same species) with the following evolutionary order (from older to more recent): *Aquifex*, 4 PPIs from *Aquifex aeloicus*; *Thermotoga*, 4 PPIs from *Thermotoga maritima*; Gram positive, 52 PPIs

from *Staphylococcus aureus*; *Cyanobacteria*, 73 PPIs from *Anabaena variabilis*; and *Proteobacteria*, 40 PPIs from *Acidovorax avenae*. There is an additional class (*Acidobacteria*, 46 PPIs) which has been more controversial in terms of bacterial evolution since their discovery [46]. There are studies which relate many of them to different subphyla of *Proteobacteria* (see [47]) and also to *Actinobacteria* [48]. However, the *Candidatus Koribacter* genus of *Acidobacteria* is not included in the latter classifications [47]. One of them is the *Candidatus Koribacter versatilis Ellin345*, despite being Gram negative like some *Proteobacteria*. In addition, *Ellin345* has been recently placed very early in the phylogenic tree [49]. In Table I we show the origin of all the 217 PPIs analyzed in this experiment, including the ones related to *Ellin345*. In Fig. 4 we show one example of PPI corresponding to each of the six species with their complexities obtained with our method.

B. Experimental results

In order to provide a fair comparison of our method (heat flow and thermodynamic depth) with previous measures of structural complexity we have chosen three spectral methods from the literature (Estrada's network heterogeneity index [24], Laplacian energy [23], and von Neumann entropy [5,6]). We have also embedded them in the thermodynamic depth approach. We choose f^∞ such that for each of the measures $f^\infty = f(B_*)$, to compute a complexity trace from the van der Waerden matrix. Unfortunately, this makes no sense in the case of the Estrada heterogeneity index since f^∞ is always zero. In this case we have chosen $f^\infty = f(G)$, that is, the complexity of the macrostate. In the case of both the Laplacian energy and the von Neumann entropy we have chosen $f^\infty = f(C_n)$, that is, the corresponding measure of the complete graph, since both quantities rely on the Laplacian spectrum. The difference between this choice and $f^\infty = f(B_*)$ is simply a scale factor since all of the nonzero Laplacian eigenvalues are equal for a complete graph (uniformly attributed or nonattributed). However, we prefer to take $f^\infty = f(C_n)$ because both measures (the Laplacian energy and the von Neumann entropy) rely on the volume of the graph $2|E|$, and we are working with nonattributed graphs.

Before addressing the main experimental evaluation in this paper, we provided the results of a toy experiment for the purposes of illustration. To this end we analyze the graphs shown in Fig. 1 using the complexity measures discussed above and incorporated into the proposed TD framework. We commence by reminding the reader of the mathematical definition of each measure studied. Given a graph $G = (V, E)$, the Estrada heterogeneity (EHN) index is formally defined as $I_{\text{EHN}}(G) = \sum_{i \sim j} (d_i^{-1/2} - d_j^{-1/2})^2$, where d_i and d_j are the degrees of the adjacent vertices i and j. The Laplacian energy (LEN) is defined as $I_{\text{LEN}}(G) = \sum_{i=1}^{n} |\lambda_i - 2|E|/n|$ where $n = |V|$ and λ_i are the eigenvalues of the graph Laplacian. Similarly, the von Neumann entropy (VNE) is $I_{\text{VNE}}(G) = -\sum_{i=1}^{n} \frac{\lambda_i}{2|E|} \log_2 \frac{\lambda_i}{2|E|}$ with the same eigenvalues as $I_{\text{LEN}}(G)$. In Table II, where the values inconsistent with regularity are shown in bold, we show the TD complexities for each complexity measure on each graph. The table shows

that the heat diffusion is consistent with regularity. However the EHN index is consistent only at high complexities, and underestimates in the mid to high and mid to low complexity ranges. Both the LEN and VNE indices are coherent neither at high complexity (which is underestimated) nor at mid to low complexity (which is overestimated). The EHN descriptor is highly dependent on the number of edges and this usually produces a large variability (within the TD process) for highly connected networks, such as Grid8N10 which has mid to high complexity. The underestimation of the complexity of Line10 is due to the fact that almost all histories have similar EHN variability, which implies that the final TD variability is low. Both the LEN and VNE measures rely on deviations either from the average degree $2|E|/n$ (LEN) or the volume $2|E|$ (VNE) of the graph. The eigenvalues of the Laplacian satisfy the trace condition $\sum_{i=1}^{n} \lambda_i = 2|E|$. As a result both the LEN and VNE indices depend on the global network parameter ($2|E|$), which is not always correlated with regularity. This limits their ability to capture finer variations in complexity. For instance, the VNE measure returns the same complexity for regular and complete graphs only for large n, while the LEN index returns the graph energy $\sum_{i=1}^{n} |\lambda_i|$ for regular graphs. This lack of flexibility is also observed in the experimental results for PPIs reported later in this paper.

