



**HAL**  
open science

# A Joint Spectral Similarity Measure for Graphs Classification

Hadj-Ahmed Bay-Ahmed, Abdel Boudraa, Delphine Dare-Emzivat

► **To cite this version:**

Hadj-Ahmed Bay-Ahmed, Abdel Boudraa, Delphine Dare-Emzivat. A Joint Spectral Similarity Measure for Graphs Classification. Pattern Recognition Letters, 2019, 120, 10.1016/j.patrec.2018.12.014 . hal-02138297

**HAL Id: hal-02138297**

**<https://hal.science/hal-02138297>**

Submitted on 23 Jul 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A Joint Spectral Similarity Measure for Graphs Classification

Hadj-Ahmed **BAY-AHMED**, Abdel-Ouahab **BOUDRAA\*\***, Delphine **DARE-EMZIVAT**

<sup>a</sup>IRENav (EA 3634), Ecole Navale/Arts-Metiers ParisTech, CC 600, 29240 Brest Cedex 9, France.

## ABSTRACT

In spite of the simple linear relationship between the adjacency  $A$  and the Laplacian  $L$  matrices,  $L=D-A$  where  $D$  is the degrees matrix, these matrices seem to reveal informations about the graph in different ways, where it appears that some details are detected only by one of them, as in the case of cospectral graphs. Based on this observation, a new graphs similarity measure, referred to as joint spectral similarity (JSS) incorporating both spectral information from  $A$  and  $L$  is introduced. A weighting parameter to control the relative influence of each matrix is used. Furthermore, to highlight the overlapping and the unequal contributions of these matrices for graph representation, they are compared in terms of the so called Von Neumann entropy (VN), connectivity and complexity measures. The graph is viewed as a quantum system and thus, the calculated VN entropy of its perturbed density matrix emphasizes the overlapping in terms of information quantity of  $A$  and  $L$  matrices. The impact of matrix representation is strongly illustrated by classification findings on real and conceptual graphs based on JSS measure. The obtained results show the effectiveness of the JSS measure in terms of graph classification accuracies and also highlight varying information overlapping rates of  $A$  and  $L$ , and point out their different ways in recovering structural information of the graph.

## 1. Introduction

Graph spectral analysis is one of the hot topics in data processing community, motivated by the prominent need to develop new mathematical tools to process networked and structured data. These data are generated from various sources, as sensor, social, biological or transportation networks, where the information resides in complex and irregular structures. For this purpose, eigen-spectrum of matrices associated with graphs are often closely studied. Recent works of the literature have emphasized the importance of matrix representations for graph characterization, pointing out the advantages and the drawbacks of some spectra associated with graphs [1],[2], including, those of adjacency ( $A$ ), Laplacian ( $L$ ), signless Laplacian  $|L|$  and distance ( $D^G$ ) matrices. The spectrum of  $L$  matrix is indeed widely studied in spectral graph theory [3], in reason of the symmetry and positive semi-definiteness of the matrix, which is useful for determining cuts and inherent

graph components. Otherwise, the spectrum of  $A$  matrix is mainly used for the study of regularity [4], isomorphisms [5] and bipartition [6] of graphs. The question of choosing either  $A$  or  $L$  matrix for graph representation is still a subject of debate. For instance, in graph signal processing theory, Shuman *et al.* [7] define the graph Fourier basis as the eigenbasis of  $L$  matrix, while Sandryhaila and Moura [8] prefer the eigenbasis of  $A$  obtained via a Jordan decomposition. This difference can be justified in part by the nature itself of the decomposition basis, and also by the fact that not all graphs are determined by their spectra and there is a family of graphs that shares the same spectrum in respect to some matrix representation, commonly called cospectral graphs [9]. The distinction between these cospectral graphs or between very similar graphs via their spectra is a tough task as stated in [1]. A big part of graph comparison algorithms aims to sort them using some structural similarity criteria without implicitly resorting to spectral analysis. This is the case of the aligned subtree kernel [10] that incorporates explicit subtree correspondences between the compared graphs, as well as other substructure based kernels [11] [12] [13] [14], which decompose the graphs into smaller sub-graphs and try to find the optimal bijection between them.

