

# Polytopal Graph Complexity, Matrix Permanents, and Embedding

Francisco Escolano<sup>1</sup>, Edwin R. Hancock<sup>2</sup>, and Miguel A. Lozano<sup>1</sup>

<sup>1</sup> University of Alicante, Departamento de Ciencia de la Computación e Inteligencia Artificial, Spain

<sup>2</sup> University of York, Department of Computer Science, UK

**Abstract.** In this paper, we show how to quantify graph complexity in terms of the normalized entropies of convex Birkhoff combinations. We commence by demonstrating how the heat kernel of a graph can be decomposed in terms of Birkhoff polytopes. Drawing on the work of Birkhoff and von Neuman, we next show how to characterise the complexity of the heat kernel. Finally, we provide connections with the permanent of a matrix, and in particular those that are doubly stochastic. We also include graph embedding experiments based on polytopal complexity, mainly in the context of Bioinformatics (like the clustering of protein-protein interaction networks) and image-based planar graphs.

## 1 Introduction

The quantification of the *intrinsic complexity* of undirected graphs has attracted significant attention due to its fundamental practical importance, not only in pattern recognition but also other areas such as control theory and network analysis. Such a quantification not only allows the complexity of different graph structures to be compared, but also allows the complexity to be traded against goodness of fit to data when a structure is being learned. Previous complexity characterizations include: a) the number of spanning trees, its connections with the Laplacian spectrum, b) methods based on the path-length chromatic decomposition, c) the number of Boolean operators necessary to build the graph from generator graphs, and more recently, d) the notion of linear complexity of any of the associated adjacency matrices.

In pattern recognition and machine learning the problem of quantifying graph complexity is not only deeply related to embedding methods [1][2] for structural classification or indexing [3], but is also key to the process of constructing graph prototypes [4][5]. In this paper we explore the connection between convex polytopes (and those of the Birkhoff type in particular), heat kernels in graphs and graph structure. Our contribution is to propose and characterize a novel multi-dimensional measure of complexity. This measure is motivated by the notion of *graph entropy* introduced by Körner in [6], which provides a useful framework based on convex polytopes. Unfortunately, the graph entropy is defined for graphs with a probability distribution function over nodes. In this paper we

focus on more general unweighed graphs. We commence with a characterisation based on the heat kernel of a graph, and show how the heat kernel can be decomposed into Birkhoff polytopes. Next, we use the work of Birkhoff and von Neuman to characterize the complexity of the decomposition, and then establish links to the permanents of a graph. We experiment with the complexity measure for the problem of graph-embedding, and provide examples on data provided by bioinformatics and computer-vision. A preliminary version of the ideas presented here appear in [7]. Here we provide a deeper theoretical and experimental analysis of the proposed method.

## 2 The Birkhoff Polytope

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

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

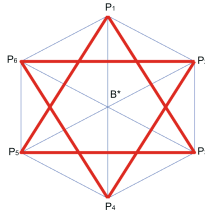
Such a polytope is convex, has dimension  $(n - 1)^2$  and its extreme points are the set of  $n!$  permutation matrices  $P = [p_{ij}]_{n \times n}$  (entries are 0 except for a single 1 in each row and column). The *Birkhoff-von Neumann (BvN) theorem* [8] states that every doubly stochastic matrix (*DSM*) can be expressed as a convex combination of permutation matrices (*PM*):

$$B = \sum_{\alpha} p_{\alpha} P_{\alpha}, \forall B \in \mathcal{B}_n \text{ and } \begin{array}{l} \sum_{\alpha} p_{\alpha} = 1 \\ p_{\alpha} \geq 0 \forall \alpha \end{array}. \tag{2}$$

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  $1/n$ .