As we have different TD complexities for each method, histogramming reveals typically long tailed distributions with most of the TDs concentrated at a given point. *Are these points ordered according to the evolutive order?* This question can be answered by studying the cumulative distributions instead of the probability density functions (Fig. 5 where the abscissa bounds are set in order to improve the results of each method). In this case, reaching the top (cumulative = 1) soon indicates low TD whereas reaching it later indicates high TD. This rationale motivates the measurement of the area under the curve (AUC) for each of them: the greater the AUC the (statistically) simpler the PPIs. We show these AUCs in Table III where incorrect (not coherent with evolution) figures are shown in bold type. The joint analysis of both the cumulatives and their AUCs shows that the TD–heat diffusion measure yields the correct results; it is followed by TD–Estrada's heterogeneity which underestimates the complexity of both the Gram-possitive and *Acidobacteria* species. The TD–Laplacian energy measure overestimates the complexity of *Thermotoga*, *Aquifex,* and *Proteobacteria*, whereas it underestimates that of *Acidobacteria*. Finally, the TD–von Neumann entropy measure overestimates the complexity of *Thermotoga* and *Aquifex* and underestimates that of *Acidobacteria*.

The results obtained above are consistent with the degree dependency between network size and complexity corresponding to each method. As we show in Fig. 6, the two best methods (heat flow and heterogeneity) are quite independent of size, whereas the others (Laplacian energy and von Neumann entropy) are high dependent on the size of the network. As we show in Table I, most of the PPI classes are quite heterogeneous (high deviation), and there are networks which are smaller than others but more complex, and vice versa (see Fig. 4).

Thus, we can conclude that TD combined with heat flow is a good tool in principle for analyzing the complexity of networks.

TABLE I. Table with the bacterial species studied in our main experiment (first column). We include the mean and deviation of the PPI size (second column) and all the query proteins used (their respective numbers of nodes are inside parentheses) for building PPIs with the settings described in text.