\*\*Corresponding author: Tel.: 33 2 98 23 40 37;

e-mail: abdel.boudraa@ecole-navale.fr (Abdel-Ouahab BOUDRAA)

In spite that, spectral similarity measures remain attractive and hold many strengths [2]. We propose in this work to tackle the problem of graph classification using a spectral measurement of similarity called Joint Spectral Similarity (JSS) which compares two graphs using jointly the spectra of A and L. Such measure is interesting to distinguish graphs that have close spectral properties or even for sorting cospectral graphs which is a hard task when using only the spectrum of A or L at once. To understand the contribution of each matrix in graph information recovering, we investigate the framework of structural complexity measurements of graphs, more specifically, energetic and entropic measurements [15]. Inspired from information theory and statistical mechanics, these measures capture similarities and differences between networks and quantify the organization level of the underlying graph structure. Among them, we focus on Laplacian graph energy and VN entropy that can be used as a measure of regularity in graphs [16],[17].

In this paper, we show the effectiveness of the JSS similarity measure for graphs classification. We highlight the graph's representation disparity between A and L matrices illustrated via entropy, connectivity and complexity measures. Combined in a heat kernel, the JSS measure shows promising results in real world graph data, built from chemical components and in real time series. These results are compared to the state-of-art graph kernels in terms of classification accuracy and computing CPU time.

## 2. Graphs basics

An undirected graph is defined with a finite non-empty set  $V$  of  $N$  elements called vertices (or nodes), and with a set  $E$  of  $M$  unordered pairs of distinct elements of  $V$  representing the connexions between vertices. A labelled graph with  $N$  vertices and  $M$  edges is represented with an  $(N \times N)$  A matrix ( $A = [a_{ij}]$ ), where the element  $a_{ij}$  is one if the vertex  $v_i$  is connected to the vertex  $v_j$  and zero otherwise. Due to the symmetric relationship between nodes in undirected graphs, their adjacency matrix is also symmetric with zeros as diagonal entries. An alternative representation of graphs is given by Laplacian matrix  $L = D - A$ , where  $D$  is a diagonal matrix in which, the  $i^{th}$  entry is the degree  $d_i$  of the node  $v_i \in V$ , representing its number of neighbors. In the case of a weighted graph, the degree  $d_i$  is equal to the sum of the weights  $w_{ij}$  of edges related to  $v_i$ , and the Laplacian matrix takes the form  $L = D - A$ , where  $A$  is the weighted A matrix.

## 3. Graphs representation via A and L spectra

The matrices A and L are compared in terms of entropic information, connectivity and complexity measures. A proposition is also formulated, giving corrective terms in the VN entropy expression due to changes in graph structure related to A and L.

### 3.1. Entropic information measure

In quantum mechanics, objects modify their states according to the presence or not of an observer. They take only one state

among other possible states, called pure states. But, it happens also that the observer measures a mixture of states, then the object is said to be in a mixed state or in a superposition. Such phenomenon is usually studied using the so called density matrix  $\rho$ , which has the occurrence probability of each pure state in the mixture as diagonal elements, and the confusion degree between them in the off-diagonal elements. This matrix is symmetric and positive semi-definite of trace one. To measure the uncertainty degree (mixedness) of a quantum system in a superposition state, Von Neumann introduced an entropic measure [16] which takes the form  $-Tr(\rho \log_2 \rho)$ , which includes the diagonal elements of  $\rho$  and determines the amount of unknown or missing information about the real state of the system. The off-diagonal elements are interpreted as confusion or noise degree in the mixed state. The link between graphs and quantum mechanics has been established in [18], [19]. The beginning was from the fact that L matrix is symmetric (for undirected graph) and positive semi-definite. Hence, once normalized in appropriate way, it can be treated as a density matrix of a quantum system in a superposition state. Considering nodes as pure states in the standard basis of an Hilbert space of dimension  $N$ , the graph is seen as a superposition (mixture) of those pure states (nodes). Therefore, the VN entropy could be then calculated using L matrix scaled by the degree-sum of the graph:  $\rho = \frac{L}{\sum_i d_i}$ .

