What Is the Complexity of a Network? The Heat Flow-Thermodynamic Depth Approach*

Francisco Escolano[1], Miguel A. Lozano[1],
Edwin R. Hancock[2], and Daniela Giorgi[3]

[1] University of Alicante
{sco,malozano}@dccia.ua.es
[2] University of York
erh@cs.york.ac.uk
[3] IMATI-CNR Genova
daniela@ge.imati.cnr.it

Abstract. In this paper we establish a formal link between network complexity in terms of Birkhoff-von Neumann decompositions and heat flow complexity (in terms of quantifying the heat flowing through the network at a given inverse temperature). We propose and proof characterization theorems and also two fluctuation laws for sets of networks. Such laws emerge from studying the implicacions of the Fluctuation Theorem in heat-flow characterization. Furthermore, we also define heat flow complexity in terms of thermodynamic depth, which results in a novel approach for characterizing networks and quantify their complexity In our experiments we characterize several protein-protein interaction (PPI) networks and then highlight their evolutive differences, in order to test the consistence of the proposed complexity measure in terms of the second law of thermodynamics.

1 Introduction

The quantification of the *intrinsic complexity* of networks has attracted significant attention, in a number of fields including complexity science, pattern recognition and machine learning, due to its fundamental practical importance. Some complexity characterizations rely on spectral graph theory (see [1] for applications in computational biology, [2] for biological, social and other kinds of networks, and [3][4] for applications to pattern recognition). The work presented herein concerns the global analysis of structural patterns but not their fine discriminability. For instance, two undirected complete graphs (the simplest ones according to our approach) of very different size should have a similar complexity; however in terms of their discrimination, they will be different for an inexact graph matching strategy. However, complexity can be used as a MDL-principled measure for graph learning. In addition, fine discrimiability methods like matching are not suitable for finding global characterizations of structural patterns

* Funded by the Project TIN2008- 04416/TIN of the Spanish Government.

E.R. Hancock et al. (Eds.): SSPR & SPR 2010, LNCS 6218, pp. 286–295, 2010.

like identifying clusters corresponding to sub-populations. In the particular case of Protein-Protein Interaction (PPI) networks, we have found in a preliminary study that networks with similar complexity are quite different in terms of edit distance. Therefore our contribution fits a global (low-frequency) methodology for analysis of graphs. In this regard, spectral graph theory is a recurrent formal tool. Recent extensive use of spectral graph theory is due to: a) that it explains some previous approaches (the number of spanning trees, path-length distribution, clusterization, and so on) from a random walks perspective and b) that it is flexible enough to allow the development of new characterizations. In this paper we explore the connection between convex polytopes (and those of the Birkhoff type in particular), heat kernels in graphs, the well known thermodynamic depth approach to complexity, and network complexity itself. Some work in this direction has been done recently [3], but no formal connections between polytopes and heat-flow characterization of structural entropy [3] has been developed so far. Our main contribution here is to formally specify the complexity profiles of both approaches to structural complexity, showing that they have a qualitatively similar behavior and that the complexity corresponding to the maximum entropy (ME) Birkhoff-von Neumann decomposition is derived from that corresponding to the maximum flow. Thus, the phase-transition point always exists and it is characterized by such maxima. Moreover we establish links between heat flow complexity, the fluctuation theorem and thermodynamic depth. We also apply this characterization to PPI networks.

2 Polytopal vs. Heat Flow Complexity

Theorem 1 *(*Birkhoff-von Neumann (BvN) *[5]). Let $\mathcal{B}_n$ be is the set of doubly stochastic matrices $B = [b_{ij}]_{n\times n}$ of dimension $n \times n$ (*Birkhoff polytope*). Then every doubly stochastic matrix (*DSM*) B can be expressed as a convex combination of permutation matrices (*PM*):*

$$B = \sum_{\alpha} p_\alpha P_\alpha,\ \forall B \in \mathcal{B}_n \text{ and } \begin{array}{l} \sum_\alpha p_\alpha = 1 \\ p_\alpha \geq 0\ \forall \alpha \end{array} .$$

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 van der Waerden constant matrix B_* with all entries equal to $1/n$.