## 3 The Link: DSMs and Heat Kernels

The *BvN* decomposition has been applied to traffic problems [9](where the  $P_{\alpha}$  are crossbar switch configurations and  $p_{\alpha}$  are the probabilities of choosing such configurations). Moreover, Birkhoff polytopes are special cases of *transportation* polytopes. Consider an undirected graph  $G = (V, E)$  where  $|V| = n$ , and adjacency matrix  $A$ , as the model of a network. The so called  $n \times n$  *heat/diffusion kernel* of the graph is found by the matrix exponentiation  $K(G) = \exp(-\beta \mathcal{L})$ , (with  $\beta > 0$ ) of the graph Laplacian  $\mathcal{L} = A - D$  where  $D = \text{diag}(d_i = \sum_{j \in V} A_{ij}, i \in V)$  is the degree matrix. It turns out that  $K$  is a *DSM* where  $K_{ij}$  encodes the probability of reaching vertex  $j$  from vertex  $i$ , and viceversa, through lazy random walks [10]. Therefore, the decomposition  $K(G) = \sum_{\alpha} p_{\alpha} P_{\alpha}$  will not only map a



$$P_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad P_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad P_3 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad P_4 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \quad P_5 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \quad P_6 = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} .$$

$$\Delta_1 = p_1 P_1 + p_3 P_3 + p_5 P_5 = \begin{bmatrix} p_1 & p_3 & p_5 \\ p_5 & p_1 & p_3 \\ p_3 & p_5 & p_1 \end{bmatrix} \quad p_1 + p_3 + p_5 = 1$$

$$\Delta_2 = p_2 P_2 + p_4 P_4 + p_6 P_6 = \begin{bmatrix} p_2 & p_4 & p_6 \\ p_4 & p_6 & p_2 \\ p_6 & p_2 & p_4 \end{bmatrix} \quad p_2 + p_4 + p_6 = 1$$

**Fig. 1.** The  $\mathcal{B}_3$  polytope. It has  $6 \cdot 5/2 = 15$  edges (all of them extremal, which is not the usual case). There are 6 long edges with  $D^2 = Tr(A - B)(A^T - B^T) = 6$  of the triangles sides and 9 short ones with length  $D^2 = 4$ . All the 2-faces of  $\mathcal{B}_3$  are triangles with one long and two short edges. Finally  $Tr(\Delta_1 - B_*)(\Delta_2^T - B_*^T) = 0$

graph into  $\mathcal{B}_n$  but also the probability distribution induced by the convex combination will *weight with each  $p_\alpha$  the interactions between vertices encoded by the corresponding  $P_\alpha$* . An interesting example is the complete graph of  $n$  vertices  $C_n$  whose diffusion kernel is the barycenter  $K(C_n) = B_*$ , that is, the van der Waerden matrix. In addition, the probability distribution, with exactly  $n$  coefficients, associated to the *BvN* decomposition is the uniform one  $\mathcal{U}_n$  where  $p_\alpha = 1/n, \forall \alpha$  (see for instance  $\mathcal{B}_3$  in Fig. 1).

## 4 Polytopal Complexity

### 4.1 The *BvN* Constructive Decomposition

The most intuitive way of computing the *BvN* is to exploit the constructive proof of the theorem [11] follows: (i) Initially pose  $K(G)$ , which is a *DSM*, as a convex combination of a permutation matrix and another *DSM*  $K(G) \equiv K_1 = \lambda_1 P_1 + (1 - \lambda_1) K_2$ ; and (ii) Iteratively decompose  $K_2 = \lambda_2 P_2 + (1 - \lambda_2) K_3$  until the final *DSM* is a permutation. Let  $\gamma$  be the number of permutations needed to encode  $K$ , satisfying  $2 < \gamma < n^2$ , and let  $i$  be the step number of the iterative process. With these ingredients, the decomposition is

$$K(G) = \lambda_1 P_1 + (1 - \lambda_1) \lambda_2 P_2 + \dots = \sum_{\alpha=1}^{\gamma} \underbrace{\left\{ \prod_{i=1}^{\alpha-1} (1 - \lambda_i) \right\}}_{p_\alpha} \lambda_\alpha P_\alpha , \quad (3)$$