Let  $\lambda_N \geq \dots \geq \lambda_2 \geq \lambda_1 = 0$  and  $\tilde{\lambda}_N \geq \dots \geq \tilde{\lambda}_2 \geq \tilde{\lambda}_1 = 0$  be respectively the eigenvalues of L and  $\rho$ , the VN entropy takes the form :

$$S(\rho) = - \sum_{i=1}^N \tilde{\lambda}_i \log_2 \tilde{\lambda}_i = -Tr(\rho \log_2 \rho) \quad (1)$$

with  $0 \log_2(0) = 0$  by convention. Otherwise, Bai *et al.* [20] [21] used an alternative density matrix  $\sigma$  based on both continuous-time and discrete-time quantum walk to compare graphs. Their idea is to build a kernel which compares the density matrices associated to graphs using the Quantum Jensen-Shannon divergence as a similarity measure:  $D_{QJS}(\sigma_1, \sigma_2) = S[1/2(\sigma_1 + \sigma_2)] - 1/2 S(\sigma_1) - 1/2 S(\sigma_2)$ , where  $\sigma_1, \sigma_2$  are a pair of density matrices. The density matrix  $\sigma$  is calculated using a transition matrix for the quantum walk which includes information about the degrees and the structure of the graph. This approach does not require any calculation of eigenvalues nor those of A neither of L. Since the graph VN entropy is interpreted as a measure of the information content, we wonder about the contribution of each matrix (A and L) in the measured quantity of information. A part of the answer comes from quantum perturbation theory where the sensitivity of eigenvalues to density matrix perturbations has been studied. Some of these studies have been extended by Chen to include the case of perturbations on VN entropy [22]. This opens the way to project the study on to the general case of graphs. The goal is to determine the cost of structural modifications of the graph in terms of information quantity, and to clearly understand the relationship between A and L in a purely entropic framework. To this end, A matrix is perturbed and the VN entropy of the perturbed density matrix  $\rho$  is computed.

**Proposition 1.** *Let G be a weighted graph whose the unper-*

turbed density matrix is  $\rho_0$ . If  $\rho_0$  undergoes a perturbation  $\rho^*$ , the VN entropy of the resulting perturbed matrix,  $\rho = \rho_0 + \xi \rho^*$ , expanded up to the second order is given by

$$S(\rho) = S(\rho_0) + \eta_{\text{Diagonal}}(\rho^*) + \eta_{\text{Off-Diagonal}}(\rho^*) \quad (2)$$

where

$$\begin{aligned} \eta_{\text{Diagonal}}(\rho^*) &= -\xi \sum_n \rho_{nm}^* \log_2 \tilde{\lambda}_n - \frac{1}{2} \xi^2 \sum_n \frac{(\rho_{nm}^*)^2}{\tilde{\lambda}_n} + o(\xi^2) \\ \eta_{\text{Off-Diagonal}}(\rho^*) &= -\xi^2 \sum_n \sum_{m \neq n} \frac{|\rho_{nm}^*|^2}{\tilde{\lambda}_n - \tilde{\lambda}_m} \log_2 \tilde{\lambda}_n + o(\xi^2) \end{aligned} \quad (3)$$

with  $\xi$  is a scalar parameter supposed to be small, and  $\tilde{\lambda}_n \neq \tilde{\lambda}_m \forall n \neq m$  are the eigenvalues of  $\rho_0$ .

**Proof :**

Since  $\rho$  is a density matrix, the perturbations are introduced in such way that  $\text{Tr}(\rho) = 1$  and  $\rho = \rho^T$ . Restricting Taylor expansion to the second order [17], the corrections to the eigenvalues are written as :

$$\tilde{\lambda}_n^{(t)} = \tilde{\lambda}_n + \xi \tilde{\lambda}_n^{(1)} + \xi^2 \tilde{\lambda}_n^{(2)} + o(\xi^2) \quad (4)$$

where the first and second order eigenvalue perturbations are given by

$$\tilde{\lambda}_n^{(1)} = \rho_{nm}^*, \quad \tilde{\lambda}_n^{(2)} = \sum_{m \neq n} \frac{|\rho_{nm}^*|^2}{\tilde{\lambda}_n - \tilde{\lambda}_m} \quad (5)$$

Therefore, the perturbed VN entropy can be written as

$$S(\rho) = -\text{Tr}(\rho \log_2 \rho) \approx -\sum_n \tilde{\lambda}_n^{(t)} \log_2 \tilde{\lambda}_n^{(t)} \quad (6)$$

On the other hand, the Taylor expansion of  $S(\rho)$  around zero up to second order is written as

$$S(\rho) \approx S(\rho_0) + \xi \frac{dS(\rho_0)}{d\xi} + \frac{1}{2} \xi^2 \frac{d^2 S(\rho_0)}{d\xi^2} + o(\xi^2) \quad (7)$$