Theorem 2 *(*Agrawal, Wang & Ye *[6]). The maximum entropy (ME) BvN decomposition of a DSM B is the solution to the left optimization problem below (primal) whose dual one is on the right:*

$$\begin{array}{ll} \min \ \sum_{\alpha\in\mathcal{S}_n} p_\alpha(\log p_\alpha - 1) & \max\ B : Y - 1 \\ s.t.\ \sum_\alpha p_\alpha P_\alpha \leq B & s.t.\ \sum_\alpha e^{(Y:P_\alpha)} P_\alpha \leq B \\ p_\alpha \geq 0 & 0 \geq Y_{ij} \geq -n\frac{\log n}{b_{min}}\ \forall i,j \end{array}$$

where $\mathcal{S}_n$ *is the set of permutations of* $\{1, 2, \ldots, n\}$, $X : Z = \sum_{ij} X_{ij} Z_{ij} = trace(XZ^T)$ *is the Frobenius inner product,* $Y \in R^{n\times}$, *a matrix of Lagrange multipliers each corresponding to one constraint (component) in* $B = \sum_\alpha p_\alpha P_\alpha$, *and* $b_{min} = \min\{B_{ij}\}$.

In [6] it is shown how to solve approximately the dual of the ME problem. In practice, however, instead of finding a unique representation for B it is preferable to obtain greedily just one of them. To that end, the constructive proof of the BvN theorem is used. This is the origin of *polytopal complexity* [3].

Definition 1 *(*Polytopal Complexity *[3]). Given* $G = (V, E)$, *an undirected and unweighted graph 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 *as the* β*-dependent function*

$$\mathcal{BC}^\beta(G) = \frac{H(\mathcal{P})}{\log_2 n} = \frac{\log_2 \gamma + D(\mathcal{P}||\mathcal{U}_\gamma)}{\log_2 n}\,, \tag{1}$$

where $\mathcal{P} = \{p_1, \ldots, p_\gamma\}$ *is the probability density function (pdf) induced by the decomposition,* $H(.)$ *the entropy and* $D(.)$ *the Kullback-Leibler divergence* $D(\mathcal{P}||\mathcal{Q}) = \sum_\alpha p_\alpha \log \frac{p_\alpha}{q_\alpha}$.

In [3] it is argued that the typical signature is heavy tailed, monotonically increasing from 0 to $\beta^+ \equiv \arg\max\{\mathcal{BC}^\beta(G)\}$ and either monotonically decreasing or stable from β^+ to ∞ where $\mathcal{BC}^\beta(G) = 1$ is reached. Thus, β^+ represents the most significant *topological phase transition* regarding the impact of the diffusion process in the topology of the input graph. However, no characterization theorem has been enunciated so far in order to validate the latter assumptions. In addition, in [4] it is showed that the $O(n^5)$ computational complexity of the greedy BvN decomposition for each β precludes the use of the descriptor for the practical analysis of complex networks. Thus, a new descriptor, qualitatively similar but more efficient than the current one, and also providing a simpler analytical framework, is needed.

Definition 2 *(*Heat Flow Complexity *[4]). Given* $G = (V, E)$ *with* $|V| = n$ *and adjacency matrix A. The diffusion kernel is* $K^\beta(G) = \exp(-\beta\mathcal{L}) \equiv \Phi\Lambda\Phi^T$, *being* $\Lambda = diag(e^{-\beta\lambda_1}, e^{-\beta\lambda_2}, \ldots, e^{-\beta\lambda_n})$, *and* $\lambda_1 = 0 \leq \lambda_2 \leq \ldots \leq \lambda_n$ *are the eigenvalues of* $\mathcal{L}$. *Therefore, the* total heat flowing through the graph *at a given* β *is:*

$$F^\beta(G) = \sum_{i=1}^{n}\sum_{j\neq i}^{n} \delta_{ij} \underbrace{\left(\sum_{k=1}^{n} \phi_k(i)\phi_k(j)e^{-\lambda_k\beta}\right)}_{K_{ij}^\beta}\,, \tag{2}$$

where $\delta_{ij} = 1$ *iff* $(i, j) \in E$. *Then,* heat flow complexity *is defined as:*