Species	Size[a]	Proteins[b]
Aquifex aeolicus	204.2500 ± 2.7538	TM_0127(201) TM_0400(203) TM_0853(207) TM_1258(206)
Thermotoga maritima	206 ± 3.1623	hksP1 (210) hksP2 (207) hksP3 (203) hksP4 (204)
Staphilococcus aureus	75.6346 ± 67.1203	MW0199 (48) NWMN_1327 (200) NWMN_1741 (54) NWMN_2523 (14) SA0216 (39) SAB0162c (46) SAB1782c (56) SAB2499c (50) SACOL0202 (50) SACOL1739 (201) SACOL1906 (54) SACOL2645 (15) SAOUHSC_00185 (48) SAOUHSC_01799 (109) SAOUHSC_01981 (40) SAOUHSC_02099 (202) SAOUHSC_02955 (14) SAR0215 (43) SAR1940 (54) SAS0199 (49) SAS1770 (34) SAUSA300_0218 (45) SAUSA300_1799 (52) SAUSA300_2558 (14) SAV0224 (35) SAV1321 (204) SAV1849 (205) SAV2624 (205) SaurJH1_0214 (46) SaurJH1_1937 (42) SaurJH1_1973 (186) SaurJH9_0058 (31) SaurJH9_0208 (30) SaurJH9_0245 (16) SaurJH9_0729 (32) SaurJH9_1475 (32) SaurJH9_1749 (37) SaurJH9_1903 (17) SaurJH9_1939 (32) SaurJH9_2115 (22) SaurJH9_2647 (18) arlS (109) graS (32) hssS (109) kdpD (44) lytS (18) nreB (137) phoR (200) saeS (112) tycG (206) vraS (202) yhcS (43)
Anabena variabilis	75.4521 ± 68.4091	Ava_0024 (66) Ava_0055 (238) Ava_0062 (224) Ava_0064 (20) Ava_0066 (33) Ava_0066 (33) Ava_0355 (16) Ava_0413 (66) Ava_0505 (224) Ava_0521 (66) Ava_0612 (31) Ava_0647 (211) Ava_0792 (54) Ava_0799 (153) Ava_1003 (13) Ava_1005 (13) Ava_1149 (13) Ava_1149 (13) Ava_1168 (31) Ava_1175 (207) Ava_1191 (101) Ava_1210 (66) Ava_1486 (16) Ava_1559 (13) Ava_1628 (17) Ava_1719 (16) Ava_1954 (66) Ava_2027 (55) Ava_2149 (66) Ava_2152 (66) Ava_2176 (66) Ava_2328 (24) Ava_2420 (17) Ava_2439 (214) Ava_2466 (80) Ava_2524 (167) Ava_2562 (52) Ava_2563 (36) Ava_3003 (9) Ava_3004 (13) Ava_3207 (66) Ava_3226 (202) Ava_3368 (60) Ava_3467 (204) Ava_3526 (66) Ava_3721 (66) Ava_3779 (224) Ava_3850 (47) Ava_3852 (41) Ava_3854 (35) Ava_3865 (31) Ava_4086 (31) Ava_4135 (43) Ava_4136 (66) Ava_4267 (66) Ava_4305 (66) Ava_4325 (13) Ava_4326 (17) Ava_4343 (13) Ava_4345 (13) Ava_4401 (66) Ava_4432 (44) Ava_4433 (66) Ava_4457 (202) Ava_4696 (36) Ava_4723 (208) Ava_4723 (40) Ava_4783 (40) Ava_4885 (66) Ava_4928 (205) Ava_B0028 (37) Ava_B0190 (101) Ava_B01974 (101) Ava_B0208 (55) Ava_C0117 (32)
Acidovorax avenae	132.7556 ± 85.4209	Aave_0042 (161) Aave_0173 (126) Aave_0317 (6) Aave_0867 (202) Aave_0874 (204) Aave_0905 (205) Aave_1122 (8) Aave_1444 (208) Aave_1494 (9) Aave_1948 (211) Aave_2036 (56) Aave_2118 (201) Aave_2218 (76) Aave_2236 (231) Aave_2267 (203) Aave_2461 (207) Aave_2525 (27) Aave_2535 (25) Aave_2632 (205) Aave_2974 (232) Aave_2975 (155) Aave_2976 (155) Aave_3001 (149) Aave_3093 (230) Aave_3275 (161) Aave_3541 (203) Aave_3564 (26) Aave_3778 (3) Aave_3863 (8) Aave_3993 (204) Aave_3996 (204) Aave_4137 (202) Aave_4230 (8) Aave_4320 (21) Aave_4378 (220) Aave_4381 (8) Aave_4526 (209) Aave_4637 (161) Aave_4654 (201) Aave_4732 (206)
Acidobacterium sp. (*Ellin345*)	84.5435 ± 73.6875	Acid345_0310 (118) Acid345_0375 (150) Acid345_0419 (20) Acid345_0474 (136) Acid345_0507 (6) Acid345_0508 (6) Acid345_0821 (48) Acid345_0843 (20) Acid345_0938 (8) Acid345_0973 (77) Acid345_1393 (203) Acid345_1524 (215) Acid345_1590 (202) Acid345_1659 (29) Acid345_1692 (90) Acid345_1733 (24) Acid345_1771 (209) Acid345_1812 (8) Acid345_1814 (8) Acid345_2403 (30) Acid345_2411 (106) Acid345_2412 (140) Acid345_2442 (39) Acid345_2554 (25) Acid345_2603 (53) Acid345_2781 (140) Acid345_2895 (185) Acid345_2989 (210) Acid345_3062 (201) Acid345_3087 (77) Acid345_3100 (29) Acid345_3292 (20) Acid345_3488 (20) Acid345_3503 (104) Acid345_3712 (128) Acid345_3715 (20) Acid345_3749 (40) Acid345_3837 (20) Acid345_3845 (7) Acid345_4032 (210) Acid345_4055 (91) Acid345_4063 (120) Acid345_4169 (206) Acid345_4289 (77) Acid345_4461 (7) Acid345_4561 (7)