Chen in [22] gives the first and second order derivatives of  $S(\rho_0)$

$$\frac{dS(\rho_0)}{d\xi} = -\sum_n \rho_{nm}^* \log_2 \tilde{\lambda}_n \quad (8)$$

$$\frac{d^2 S(\rho_0)}{d\xi^2} = -\sum_n \frac{(\rho_{nm}^*)^2}{\tilde{\lambda}_n} - 2 \sum_n \tilde{\lambda}_n^{(2)} \log_2 \tilde{\lambda}_n \quad (9)$$

By substituting (5), (8) and (9) in equation (7), we get

$$\begin{aligned} S(\rho) \approx S(\rho_0) &- \underbrace{\xi \sum_n \rho_{nm}^* \log_2 \tilde{\lambda}_n - \frac{1}{2} \xi^2 \sum_n \frac{(\rho_{nm}^*)^2}{\tilde{\lambda}_n}}_{\eta_{\text{Diagonal}}} \\ &+ \underbrace{\xi^2 \sum_n \sum_{m \neq n} \frac{|\rho_{nm}^*|^2}{\tilde{\lambda}_n - \tilde{\lambda}_m} \log_2 \tilde{\lambda}_n}_{\eta_{\text{Off-Diagonal}}} + o(\xi^2) \end{aligned} \quad (10)$$

A careful examination of equation (10) shows that the corrective terms  $\eta_{\text{Diagonal}}$  and  $\eta_{\text{Off-Diagonal}}$ , include respectively both diagonal and off-diagonal perturbation elements. Since changes

in edge weights of A lead to changes in node degrees, perturbations affect at first place off-diagonal elements of A. Thus, it is expected that changes in weights introduce only off-diagonal elements in the entropy expression (10). But, diagonal elements related to node degrees changes appear also. Due to the linear relationship  $L=D-A$ , these diagonal elements are introduced to L via D and thus they may explain why L matrix is more sensitive to structural changes of the graph than A.

### 3.2. Connectivity measure

In spectral graph theory, L matrix is often used for structural properties study of graphs. This is essentially motivated by its algebraic characteristics as the non-negativeness of its eigenvalue spectrum, or its interesting quadratic form written as  $\mathbf{y}^T \mathbf{L} \mathbf{y} = \sum_{u \sim v} (\mathbf{y}(u) - \mathbf{y}(v))^2$ , with  $\mathbf{y}$  is a function that assigns to each vertex  $v$  of the graph a real value  $\mathbf{y}(v)$ , and  $\sum_{u \sim v}$  denotes the sum over all unordered pairs  $(u, v)$  for which  $u$  and  $v$  are adjacent. This quadratic form is useful for getting a variational characterizations of the eigenvalues in terms of *Rayleigh quotient* [3]. The eigenpairs  $(\lambda_k, \mathbf{f}_k)$  of L can be formulated as convex optimization problems

$$\lambda_k = \min_{\substack{\mathbf{y} \perp \mathbf{f}_1, \dots, \mathbf{f}_{k-1} \\ \mathbf{y} \neq \mathbf{0}}} \frac{\mathbf{y}^T \mathbf{L} \mathbf{y}}{\mathbf{y}^T \mathbf{y}} = \min_{\substack{\mathbf{y} \perp \mathbf{f}_1, \dots, \mathbf{f}_{k-1} \\ \mathbf{y} \neq \mathbf{0}}} \frac{\sum_{u \sim v} (\mathbf{y}(u) - \mathbf{y}(v))^2}{\sum_v \mathbf{y}(v)^2} \quad (11)$$

$$\mathbf{f}_k = \arg \min_{\substack{\mathbf{y} \perp \mathbf{f}_1, \dots, \mathbf{f}_{k-1} \\ \mathbf{y} \neq \mathbf{0}}} \frac{\mathbf{y}^T \mathbf{L} \mathbf{y}}{\mathbf{y}^T \mathbf{y}} = \arg \min_{\substack{\mathbf{y} \perp \mathbf{f}_1, \dots, \mathbf{f}_{k-1} \\ \mathbf{y} \neq \mathbf{0}}} \frac{\sum_{u \sim v} (\mathbf{y}(u) - \mathbf{y}(v))^2}{\sum_v \mathbf{y}(v)^2} \quad (12)$$