$$\mathcal{FC}^\beta(G) = \frac{\log_2(1 + F^\beta(G))}{\log_2 n}\,. \tag{3}$$

3 Characterization of Polytopal and Flow Complexity

3.1 Characterization of Phase Transition

Theorem 3 *(*Phase-Transition Point*). Let $G = (V, E)$ be a graph with $|V| = n$ and edge-set E. Then, there exists a unique finite inverse temperature $\beta^+ \geq 0$ so that β^+ is the maximal value for which the sum of the off-diagonal elements of the diffusion kernel (or Gram matrix) on graph G is less that the sum of the on-diagonal elements. In other words, there exists an unique $\beta^+ \geq 0$ so that $\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_{ij}^{\beta^+} < trace(K^{\beta^+})$, and $\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_{ij}^{\beta} \geq trace(K^{\beta})\ \forall \beta > \beta^+$.*

Proof. Let us analyze the behavior of the function $\Xi^\beta = trace(K^\beta) - \sum_{i=1}^{n}\sum_{j\neq i}^{n} K_{ij}^{\beta}$. The analysis of the limiting cases $K^0 = I_n$ and $K^\infty = \mathcal{B}_*$ yields $\Xi^0 = n$ and $\Xi^\infty = -n$. Actually $-n$ 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. Thus, the PTP exists just before the zero-crossing $\Xi^\beta = 0$ and it is unique. □

The existence of a unique PTP is key to relating heat flow and maximal entropy.

Theorem 4 *(*Phase-Transition*). Let $\beta^+ > 0$ define a PTP. Then, the heat flow $F^{\beta^+}(G)$ corresponding to the PTP is maximal among all choices of β. Moreover, this implies that the entropy $H^{\beta^+}(\mathcal{P})$ with $\mathcal{P} = \{p_1, \ldots, p_\gamma\}$ corresponding to the maximal entropy BvN decomposition of $K^{\beta^+}(G) = \sum_{\alpha=1}^{\gamma} p_\alpha P_\alpha$ is maximal over β.*

Proof (Flow Maximality at PTP). Consider $\beta < \beta^+$ and suppose that $F^\beta > F^{\beta^+}$, that is, $\sum_{ij}\delta_{ij}K_{ij}^{\beta} > \sum_{ij}\delta_{ij}K_{ij}^{\beta^+}$. We can write $\sum_{ij}\delta_{ij}K_{ij} = A : K$, where A the adjacency matrix of G and $X : Y = \sum_{ij} X_{ij}Y_{ij}$ denotes here the Frobenius inner product. It follows that $A : K^\beta > A : K^{\beta^+}$. All the off-diagonal elements of K_β decrease at β, with respect to their values at β^+ due to the diffusion process. As a result, the sum of off-diagonal elements of K^β is smaller than the sum of off-diagonal elements of K^{β^+}. Moreover, as on-diagonal elements are zero on A, we have that $A : K^\beta \leq A : K^{\beta^+}$ which is a contradiction. Therefore $F^\beta \leq F^{\beta^+}$.

Consider now the case $\beta > \beta^+$ and also $F^\beta > F^{\beta^+}$. Then, we should have that $A : K^\beta > A : K^{\beta^+}$ which is consistent with the fact that the sum of off-diagonal elements is more and more greater or equal than the sum of on-diagonal elements as β increases. This is due to the fact that off-diagonal values which are not associated to an edge in the graph increase whereas on-diagonal ones decrease. However the individual values of both diagonal and off-diagonal elements are bounded by $1/n$, and tend to such value as β increases. Furthermore, when all values reach $1/n$ at a given inverse temperature, such equilibrium state remains constant for greater values of the inverse temperature. If the equilibrium is reached later than β, only off-diagonal elements which are not associated to an edge (but to a path) increase. However, edge-associated off-diagonal elements decrease which implies $A : K^\beta < A : K^{\beta^+}$, that is $F^\beta < F^{\beta^+}$ which is a contradiction. If β corresponds to an inverse temperature beyond the equilibrium

value, then we have that $F_\beta = \frac{2|E|}{n}$ which must be greater than F_{β^+} (where the sum of the on-diagonal elements is greater than that of the off-diagonal elements) and the off-diagonal elements associated with edges have a greater value since $\beta^+ < \beta^1$. Therefore we have again a contradiction. The limiting case is that equilibrium is reached at $\beta = \beta^+$. In that case we have also contradiction because $F^\beta = F\beta^+ = \frac{2|E|}{n}$. From the contradiction in the two cases $\beta < \beta^+$ and $\beta > \beta^+$, we have $F^{\beta^+} > F^\beta$ for all values of $\beta \in [0, +\infty)$.