[a]$\mu \pm \sigma$.
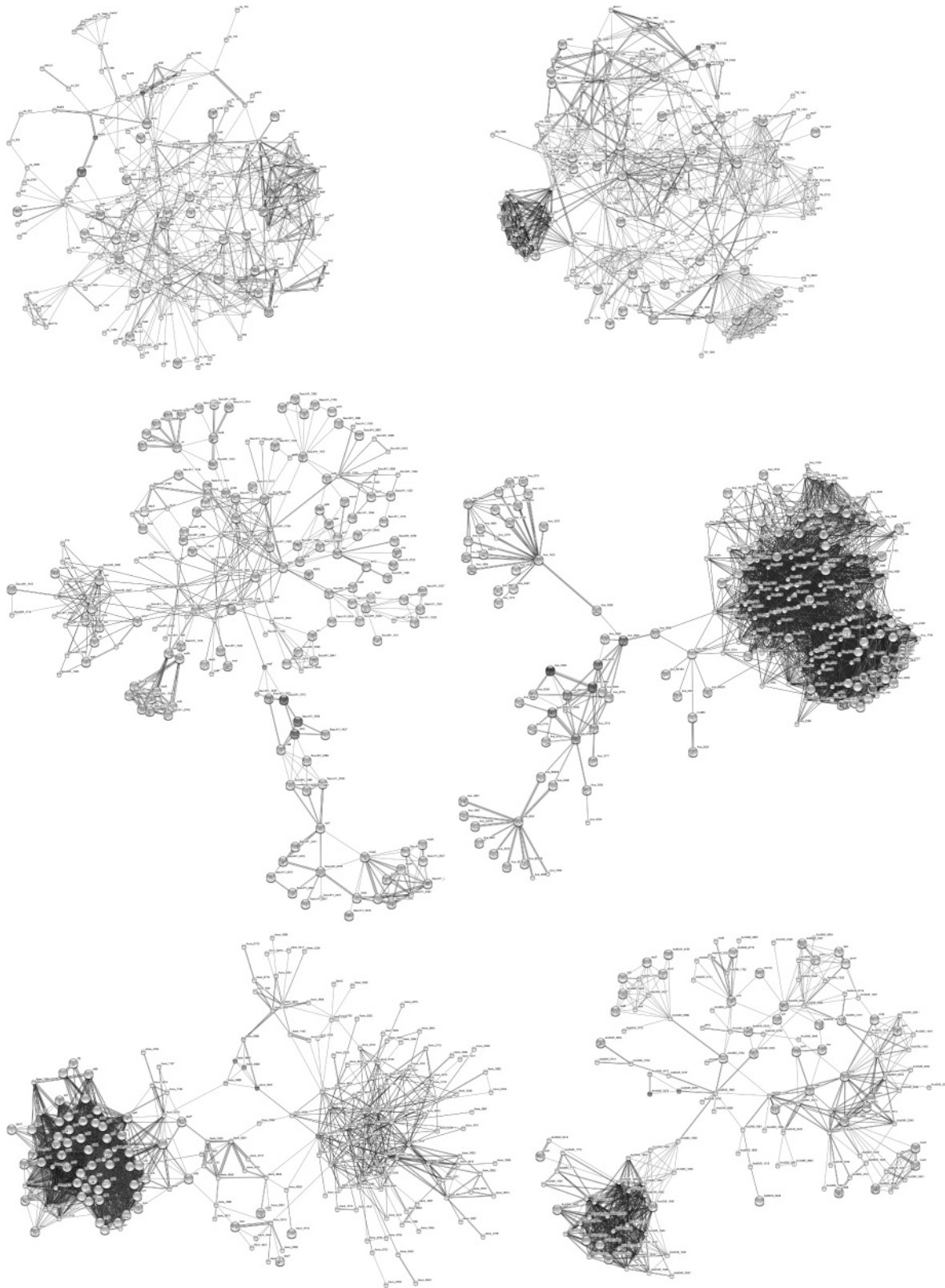
[b]With $|V|$ in parentheses.