It can be easily deduced from equation (11) that 0 is an eigenvalue of L corresponding to the constant eigenvector  $\mathbf{f}_1 = \mathbf{1}$ . The second eigenvector must be orthonormal to the constant eigenvector ( $\langle \mathbf{f}_2, \mathbf{1} \rangle = \sum_v \mathbf{f}_2(v) = 0$ ). Therefore, a null second eigenvalue implies that the graph is composed of two disconnected components. This leads to the fundamental result of graph algebra which states that the multiplicity of zero eigenvalue corresponds to the number of the disconnected clusters in the graph. In connected graphs, the second eigenvalue is non-zero and corresponds to the minimal cost of the connections which can be lost in case of segmentation of the graph in two connected sub-graphs. This quantity has been defined by Fiedler as the algebraic connectivity of the graph, the corresponding second eigenvector is called Fiedler vector [23]. In the following example, we consider an undirected and unweighed random graph that contains implicitly two communities. We observe that the A matrix hides almost all indications about the existence of such communities or at least they remain difficult to detect (Fig. 2(a)). Fig. 1(b) shows that the Fiedler vector of L matrix associates to the nodes values of opposite signs in order to cluster them into subgroups. Using those labels, we reordered the rows and columns of A matrix in a way that reveals more structural information about the graphs (Fig. 2(b)). This example illustrates the interest of L matrix for recovering homogeneous clusters of nodes.

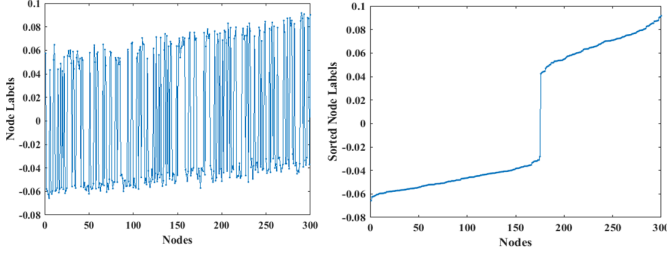


Fig. 1: Laplacian's second eigenvector (Fiedler vector) of a connected random graph with 300 nodes. In the left: labels associated to nodes before sorting (a). In the right: After sorting (b).

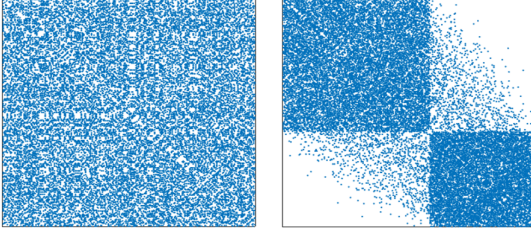


Fig. 2: Adjacency matrix of the connected random graph with 300 nodes. In the left: the original A matrix (a). In the right: the reordered A matrix using the Fiedler labels order (From L matrix) (b).

### 3.3. Complexity measure

Based on facts observed in molecular theory, Gutman *et al.* [24] introduced the A based graph energy (AE) as

$$AE(G) = \sum_{k=1}^N |\mu_k|, \quad LE(G) = \sum_{k=1}^N \left| \lambda_k - \frac{2M}{N} \right| \quad (13)$$

where  $\mu_k$  are the eigenvalues of A. This spectrum-based graph invariant has been largely studied in both chemical and mathematical literature. An AE-like quantity LE, defined in terms of Laplacian and preserving the main features of the original AE has been proposed in [25]

where  $\lambda_k$  are the eigenvalues of L and  $2M/N$  is the average vertex degree. Both LE and AE measure the graph complexity. We illustrate this fact by using Erdős-Rényi random graphs parametrized with a connectivity probability parameter  $p$ . As  $p$  increases from zero to one, the model becomes more and more likely to include graphs with more edges and less and less likely to include graphs with fewer edges. As shown in Fig. 3, the quantities LE and AE are effectively relevant for measuring the complexity of graphs. Their respective curves take a smooth upward rate when the Erdős-Rényi probability increases. The growth tendency is very different, revealing the disparity between A and L in recovering the structural changes in the graphs, thus their complexity. Moreover, according to these findings, LE estimates more accurately the connectivity of the graph, and performs better than the smallest second eigenvalue ( $\lambda_2$ ) of L in Erdős-Rényi random graphs.