Proof (Entropy Maximality at PTP). Let H^β the entropy corresponding to the maximum entropy BvN decomposition for a given β. Now, we have to prove that $F^{\beta^+} > F^\beta \Rightarrow H^{\beta^+} > H^\beta$, for any β. The maximum entropy BvN decomposition yields $p_\alpha = e^{Y:P_\alpha}$, and $Y \in \mathbb{R}^{n\times n}$ is the matrix of Lagrange multipliers satisfying the condition $0 \leq K_\beta : Y = trace(K_\beta Y^T) \leq -n \log n$ (see proof of Lemma 5 in [6]). Such a BvN decomposition is unique for the given value of β, and the Lagrange multipliers which correspond to dual variables associated to the $n \times n$ constraints $K^\beta(G) = \sum_{\alpha=1}^{\gamma} p_\alpha P_\alpha$. Let $k_{min} = \min_{ij}\{K^\beta_{ij}\}$ be the minimal component in K_β. Then, every multiplier satisfies the bound $0 \geq Y_{ij} \geq -\frac{n \log n}{k_{min}}$. Consequently, those kernel elements that are zero or close-to-zero may enlarge the bounds (see the dual problem) up to $-\infty$ (when $k_{min} = 0$). These large bounds imply that $p_\alpha \to 0$ for some value of α (the exponential argument in $e^{Y:P_\alpha}$ may be $-\infty$), but not necessarily to all of them because of the different structures of the associated permutation matrices P_α in each case. This occurs at every β for the same graph G. In the limiting cases of $\beta = 0$ and $\beta \to +\infty$ we have, respectively, $p_\alpha = 1$ for the unique $P_\alpha = I_n$ and $p_\alpha = 1/n$ (all kernel components are $1/n$) for the n permutation matrices, where $H^0 = 0$ and $H^{+\infty} = \log_2 n$. The respective flows are $F^0 = 0$ and $F^{+\infty} = \frac{2|E|}{n}$.

Proving that $H^\beta < H^{\beta^+}$ for each $\beta \neq \beta^+$ is equivalent to prove $-n \log n \leq K^\beta : Y^\beta < K^{\beta^+} : Y^{\beta^+} \leq 0$ for each $\beta \neq \beta^+$, since we are maximizing $K : Y - 1$ in the dual problem, being Y^β and Y^{β^+} respectively the optimal Lagrange multipliers corresponding to the maximum entropy BvN decompositions at β and β^+. This means that the multipliers (which are all negative) are set to their maximal (close-to-zero) values provided that the decomposition constraints are satisfied. Given their theoretical bounds $0 \geq Y^{\beta^+}_{ij} \geq \frac{-n \log n}{k^{\beta^+}_{min}}$ and $0 \geq Y^\beta_{ij} \geq \frac{-n \log n}{k^\beta_{min}}$, the Lagrange multipliers can be arbitrarily close to zero. Each multiplier is related to a kernel component (the Frobenius inner product is the sum of the elements of the matrix resulting from the Hadamard product) and both kernels are DSMs. Hence, we must only set $Y^{\beta^+}_{ij}$ and Y^β_{ij} to their minimal values when $K^{\beta^+}_{ij} = 0$ and $K^\beta_{ij} = 0$ so that each Frobenius product is maximized (given that p_α is defined by the exponential of $Y : P_\alpha$).

For $\beta < \beta^+$, As β^+ defines a PTP, we have that the sum of the $n^2 - n$ off-diagonal values in K^β is lower than the n on-diagonal elements. Therefore we obtain K^β :

[1] Furthermore, for large β we have that $K_\beta = e^{-\beta\lambda_2}\phi_2\phi_2^T$, where ϕ_2 is the Friedler vector.