if  $\gamma = 1$  or it is even, and it is the expression above from which we remove the last  $\lambda_\alpha$  otherwise. What is more important, from the point of view of graph theory, is the way permutations are selected at each step. Let  $K$  be the *DSM* which we wish to decompose. Such a matrix induces a weighted bipartite graph  $H = (V', E', W)$  where  $V' = \{s_1, \dots, s_n, r_1, \dots, r_n\}$  and there is an edge  $e_{ij} = (s_i, r_j) \in E'$  when  $K_{ij} \neq 0$  and has weight  $W_{ij} = K_{ij}$ . As  $H$  is bipartite,  $V'$  is disjointly partitioned into  $S = \{s_i\}$  and  $R = \{r_j\}$ . Thus, for any vertex  $s_i$  its neighbors satisfy the condition  $N(s_i) \subset R$ , and for  $r_j$  we have  $N(r_j) \subset S$ . Given that  $K$  is a *DSM*, we have that  $\sum_{u \in N(s_i)} W_{iu} = \sum_{v \in N(r_j)} W_{vj} = 1$ . Therefore, if  $A \subset R$ , the sum of weights connecting vertices  $x \in A$  and  $y \in N(A)$  is exactly  $|A| = \sum_{x \in A, y \in N(A)} W_{xy} = \sum_{x \in A} \sum_{z \in N(y)} W_{zx} = \sum_{x \in A} 1$ . Similarly, the size of  $N(A) \subset S$  is given by the sum of the weights connecting the vertices of  $N(A)$  and the vertices of  $N(N(A))$  which are a superset of the vertices of  $A$ . Therefore,  $|N(A)| \geq |A|$ . This allows us to apply the *Hall's Marriage Theorem* on  $H$ , and this in turn ensures the existence of a *perfect matching*. This matching contains  $n$  edges, being the largest matching possible, and which can be computed with the Kuhn-Munkres algorithm. Finally, the latter match may be encoded by the permutation  $P$ , which is the one selected for decomposing  $K$ .

Then, the corresponding  $\lambda > 0$  is selected among the minimum of the components determined by  $P$ , that is  $\lambda = \min_{i,j=1,\dots,n} \{K_{ij} : P_{ij} \neq 0\}$ . Such a selection ensures the non-negativity of the matrix  $M = K - \lambda P$  whose rows and columns sum to  $1 - \lambda$ . Therefore,  $K' = \frac{1}{1-\lambda} M$  is a *DSM*. In addition, from the  $n$  components of  $K_{ij}$  satisfying  $P_{ij} = 1$ , those with value  $\lambda$  will be zero in  $M$  and, obviously, in  $K'$ . This implies that the number of non-zero components will decrease with each step and, thus, the number of steps is bounded by  $n^2$ .

For instance, the *BvN* decomposition of  $B_*$  in  $\mathcal{B}_3$  using the latter algorithm is given by:

$$\begin{aligned}
 K &= \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \frac{2}{3} \frac{1}{(1-\frac{1}{3})} \begin{bmatrix} 0 & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & 0 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \frac{2}{3} \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} = \\
 &= \frac{1}{3} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \frac{2}{3} \left\{ \frac{1}{2} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \right\} = \frac{1}{3} P_1 + \frac{1}{3} P_4 + \frac{1}{3} P_6
 \end{aligned}$$

### 4.2 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  $\beta$ -graph complexity of  $G$  as

$$\mathcal{C}_\beta(G) = \frac{H(\mathcal{P})}{\log_2 n} = \frac{\log_2 \gamma + D(\mathcal{P} || \mathcal{U}_\gamma)}{\log_2 n}, \tag{4}$$

where  $\mathcal{P} = \{p_1, \dots, p_\gamma\}$  is the probability density function (pdf) induced by the decomposition,  $H(\cdot)$  the entropy and  $D(\cdot)$  the Kullback-Leibler divergence

$D(\mathcal{P}||\mathcal{Q}) = \sum_{\alpha} p_{\alpha} \log \frac{p_{\alpha}}{q_{\alpha}}$ . The latter definition considers: a) the size of the graph, b) the number of components of the decomposition and c) the information content of the induced pdf. Moreover, as the pdf induced by the complete graph  $C_n$  is  $\mathcal{U}_n$ , our definition of graph complexity is actually the entropy ratio  $H(\mathcal{P})/H(\mathcal{U}_n)$ . Independent of  $n$ , we have  $\mathcal{C}_{\beta}(C_n) = 1$ , and as a result a complete graph has unit  $\beta$ -graph complexity. Also,  $\mathcal{C}_{\beta}(I_n) = 0$  where  $I_n$  is the graph with all its  $n$  vertices isolated (without neighbors).