FIG. 4. PPI examples for protein histidine kinase. From top left to bottom right: *Aquifex* (hksP2, 207 nodes, TD = 57.8926), *Thermotoga* (TM_0127, 201 nodes, TD = 69.2323), Gram positive (SaurJH1_1973, 186 nodes, TD = 85.6689), *Cyanobacteria* (Ava_0062, 224 nodes, TD = 4.6383), *Proteobacteria* (Aave_0867, 202 nodes, TD = 58.3774), and *Acidobacteria* (Acid345_0474, 136 nodes, TD = 618.1457).
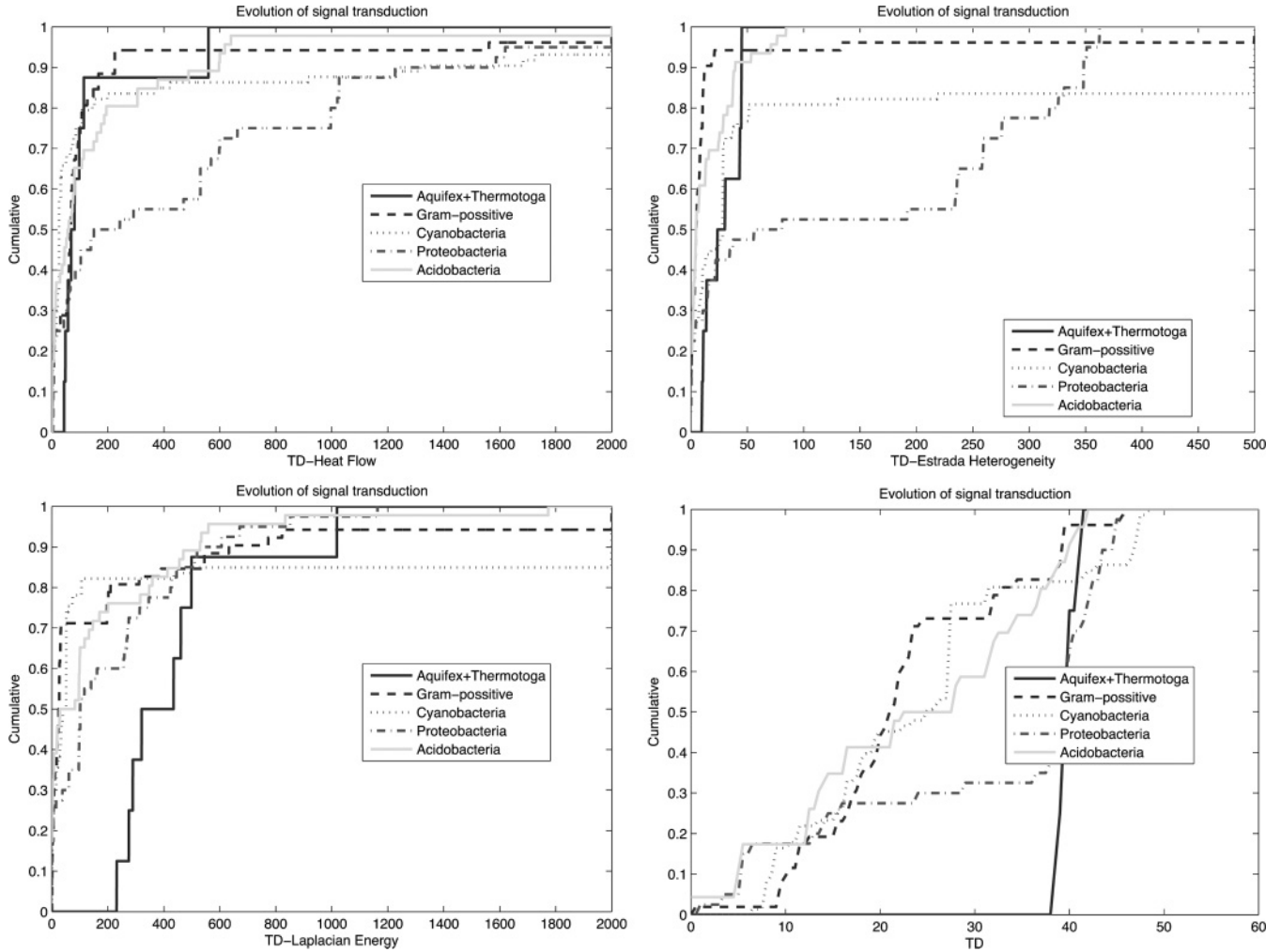
FIG. 5. Cumulatives for heat flow (top left), Estrada's heterogeneity (top right), Laplacian energy (bottom left), and von Neumann entropy (bottom right).