$Y^{\beta} < K^{\beta^+} : Y^{\beta^+}$ which is due to the fact that, although the multipliers are chosen as close to zero as possible, the most negative multipliers must be assigned to the lower elements in K^{β} in order to maximize the Frobenius product. Although the less negative elements correspond with the (dominant) diagonal elements of K^{β}, they become more closer to zero than at β^+. There will be an increasing number of zero elements as $\beta \to 0$, since in these conditions we have $K^{\beta} = (I_n - \mathcal{L}\beta)$ which means that the on-diagonal elements will be closer to the unity, and we have the freedom to assign negative multipliers to increasingly small off-diagonal elements. The latter assigment yields a small $K_{\beta} : Y_{\beta}$. However, as we approach β^+, where off-diagonal elements start to dominate, it is more convenient to assign the close-to-zero multipliers to dominant elements and then the Frobenius product increases.

When $\beta > \beta^+$, the sum of off-diagonal values is greater to the sum of on-diagonal ones until the equilibrium point is reached. If in addition $A : K^{\beta^+} > A : K^{\beta}$ before equilibrium and recalling that on-diagonal elements at β are smaller than their values at β^+, we obtain $K^{\beta^+} : Y^{\beta^+} > K^{\beta} : Y^{\beta}$. This is due to: (i) that it is desirable to assign the closer-to-zero multipliers the off-diagonal elements, and the more negative ones to the diagonal in order to maximize the Frobenius product; (ii) that the latter assignment is increasingly infasible as β grows because of the increasing number of constraints over these multipliers as $\Xi^{\beta} = trace(K^{\beta}) - \sum_{i=1}^{n} \sum_{j \neq i}^{n} K_{ij}^{\beta}$ decreases. Under this latter condition, heat flow increases through the edges and establishes virtual paths (reachability) between those node pairs not connected by edges. As a result, there is an increase of the off-diagonal elements associated with indirect paths (rather than connecting edges) . An increasing number of close-to-zero multipliers are needed for the latter elements in order to maximize the Frobenius product. However, not all off-diagonal elements can have a close-to-zero multiplier and some of them will be very negative. If β is closer to β^+ than to the equilibrium point, the off-diagonal elements associated to indirect paths can be very negative and thus $K^{\beta^+} : Y^{\beta^+} > K^{\beta} : Y^{\beta}$. As β reaches the equilibrium point all the elements tend to $1/n$, which implies that all multipliers are almost equal but less or equal to any multiplier at β^+. Then, again $K^{\beta^+} : Y^{\beta^+} > K^{\beta} : Y^{\beta}$ even beyond the equilibrium point. Therefore, for $\beta \neq \beta^+$ we have $H^{\beta^+} \geq H^{\beta}$. □

3.2 The Fluctuation Laws

The fluctuation theorem (FT) states that the probability of *destroying entropy* in an isolated (macroscopic or microscopic) system decreases exponentially with time.Herein, as the β inverse temperature is assimilated to time t, and entropy H_{β} is assimilated to heat flow F_{β}, we do not have the case of *destroying entropy* in the sense of having a negative entropy. However, due to the existence of a PTP, for each flow trace F_{β} a *entropy production phase* $[0, \beta^+]$ and an *entropy stabilization/destruction phase* $(\beta^+, \beta^{max}]$. Entropy decayment after the PTP is due to the structure imposed by the network (but in the complete graph where there is no topological constrain) because structure means information. However, is there a formal relation between the rate of entropy production and that of entropy reduction in a set of networks representing, for instance, the same

phenomenon like a PPI? We have found that (see the experimental section) for many sets of networks such a relation exists and also that it is linear. Moreover, besides linearity, there also exists an exponential decay.