As we will show in the experimental section,  $\beta$ -graph complexity is *quasi-invariant* to graph permutations, that is  $\mathcal{C}_{\beta}(G) \approx \mathcal{C}_{\beta}(Q^T A Q)$ , where  $A$  is the adjacency matrix of  $G$  and  $Q$  any permutation of order  $n$ . Quasi-invariance is fulfilled despite the fact that the *BvN* decomposition does not yield such invariance in the coefficients. As we will see, the intra-class variability is also useful for discriminating graph classes. However, such a definition is  $\beta$ -dependent and a more complete multi-dimensional descriptor is more convenient to describe the full complexity of the graph. Thus, we define the *BvN graph complexity* as

$$\mathcal{C}(G) = \left\{ \begin{array}{l} [\mathcal{C}_{\beta_1}(G), \dots, \mathcal{C}_{\beta_r}(G)]^T \in R^r \\ \beta_1 \equiv \beta_{max} \geq \beta_2 \geq \dots \geq \beta_r \equiv \beta_{min} \\ \beta_{max} = \min\{\beta \in R^+ : \mathcal{C}_{\beta}(G) = 1\} \\ \beta_{min} = \max\{\beta \in R^+ : \mathcal{C}_{\beta}(G) = 0\} \end{array} \right\}. \tag{5}$$

The latter *complexity profile*, is thus constructed by considering that when it is treated as a function we have

$$\lim_{\beta \rightarrow \infty} \mathcal{C}_{\beta}(G) = 1 \text{ and } \lim_{\beta \rightarrow 0} \mathcal{C}_{\beta}(G) = 0 \forall G. \tag{6}$$

Moreover,  $\beta_{max}$  and  $\beta_{min}$  are computed respectively from  $\Lambda_{max} = \Phi^T K(C_n) \Phi$  and  $\Lambda_{min} = \Phi^T K(I_n) \Phi$ , given that for  $\beta = 1$  we have that  $K(G) = \exp(-\mathcal{L}) \equiv \Phi \Lambda \Phi^T$  where  $\Phi$  the eigenvector matrix of  $-\mathcal{L}$ ,  $\Lambda = \text{diag}(e^{-\lambda_n}, e^{-\lambda_{n-1}}, \dots, e^{-\lambda_1} = 1)$ , and where  $\lambda_1 = 0 \leq \lambda_2 \leq \dots \leq \lambda_n$  are the eigenvalues of  $\mathcal{L}$ . More precisely, we have that the matrix  $\Lambda_{max}$  is of the form  $\text{diag}(e^{-\beta\lambda_n}, e^{-\beta\lambda_{n-1}}, \dots, e^{-\beta\lambda_1})$ , and thus  $\beta_{max}$  may be deduced from the greatest  $\frac{\beta\lambda_i}{\lambda_i} \in R^+ : K(C_n) \approx \exp(-\beta\mathcal{L})$ . A similar rationale yields the  $\beta_{min}$  bound. Thus, the graph complexity trace  $\mathcal{C}(G)$  is a signature of the interaction between the heat diffusion process and the structure/topology of the graph as  $\beta$  (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  $P_I = \mathcal{I}_n$  and the barycenter  $B_* = K(C_n)$ . In addition, the typical signature is heavy tailed, non-monotonically increasing from  $\beta_{max}$  to  $\beta^* \approx \arg \max\{\mathcal{C}_{\beta}(G)\}$  and monotonically decreasing from  $\beta^*$  to  $\beta_{min}$ . When  $\mathcal{C}_{\beta}(G)$  is  $r$ -discretized, each  $k \in \{1, \dots, r\}$  is seen as a *percentile*. The interval  $[\beta_{max}, \beta^*)$  reflects intra-class variability, given that it encodes global aspects of heat diffusion, whereas the interval  $[\beta^*, \beta_{min}]$  reflects inter-class one, because it encodes local (more constrained as we tend to  $\beta_{min}$ ) diffusion. Both intervals  $[\beta_{max}, \beta^*) \cup [\beta^*, \beta_{min}]$  may be discretized