VII. CONCLUSIONS

In this paper, we have proposed and successfully tested a measure of graph complexity: the heat flow–thermodynamic depth measure. We have characterized formally its link with Birkhoff polytopes and polytopal complexity, also defined in this paper. The link is given by the assimilation of high-entropic decompositions to maximal flow but it has been derived through exploring the connections between

polytopal complexity and matrix permanents, and also by analyzing the dynamics of heat diffusion in graphs. In addition we have also proposed a thermodynamic depth theoretical framework for embedding any structural complexity measure. This framework reduces the risk of isospectrality when measures inspired in spectral-graph theory are used. We compare our method with three other methods in the literature: Estrada's heterogeneity, the Laplacian energy, and the von

TABLE II. Comparison of complexity measures after use of TD.

Graph	Heat[a]	EHN[b]	LEN[c]	VNE[d]
Gauss10	12.8965	6.3599	**0.0777**	**0.0016**
Grid4N10	0.6983	**5.83×10^{-5}**	0.1102	0.0103
Grid8N10	0.5054	0.0011	0.0123	0.0034
Line10	0.1509	**0.0000**	**0.1016**	4.2623
Circle10	0.0000	0.0000	0.0000	0.0000

[a]Heat flow.
[b]Estrada's heterogeneity.
[c]Laplacian energy.
[d]Von Neumann entropy.

TABLE III. Table of AUCs.

Bacteria	Heat[a]	EHN[b]	LEN[c]	VNE[d]
Aquifex and *Thermotoga*	93.29%	94.44%	**77.93%**	**34.09%**
Gram positive	91.33%	**94.61%**	88.85%	62.00%
Cyanobacteria	86.42%	79.94%	83.00%	59.49%
Proteobacteria	75.70%	70.62%	**88.86%**	48.20%
Acidobacteria	91.39%	**96.77%**	91.10%	**61.00%**

[a]Heat flow.
[b]Estrada's heterogeneity.
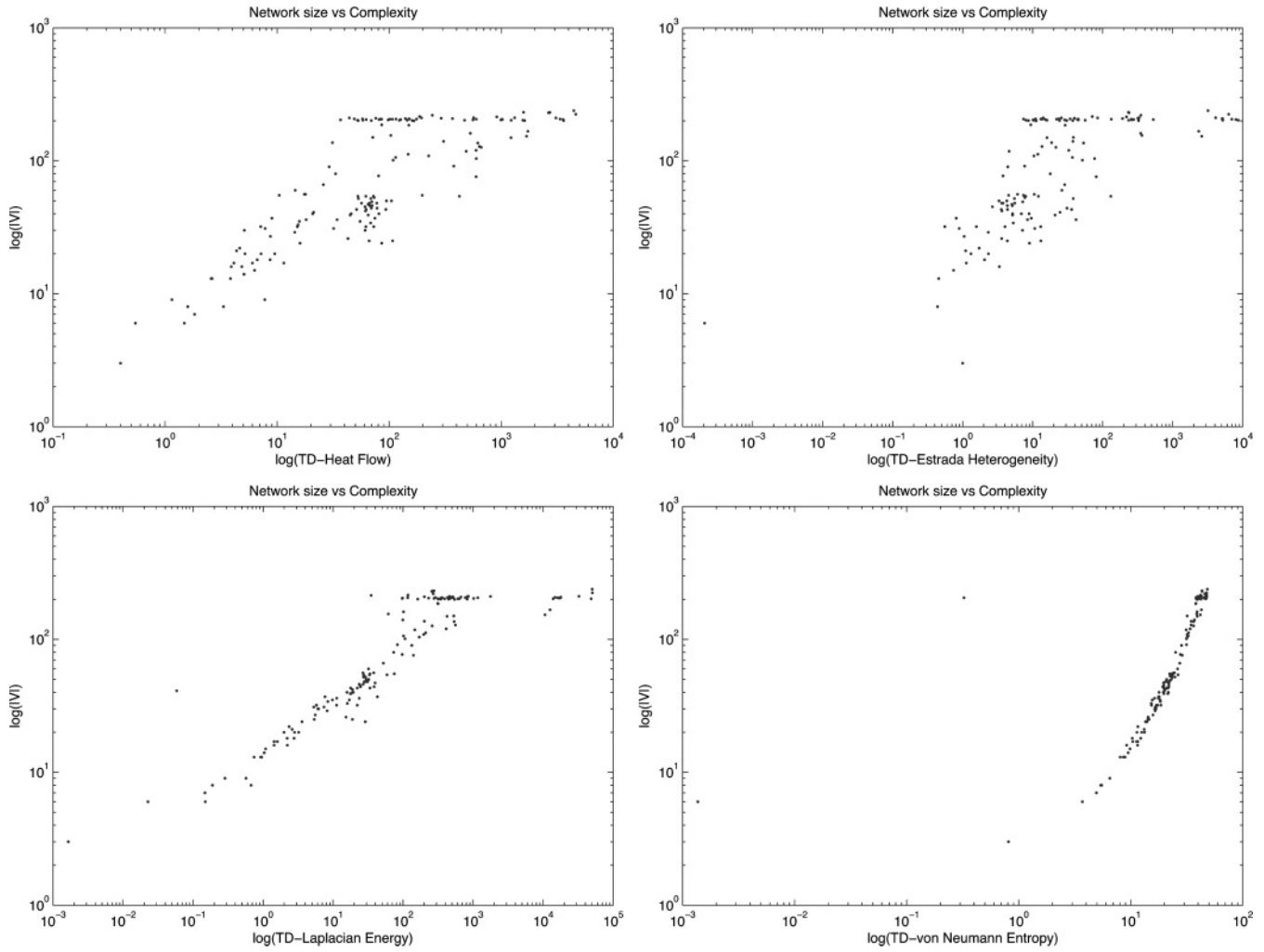[c]Laplacian energy.
[d]Von Neumann entropy.