Definition 3 *(*Fluctuation Laws*). Let $\Omega = \{G_i = (V_i, E_i)\}$ where for $G_i \in \Omega$ we have that $F_\beta(G_i)$ is the heat flow trace for $\beta \in [0, \beta^+]$ and $\langle \nabla F(G_i) \rangle_a^b$ is the average flow gradient between $\beta = a$ and $\beta = b$. Such set of networks satisties the* linear fluctuation law (LFL) *if exists $k > 0$ so that*

$$Pr\left(\langle \nabla F(G_i) \rangle_{\beta^+}^{\beta^{max}} = -k \langle \nabla F(G_i) \rangle_0^{\beta^+} \right) \approx 1,$$

that is,entropy variation decay is, with high probability, larger (in absolute value) as the entropy variation increase grows (the more entropy is produced at the beginning, the more is distroyed beyond the PTP) and variation decay is linear with respect to variation increasing along the population of networks. If in addition to satisfying LFL *, for every pair $(G_i, G_j), i \neq j$ so that $G_i, G_j \in \Omega$ and exists $\lambda > 0$ so that $Pr\left(\left| \langle \nabla F(G_i) \rangle_{\beta^+}^{\beta^{max}} - \langle \nabla F(G_j) \rangle_{\beta^+}^{\beta^{max}} \right| \right) = e^{-\lambda}$, we have that the set of networks satisfies the* linear fluctuation law with exponential decay (LFLED).

4 Heat Flow - Thermodynamic Depth Complexity

The application of thermodynamic depth (TD) to characterize network complexity demands the formal specification of the micro-states whose history leads to the macro-state (of the network). Here we define such micro-states in terms of *expansion subgraphs*.

Definition 4 *(*Node History & Expansion Subgraphs*). Let $G = (V, E)$ with $|V| = n$. 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 on $z \sim j : j \in V_{e(i)}, z \notin V_{e(i)}$, 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. Thus, in terms of TD the full graph G or the union of its connected components is the macro-state (macroscopic state). The *depth* of such macro-state relies on the variability of the causal trajectories leading to it. The higher the variability, the more complex it is to explain how the macro-state 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* (MEBB) [8]. 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 $\mathcal{X} \subseteq R^d$ is a convex domain, and d the data dimension (in this case the number of discretized β - inverse temperatures). Given two patterns (discretized functions in this case)

$\boldsymbol{f}$ and $\boldsymbol{g}$, $D_F(\boldsymbol{f}||\boldsymbol{g}) = F(\boldsymbol{f}) - F(\boldsymbol{g}) - (\boldsymbol{f} - \boldsymbol{f})^T \nabla F(\boldsymbol{f})$. Here, we use the I-Kullback-Leibler divergence $D_F(\boldsymbol{f}||\boldsymbol{g}) = \sum_{i=1}^{d} f_i \log \frac{f_i}{g_i} - \sum_{i=1}^{d} f_i + \sum_{i=1}^{d} g_i$ with $F(\boldsymbol{f}) = \sum_{i=1}^{d}(f_i \log f_i - f_i)$ (un-normalized Shannon entropy) which yields better results (more representative centroids of heat flow complexities) than other divergences/distorions like that of Itakura-Saito. When using the I-KL divergence in R^d, we have that $\nabla F(f_i) = \log f_i$ and also that $\nabla^{-1} F(f_i) = e^{f_i}$ (obviously the natural logarithm is assumed). Using these formal ingredients we define the *causal trajectory* in terms of MEBBs.

Definition 5 *(*Causal Trajectory*). Given $h_i(G)$, the heat flow complexity $\boldsymbol{f}_t = f(e^t(i))$ for the $t-th$ 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 $\boldsymbol{c}_i \in R^d$ and radius $r_i \in R$ of the MEBB $\mathcal{B}^{\boldsymbol{c}_i, r_i} = \{\boldsymbol{f}_t \in \mathcal{X} : D_F(\boldsymbol{c}_i||\boldsymbol{f}_t) \leq r_i\}$.*

Solving for the center and radius implies finding $\boldsymbol{c}^*$ and r^* minimizing r subject to $D_F(\boldsymbol{c}_i||\boldsymbol{f}_t) \leq r \; \forall t \in \mathcal{X}$ with $|\mathcal{X}| = T$. Considering the Lagrange multipliers α_t we have that $\boldsymbol{c}^* = \nabla^{-1} F(\sum_{t=1}^{T} \alpha_t \boldsymbol{f}_t \nabla F(\boldsymbol{f}_t))$. The efficient algorithm in [8] estimates both the center and multipliers. This idea is closely related to Core Vector Machines [9], and it is interesting to focus on the non-zero multipliers (and their support vectors) used to compute the optimal radius. More precisely, the multipliers define a convex combination and we have $\alpha_t \propto D_F(\boldsymbol{c}^*||\boldsymbol{f}_t)$, and the radius is simply chosen as: $r^* = \max_{\alpha_t > 0} D_F(\boldsymbol{c}^*||\boldsymbol{f}_t)$.