individually with the same rate, and then fuse both fragments of the complexity function for the purpose of pattern analysis (see Fig. 2). All profiles are aligned according to Eq. 6. In addition, the value of  $\beta^*$  may be easily approximated by accounting for the appearing of a increasing number of  $K_{ii} \approx 1$  (the decay starts just before that point). Thus,  $\beta^*$  represents the most significant *topological phase transition* regarding the impact of the diffusion process in the topology of the input graph.

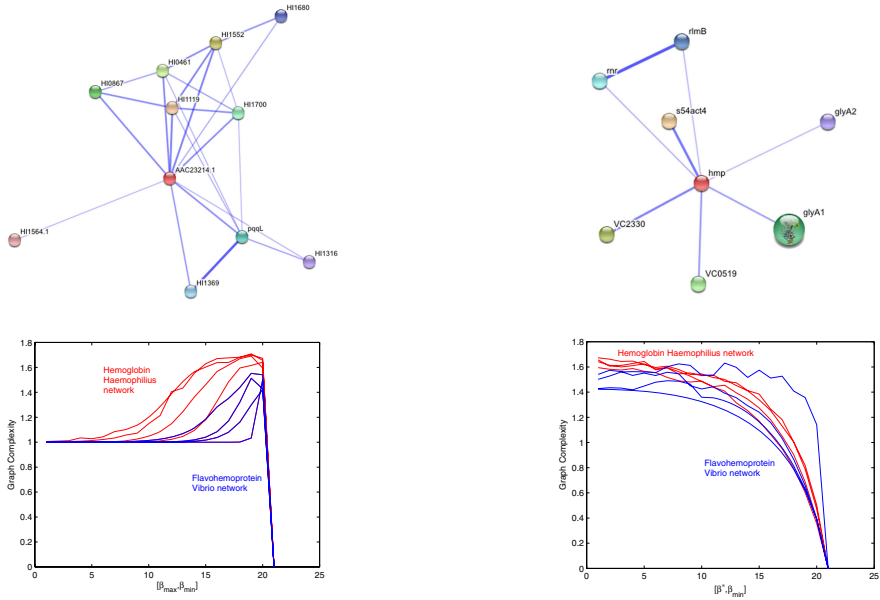
### 4.3 Connections with Matrix Permanents

The observation that the interval  $[\beta^*, \beta_{min}]$  is populated by an increasing number of elements of the heat kernel for which  $K_{ii} \approx 1$ , or equivalently an increasing number of elements for which  $K_{ij} = 0, i \neq j$  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 to a given heat kernel  $K_{\beta_{ij}}$  such that  $A_{ij} = 1$  if  $K_{\beta_{ij}} > 0$ , for a given value of  $\beta$ . The *permanent* of  $A$  is

$$per(A) = \sum_{\pi \in \mathcal{S}_n} \prod_{i=1}^n A_{i\pi(i)}, \tag{7}$$

being  $\mathcal{S}_n$  the set of permutations of  $\{1, 2, \dots, n\}$  and  $\pi(i)$  is the  $i$ -th element of the  $\pi$  permutation. As each product can be only 1 when a perfect matching exists in the bipartite graph induced by  $A$ , the overall sum counts the number of perfect matchings in the such bipartite graph. As  $\lim_{\beta \rightarrow \beta_{min}} K_{\beta} = I_n$  the minimum number of perfect matchings in the interval  $[\beta, \beta_{min}]$  is 1. However, preliminary experiments show that  $per(A)$  is nearly constant over the latter interval and sometimes close to its maximum value  $n!$ , the one for  $A = C_n$  (complete graph), before falling to 1. Consequently the  $per(A)$  profile is it not qualitatively related to the one of  $\mathcal{C}$ . However, this is not true if we consider  $per(K_{\beta})$  that yields the *sum of weighted perfect matchings*.