FIG. 6. Network size vs complexity ln-ln plots for heat flow (top left), Estrada's heterogeneity (top right), Laplacian energy (bottom left), and von Neumann entropy (bottom right). In all cases the natural logarithm is used.

Neumann entropy. These methods are also embedded in our thermodynamic depth framework. Our experiments are addressed to analysis of 217 PPI networks corresponding to several phyla of bacteria. We find that more evolved species have (statistically) more complex structure. Both heat flow and Estrada's heterogeneity are proved to be quite independent of network size. Given the relatively good results obtained with Estrada's heterogeneity, future work will include a more in-depth analysis of the formal connections between the two methods (Estrada's heterogeneity relies on a quadratic form of the Laplacian as do the eigenvectors driving heat flow). Finally, since obtaining both the flow and the node histories are highly

parallel processes, we will also investigate estimating the NC computational complexity for the proposed method.

[1] *Network Science: Complexity in Nature and Technology*, edited by E. Estrada, N. Fox, and G.-L. O. (Springer, London, 2010).

[2] A. Torsello and E. Hancock, IEEE TPAMI **28**, 954 (2006).

[3] J. Körner, in *Transactions of the Sixth Prague Conference on Information Theory* (Academia, Prague, 1973), pp. 411–250.

[4] G. Simonyi, in *Perfect Graphs* (John Wiley and Sons, New York, 2001), pp. 293–328.

[5] F. Passerini and S. Severini, e-print arXiv:0812.2597v1 [cond-mat.dis-nn].

[6] K. Anand, G. Bianconi, and S. Severini, e-print arXiv:1011.1565v2.

[7] S. Braunstein, S. Ghosh, and S. Severini, Ann. Combinatorics **10**, 291 (2006).

[8] N. Biggs, *Algebraic Graph Theory*, Cambridge Tracts in Mathematics No. 67 (Cambridge University Press, Cambridge, 1974).

[9] D. Bonchev, *Information Theoretic Indices for Characterization of Chemical Structures* (Research Studies Press, Chichester, 1983).

[10] D. Bonchev, *Complexity in Chemistry. Introduction and Fundamentals* (Taylor and Francis, London, 2003).

[11] M. Dehmer, Appl. Math. Comput. **201**, 82 (2008).

[12] J. Claussen, Physica A **375**, 365 (2007).

[13] D. Feldman and J. Crutchfield, Phys. Lett. A **238**, 244 (1998).

[14] C. H. Bennett, Found. Phys. **16**, 585 (1986).

[15] A. N. Kolmogorov, Prob. Peredachi Inf. **1**, 3 (1965).

[16] G. Chaitin, J. Assoc. Comput. Mach. **13**, 547 (1965).

[17] Pudlák and V. Rödl, Combinatorica **12**, 221 (1992).

[18] S. Jukna, Combinatorics, Probab. Comput. **15**, 855 (2006).

