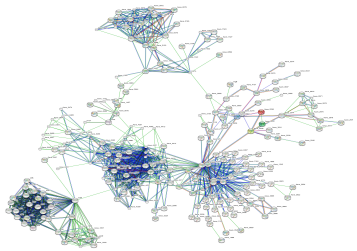
What is the Complexity of a Network? The Heat Flow-Thermodynamic Depth Approach

F. Escolano, M.A. Lozano, E.R. Hancock, D. Giorgi

University of Alicante, Spain; Univerisity of York UK, and IMATI CNR, Genova, Italy

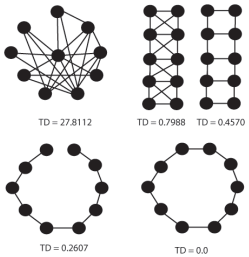
Abstract

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. 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**.



Polytopal vs Heat Flow Complexity

What is the complexity of a network/graph? How to quantify it?



(*Polytopal Complexity* [**?**]). 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

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

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 $D(\mathcal{P}||\mathcal{Q}) = \sum_\alpha p_\alpha \log \frac{p_\alpha}{q_\alpha}$.

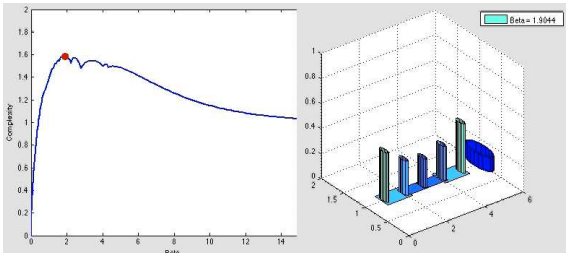
(*Heat Flow Complexity* [**?**]). Given $G = (V, E)$ with $|V| = n$ and adjacency matrix A. The diffusion kernel is $K^\beta(G) = \exp(-\beta\mathcal{L}) \equiv \Phi\Lambda\Phi^T$, being $\Lambda = diag(e^{-\beta\lambda_1}, e^{-\beta\lambda_2}, \ldots, e^{-\beta\lambda_n})$, and $\lambda_1 = 0 \leq \lambda_2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

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

Characterization of Polytopal and Flow Complexity

(*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 $\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_{ij}^{\beta} \geq trace(K^{\beta}) \; \forall \beta > \beta^+$.

(*Phase-Transition*) Let $\beta^+ > 0$ define a PTP. Then, the heat flow $calF^{\beta^+}(G)$ 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.

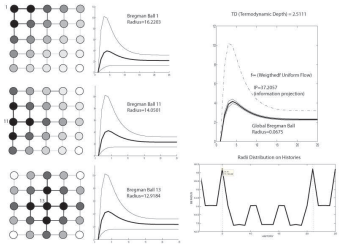


Heat Flow - Thermodynamic Depth Complexity

(*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.

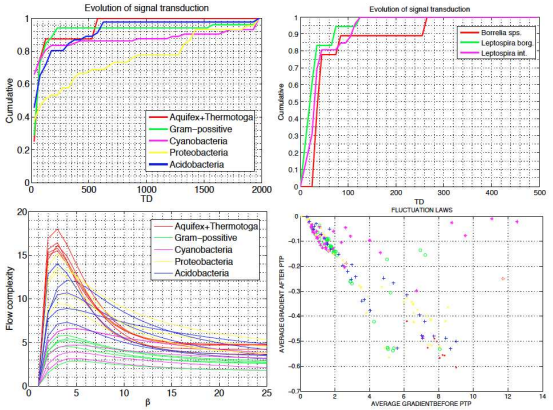
(*Causal Trajectory*) Given $h_i(G)$, the heat flow complexity $\vec{f}_t = f(e^t(i))$ for the $t-th$ expansion of i, a generator F and a Bregman divergence D_F, the *causal trajectory* leading to G (or one of its connected components) from i is characterized by the center $\vec{c}_i \in R^d$ and radius $r_i \in R$ of the MEBB $\mathcal{B}^{\vec{c}_i, r_i} = \{\vec{f}_t \in \mathcal{X} : D_F(\vec{c}_i||\vec{f}_t) \leq r_i\}$.

(*TD Network Depth*) 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 $\mathcal{B}^{\vec{c},r} = \{\vec{c}_t \in \mathcal{X}_i : D_F(\vec{c}_t||\vec{c}_i) \leq r\}$ 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 network* is given by $\mathcal{D}(G) = r \times D_{min}$.



Experimental Results

The main 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. 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 the 5 first phyla studied. However, the Acidobacterium sp. chosen seems older than Gram-Postive which seems not to be the case. We also show some $\vec{c}_i$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. We also analyze the cumulatives of three different species of the same phylum (76 PPIs of *Spirochaetes*) to check that the intra-species variability is low. Finally, we show how the PPIs analyzed in the second experiment follow the fluctuation law, and some of them like *Cianobacteria* follow the LFLED.



Conclusions

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.

References

[1] Escolano, F., Hancock, E.R., Lozano, M.A.: Birkhoff Polytopes, Heat kernels and Graph complexity. Proc. of ICPR, 1–5, (2008)

[2] Escolano, F., Giorgi, D., Hancock, E.R., Lozano, M.A., Falcidieno, B.: Flow Complexity: Fast Polytopal Graph Complexity and 3D Object Clustering. Proc. of GbRPR, 253–262 (2009)

[3] Estrada, E.: Spectral Scaling and Good Expansion Properties in Complex Networks. Europhys. Lett. 73(4), 649 (2006)