The van der Waerden conjecture states that:  $\forall B \in \mathcal{B}_n : per(B) \geq per(B_*) \equiv n!/n^n$  and  $per(B) = per(B_*) \Leftrightarrow B = B_*$ . The proofs solving the latter conjecture (which embodies a minimization problem) reveal that the permanent is a *mixed discriminant* of diagonal quadratic forms. Mixed discriminants are considered generalizations of the permanent. For the sake of polytopal complexity it is useful to note that  $\forall B \in \mathcal{B}_n : 0 < per(B) \leq 1$  and  $per(B) = 1 \Leftrightarrow B$  is a permutation matrix. Thus, an alternative definition of graph complexity relies on  $\mathcal{C}'_{\beta} = 1 - per(K_{\beta})$  defining the  $\mathcal{C}'(G)$  profile in  $[\beta_{max}, \beta_{min}]$ . However, the latter profile is typically flat and equal to 1 along  $[\beta_{max}, \beta^*]$  and it is only informative at second half of  $[\beta^*, \beta_{min}]$  see (Fig. 3). In addition, the computation of the permanent is a  $\#-P$  problem and one of the best approximate algorithms [12] takes  $O(n^{23})$  on average only for generating a sample of the Markov chain in which it it relies. However, when small graphs are available or complex

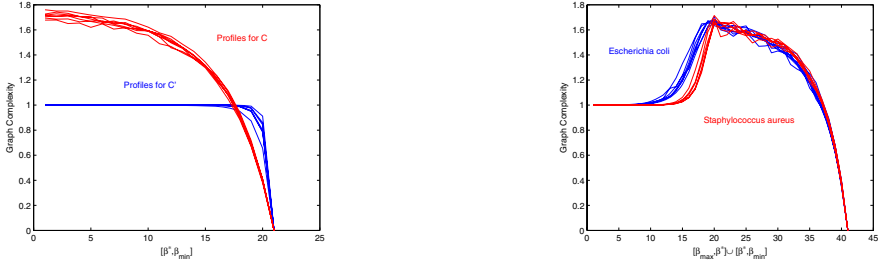


**Fig. 2.** Top: Examples of protein-protein interaction networks (STRING [14]) for hemoglobin in *Haemophilus* (left) and flavohemoprotein in *Vibrio* (right). Bottom: Complexity profiles for both nets in the ranges  $[\beta_{max}, \beta_{min}]$  (left) and  $[\beta^*, \beta_{min}]$  (right).

graphs are simplified (see below) we may use the Ryser’s algorithm which takes  $\Theta(n2^n)$  operations.

### 4.4 Non-metric Graph Simplification

For a graph  $G = (V, E)$  with  $|V| = n$  vertices, the *BvN* decomposition process takes  $O(n^2)$  iterations and in each operation solves a Kuhn-Munkres (Hungarian) search for a suitable permutation, which takes  $O(n^3)$ . Hence, we have a global complexity of  $O(n^5)$  which is not too expensive for  $n = 30$  nodes, but becomes unmanageable for  $n \geq 70$ , taking into account that this is the complexity for each percentile. For this reason, in this paper we extend the *stratified graph simplification* method proposed in [13] to the *non-metric* case. More precisely,  $G$  is augmented with one heat-source node  $\tau$  which is connected with all  $v \in Periphery(G)$ . The result is the so-called *affixation graph*  $\mathcal{A}(G) = (V \cup \{\tau\}, E \cup \{(\tau, u), (u, \tau), \forall u \in Periphery(G)\})$ . In our case, the  $Periphery(G)$  is not the convex hull of points but it is defined by computing all vertices  $u$  having as excentricity  $e(u) = \max_{v \in V} d(u, v)$  (where  $d(u, v)$  is the minimum path length between vertices  $u$  and  $v$ ) equal to the diameter of graph  $Diam(G) = \max_{u \in V} e(u)$ . From that point we compute the *average path lengths*  $D(\tau, u)$  using the normalized Laplacian and perform unsupervised clustering to