[19] D. Neal and M. Orrison, Electron. J. Combinatorics **15**, R9 (2006).

[20] M. Gell-Mann, Complexity **1**, 16 (1995).

[21] M. Gell-Mann and S. Lloyd, in *Non-extensive Entropy: Interdisciplinary Applications* (2003), pp. 387–425.

[22] Y.-Z. Song, P. Arbelaez, P. M. Hall, C. Li, and A. Balikai, in *Computer Vision–ECCV 2010, 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5–11, 2010, Proceedings, Part IV*, edited by Kostas Daniilidis, Petros Maragos, and Nikos Paragios, Vol. 6314 (Springer, 2010), pp. 694–707.

[23] I. Gutman and B. Zhou, Lin. Alg. Appl. **414**, 29 (2006).

[24] E. Estrada, Phys. Rev. E **82**, 066102 (2010).

[25] S. Lloyd and H. Pagels, Ann. Phys. **188**, 186 (1988).

[26] B. Machta and J. Machta, Phys. Rev. E **71**, 026704 (2005).

[27] J. Machta, Complexity J. **5**, 46 (2006).

[28] J. P. Crutchfield and C. R. Shalizi, Phys. Rev. E **59**, 275 (1999).

[29] G. D. Birkhoff, Univ. Nacional Tucuman Rev., Ser. A **5**, 147 (1946).

[30] R. I. Kondor and J. D. Lafferty, in *Machine Learning, Proceedings of the Nineteenth International Conference (ICML 2002),* *University of New South Wales, Sydney, Australia, July 8–12, 2002*, edited by C. Sammut and A. G. Hoffmann (Morgan Kaufmann, 2002), pp. 315–322.

[31] F. Escolano, E. R. Hancock, and M. A. Lozano, in *19th International Conference on Pattern Recognition (ICPR 2008), December 8–11, 2008, Tampa, Florida, USA* (IEEE, 2008), pp. 1–5.

[32] F. Escolano, E. Hancock, and M. Lozano, in *SSPR/SPR* (2008), pp. 237–246.

[33] H. Minc, in *Encyclopedia of Mathematics and its Applications*, Vol. 6 (Addison-Wesley, Reading, MA, 1978).

[34] G. Egorychev, Adv. Math. **42**, 299 (1981).

[35] L. G. Valiant, Theor. Comput. Sci. **8**, 189 (1979).

[36] K. Mulmuley, J. ACM **58**, 5 (2011).

[37] M. Jerrum, A. Sinclair, and E. Vigoda, J. Assoc. Comput. Mach. **51**, 671 (2004).

[38] P. B. Slater, Environ. Planning **21**, 1541 (1989).

[39] H. Qiu and E. Hancock, Pattern Recognit. **40**, 2874 (2007).

[40] S. Agrawal, Z. Wang, and Y. Ye, in *WINE*, edited by C. Papadimitriou and S. Zhang (Springer-Verlag, Berlin, Heidelberg, 2008), pp. 126–137.

[41] R. Nock and F. Nielsen, in *ECML*, edited by J. Gama, R. Camacho, P. Brazdil, A. Jorge, and L. Torgo (Springer, 2005), pp. 649–656.

[42] I. W. Tsang, A. Kocsor, and J. T. Kwok, in *ICML*, edited by Z. Ghahramani (ACM, 2007), pp. 911–918.

[43] L. Cayton, in *ICML*, edited by W. W. Cohen, A. McCallum, and S. T. Roweis (ACM, 2008), pp. 112–119.

[44] L. Jensen, M. Kuhn, M. Stark, S. Chaffron, C. Creevey, J. Muller, T. Doerks, P. Mullien, A. Roth, M. Simonovic, P. Bork, and C. von Mering, Nucleic Acids Res. **37**, D412 (2009).

[45] E. Estrada, Europhys. Lett. **73**, 649 (2006).

[46] S. B. C.R. Kuske and J. Bush, Appl. Environ. Microbiol. **63**, 3614 (1997).

[47] P. H. M. Sait and P. Janssen, Environ. Microbiol. **4**, 654 (2002).

[48] F. Ciccarelli, F. Doerks, C. von Mering, C. Creevey, B. Snell, and P. Bork, Science **311**, 1283 (2006).

[49] M. Wu and J. Eisen, Genome Biol. **9**, R151 (2008).