Definition 6 *(*TD Network Depth*). Given $G = (V, E)$, with $|V| = n$ and all the n pairs $(\boldsymbol{c}_i, r_i)$, the* heat flow-thermodynamic depth complexity *of G is characterized by the MEBB $\mathcal{B}^{\boldsymbol{c},r} = \{\boldsymbol{c}_t \in \mathcal{X}_i : D_F(\boldsymbol{c}||\boldsymbol{c}_i) \leq r\}$ and $D_{min} = \min_{f \in \mathcal{B}^{\boldsymbol{c},r}} D_F(f^\infty||f)$, where $f^\infty = f(B_*) \in R^d$ is the van der Waerden complexity trace . As a result, the* TD depth of network *is given by $\mathcal{D}(G) = r \times D_{min}$.*

The above definitions of complexity and depth are highly consistent with summarizing 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. The combinations or hierarchies of MEBBs have proved to be more effective than ball trees for nearest-neighbor retrieval [10]. In the computation of depths, the Legrendre duality (convex conjugate) is key because it establishes a one-to-one correspondence between the gradients ∇F and ∇F^{-1} due to the convexity of F. Therefore, the Bregman projection f of f^∞ on the the border of $\mathcal{B}^{\boldsymbol{c},r}$ lies on the curve $f_\theta^{-1} = \theta \nabla F(\boldsymbol{c}) + (1-\theta)\nabla F(f^\infty)$ with $\theta \in [0, 1]$ and $f_\theta = \nabla^{-1} F(f_\theta^{-1})$. The projection f be easily found (approximately) through bisection search on θ.

5 Experiments: TD of PPIs

We have designed an experimental section using PPIs extracted from STRING–8.2[2]. In a first experiment, we consider PPIs related to *histidine kinase*, a key protein in the development of signal transduction, corresponding to 10 species belonging to 10 phyla of bacteria. We select subjectively 3 PPIs (simple, complex and more-complex) from each species and compute their TDs. In 70% of the cases, TD matches intuition. When comparing with Estrada's spectral homogeneity descriptor [2] we also find that the ratio between intraclass and

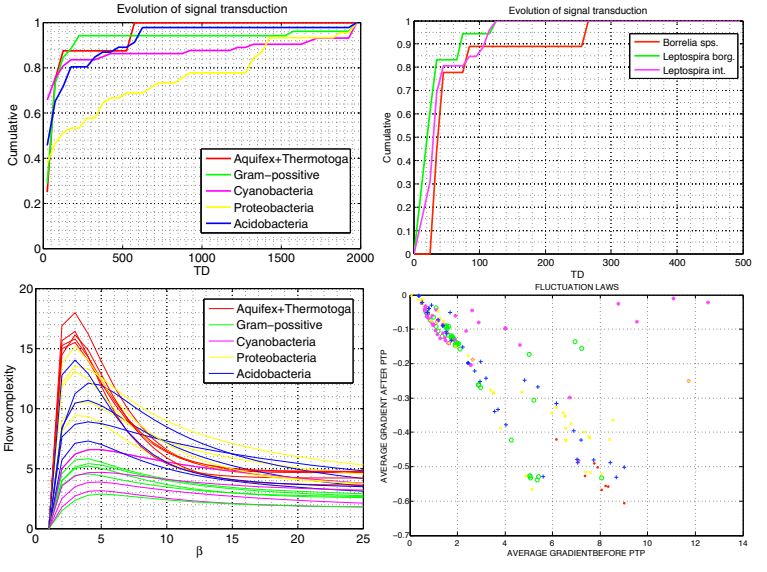

Fig. 1. PPI analysis with TD and Illustration of Fluctuation Laws (bottom-left)