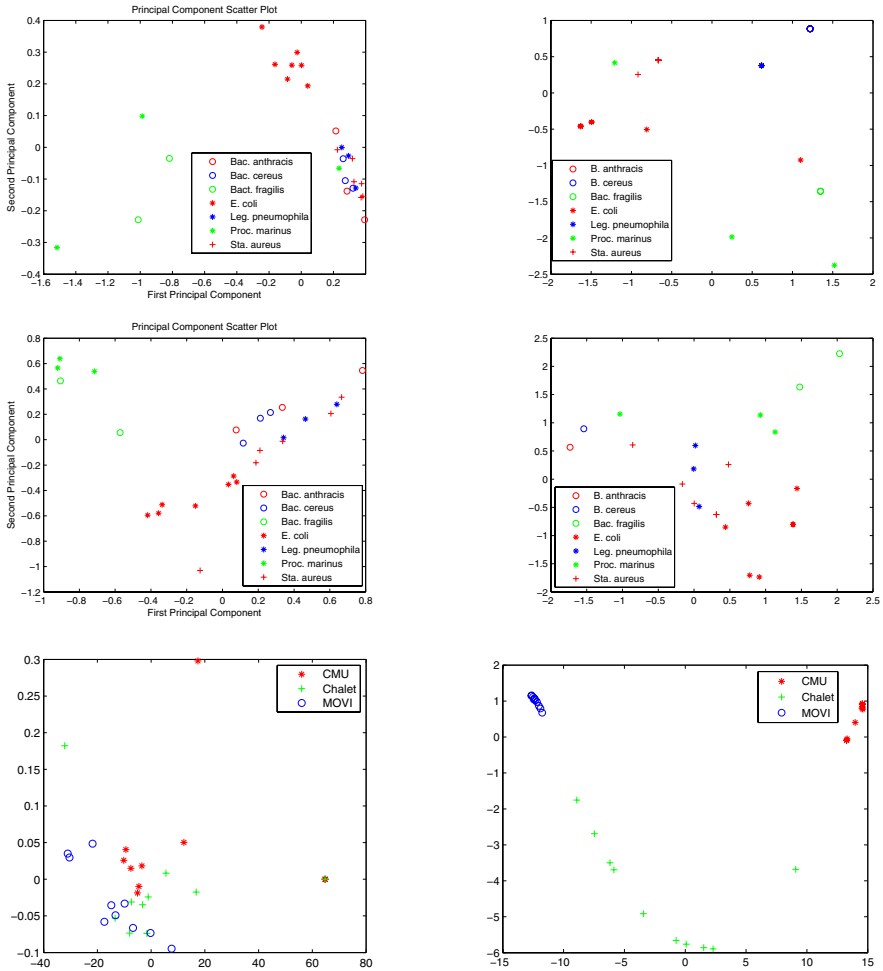
**Fig. 3.** Left: Polytopal vs Permanent-based complexity considering  $[\beta^*, \beta_{min}]$ . Right: Complete traces of  $[\beta_{max}, \beta^*] \cup [\beta^*, \beta_{min}]$  considering same discretization.

find layers in the graph. We assign a node  $l$  to each cluster and we connect nodes with adjacency relations in  $\mathcal{A}$ . So,  $\mathcal{S}(G)$  is the resulting *simplified graph*.

## 5 Experimentation

First, we report the testing of *quasi-invariance*  $C_\beta(G) = C_\beta(Q^T A Q) \pm e(Q)$ , being  $e(Q)$  an error term. For the CMU, MOVI and Chalet sequences we have a *median absolute error* (for the sake of robustness) of 0.0215, 0.0167 and 0.0150 respectively, that is  $O(2 \times 10^{-2})$ , when considering 50 permutations for each of the 30 *simplified* graphs. In this case we set  $\beta = 1$ . Second, we have tested the proposed measure in the context of *graph embedding*. In Fig. 4(left) we show the results of PCA/MDS embeddings derived from the complexity and compare them with the embeddings given by the leading eigenvalues [2] showed in Fig. 4(right). In the first two rows we perform embedding over protein-protein interaction networks from STRING. More precisely, for networks emanating from the *trpA* (Tryptophan synthase alpha chain) in 27 sub-species corresponding to 7 species (classes). In this case, interactions account for similar genetic context, and consider two cases: networks of 10 interactors with *trpA* (11 nodes) and networks of 20 ones. Our complexity method (right) allows to discriminate most of the species except *Bacillus* in the 10 interactors case. However, they are segregated when we consider 20 interactors. In the same cases, the leading eigenvalues method confuses all classes (although it works better with 20 interactors). No simplification was applied. However, in the last row, where we show the embedding of graphs from de CMU-Chalet-MOVI sequence, graphs are simplified with the method described above. In this latter case, since simplification may yield small graphs, the number of eigenvectors used for spectral embedding, the result is a complete overlap of the three graph classes. This overlap does not appear when we consider the original graphs. So in summary, polytopal complexity seems to yield good embeddings in conditions of low structural variability where the leading eigenvectors method fails, and acceptable results after applying simplification.





**Fig. 4.** Graph embedding based on  $\mathcal{C}$  vs leading eigenvalues. First row: Protein-interaction nets with 10 interactors. Second row: with 20 interactors. Third row: for graphs of the CMU-Chalet-MOVI sequence.

## 6 Conclusions

In this paper, we have proposed and successfully tested a novel measure of graph complexity. We have also extended a recently reported method of graph simplification to the non-metric case. The main question for the future is the formal characterization of the main phase transition defined by the  $\beta^*$  value. An attempt at answering this question has been provided in this paper, through exploring the connections between polytopal complexity and matrix permanents. Here the progressive loss of ability to form perfect matchings seems to explain the observed behaviour. However a complete characterization of the phase tran-

sition in terms of entropy change is a topic for future research. We will also compare this method with others in the literature.

## References

1. Robles-Kelly, A., Hancock, E.R.: A riemannian approach to graph embedding. *Pattern Recognition* (40), 1042–1056 (2007)
2. Luo, B., Wilson, R.C., Hancock, E.: Spectral embedding of graphs. *Pattern Recognition* (36), 2213–2223 (2003)
3. Shokoufandeh, A., Dickinson, S., Siddiqi, K., Zucker, S.: Indexing using a spectral encoding of topological structure. In: *IEEE ICPR*, pp. 491–497
4. Torsello, A., Hancock, E.: Learning shape-classes using a mixture of tree-unions. *IEEE Trans. on PAMI* 28(6), 954–967 (2006)
5. Lozano, M., Escolano, F.: Protein classification by matching and clustering surface graphs. *Pattern Recognition* 39(4), 539–551 (2006)
6. Körner, J.: Coding of an information source having ambiguous alphabet and the entropy of graphs. In: *Trans. of the 6th Prague Conference on Information Theory*, pp. 411–425 (1973)
7. Escolano, F., Hancock, E., Lozano, M.: Birkhoff polytopes, heat kernels, and graph embedding. In: *ICPR* (2008)
8. Birkhoff, G.D.: Tres observaciones sobre el algebra lineal. *Universidad Nacional de Tucuman Revista, Serie A* 5, 147–151 (1946)
9. Chang, C., Chen, W., Huang, H.: On service guarantees for input buffered crossbar switches: A capacity decomposition approach by birkhoff and von neumann. In: *IEEE IWQoS*, pp. 79–86 (1998)
10. Kondor, R.I., Lafferty, J.: Diffusion kernels on graphs and other discrete structures. In: *Proc. ICML* (2002)
11. Mirsky, L.: Proofs of two theorems on doubly stochastic matrices. *Proc. Amer. Math. Soc.* 9, 371–374 (1958)
12. 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)
13. Qiu, H., Hancock, E.: Graph simplification and matching using commute times. *Pattern Recognition* (40), 2874–2889 (2007)
14. von Mering, C., Huynen, M., Jaeggi, D., Schmidt, S., Bork, P., Snell, B.: String: a database of predicted functional associations. *Nuc. Acid Res.* 31, 258–261 (2003)