interclass variability is slightly better (smaller) for TD (0.6840 vs 0.7383). The second experiment consists of analyzing 222 PPIs, also related to histidine kinase, from 6 different groups (all the PPIs in the same group corresponds to the same species) with the following evolutive order (from older to more recent): *Aquifex* –4 PPIs, *Thermotoga*–4 PPIs, *Gram-Positive*–52 PPIs, *Cyanobacteria*–73 PPIs *Proteobacteria*–45 PPIs. There is an additional class (*Acidobacteria*—46 PPIs). Histogramming TDs 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 pdfs (Fig. 1-left/top). In such case, reaching the top (cumulative=1) soon indicates low TD whereas reaching it later indicates high TD. Then, it can be seen that the evolving complexity of the signal transduction mechanism driven by the histidine kinase is properly quantified by TD for

[2] http://string.embl.de/

the 5 first phyla studied. However, the Acidobacterium sp. chosen seems older than Gram-Postive which seems not to be the case. In the bottom of Fig. 1-left/bottom we show some $\boldsymbol{c}_i$s of all classes, and their intraclass variability is low (similar shape). Thus, we can conclude that TD is a good principled tool for analying the complexity of networks. In a third experiment we analyze the cumulatives of three different species of the same phylum (76 PPIs of *Spirochaetes*) to check that the intra-species variability is low (Fig. 1-right/top). Finally, we show how the PPIs analyzed in the second experiment follow the fluctuation law, and some of them (we preserve the colors of Fig. 1-left/top) like *Cianobacteria* follow the LFLED.

6 Conclusions and Future Work

In this work, there are four contributions: a) the characterization heat flow complexity in terms of information theory, b) to define structural complexity in terms of Heat Flow-Thermodynamic Depth, c) to explore connections between the heat-flow thermodynamic depth and the fluctuation theorem and d) test the formal definition in terms of characterizing the evolution of Bacteria through quantifying the TD of their PPIs. Future work includes both exploring formal links with the Ihara Zeta function and studying different kind of networks.

References

[1] Banerjee, A., Jost, J.: Graph Spectra as a Systematic Tool in Computational Biology. Discrete Applied Mathematics 157(10), 27–40 (2009)
[2] Estrada, E.: Spectral Scaling and Good Expansion Properties in Complex Networks. Europhys. Lett. 73(4), 649 (2006)
[3] Escolano, F., Hancock, E.R., Lozano, M.A.: Birkhoff Polytopes, Heat kernels and Graph complexity. In: Proc. of ICPR, pp. 1–5 (2008)
[4] Escolano, F., Giorgi, D., Hancock, E.R., Lozano, M.A., Falcidieno, B.: Flow Complexity: Fast Polytopal Graph Complexity and 3D Object Clustering. In: Proc. of GbRPR, pp. 253–262 (2009)
[5] Birkhoff, G.D.: Tres Observaciones sobre el Algebra Lineal. Universidad Nacional de Tucuman Revista, Serie A 5, 147–151 (1946)
[6] Agrawal, S., Wang, Z., Ye, Y.: Parimutuel Betting on Permuations. In: Papadimitriou, C., Zhang, S. (eds.) WINE 2008. LNCS, vol. 5385, pp. 126–137. Springer, Heidelberg (2008)
[7] Jerrum, M., Sinclair, A., Vigoda, E.: A Polynomial-time Approximation Algorithm for the Permanent of a Matrix with Nonnegative Entries. Journal of the ACM 51(4), 671–697 (2004)
[8] Nock, R., Nielsen, F.: Fitting Smallest Enclosing Bregman Ball. In: Gama, J., Camacho, R., Brazdil, P.B., Jorge, A.M., Torgo, L. (eds.) ECML 2005. LNCS (LNAI), vol. 3720, pp. 649–656. Springer, Heidelberg (2005)
[9] Tsang, I.W., Kocsor, A., Kwok, J.T.: Simple Core Vector Machines with Enclosing Balls. In: Proc. of ICLM, pp. 911–918 (2007)
[10] Cayton, L.: Fast Nearest Neighbor Retrival with Bregman Divergences. In: Proc. of ICLM, pp. 112–119 (2008)
[11] Slater, P.B.: Maximum-Entropy Representations in Convex Polytopes. Env. and Planning 21, 1541–1546 (1989)