### 3.4. Graphs Cospetrality Issue

The spectrum of different matrix representations of graphs holds a variety of informations that differs from one matrix to another [26]. This is seen in the existence of cosppectral

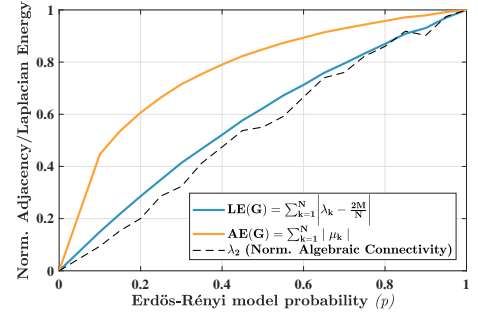


Fig. 3: A and L based graph energy evolution in terms of the graph's structural complexity. Illustrated in Erdős-Rényi random graphs with 700 nodes.

graphs, or graphs that share the same spectrum for a particular matrix, A, L or others. They are called non-DS graphs referring to "Non Determined by the Spectrum". Among many graph families, Schwenk stated that almost all trees are non-DS graphs in respect to A [27]. Furthermore, Godsil and McKay proposed a method based on edge switching to make two non-isomorphic graphs A-cospectral [9]. Likewise, many families of L-cospectral graphs were defined in literature, such those in [28]. We show in Figure 5 an example of two graphs that share the same Laplacian spectrum despite of their structural differences. As cospetrality appears to be a challenge to distinguish graphs via their spectra, it remains nevertheless surmountable by using a variety of spectral metrics simultaneously in the graph identification process. Figure 4 shows that it is possible to recognize A-cospectral graphs via their L spectrum, and inversely in Figure 5. This justifies more our idea of combining both spectra of A and L in JSS measure for comparing graphs.

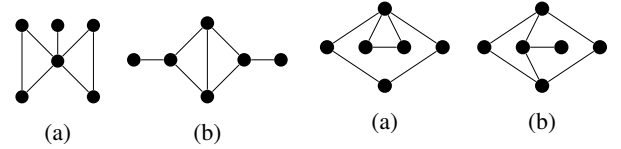


Fig. 4: Two A-cospectral graphs  
 $\sigma_a(A) = \sigma_b(A) =$   
 $[-1.9, -1, -1, 0.2, 1, 2.7]$   
 $\sigma_a(L) = [0, 1, 1, 3, 3, 6]$   
 $\sigma_b(L) = [0, 0.5, 1.2, 3.4, 4, 4.7]$

Fig. 5: Two L-cospectral graphs  
 $\sigma_a(L) = \sigma_b(L) = [0, 0.7, 2, 3, 3, 5.2]$   
 $\sigma_a(A) = [-2.1, -1, -0.5, 0, 1.2, 2.5]$   
 $\sigma_b(A) = [-2.5, -0.7, 0, 0, 0.7, 2.5]$   
 where  $\sigma(B)$  means the eigen spectrum of the matrix B.

## 4. Joint spectral similarity of graphs

Let  $\mathcal{G}$  be a set of undirected graphs, and consider two graphs  $G_1(A_1, L_1)$  and  $G_2(A_2, L_2) \in \mathcal{G}$  with  $A_1, A_2$  their adjacency matrices and  $L_1, L_2$  their Laplacian matrices where  $(\lambda_{1i}, \lambda_{2i})$  and  $(\mu_{1i}, \mu_{2i})$  are the eigenspectra of their Laplacian and adjacency matrices ordered in descending order. Information about the degree distribution is encoded mainly in the eigenvalues of L as well as the number of components of the graph, while information about walks, paths and the bipartition of the graph are in the eigenvalues of A [1]. Also to avoid the problem of cospetrality when comparing graphs via A and L spectra, we quantify the associated spectral similarity of the graphs as the trade-off between them. The introduced JSS measure aims to exploit both A and L spectral informations to better discriminate graphs. This

measure, sum of two weighted spectral components, is given by the following convex linear combination :

$$\text{JSS}(G_1, G_2) = \alpha \text{JSS}_A(G_1, G_2) + (1 - \alpha) \text{JSS}_L(G_1, G_2) \quad (14)$$

with  $\alpha \in [0, 1]$ , a weighting factor and where the components are given by :

$$\text{JSS}_L(G_1, G_2) = \sum_{i=1}^k (\lambda_{1i} - \lambda_{2i})^2, \quad \text{JSS}_A(G_1, G_2) = \sum_{i=1}^k (\mu_{1i} - \mu_{2i})^2 \quad (15)$$

where  $k = \min(N_1, N_2)$  and  $N_1, N_2$  are the numbers of eigenvalues corresponding to each graph.  $k$  represents the important common eigenvalues between the graphs, however, this criterion is particularly suitable for graphs of similar and comparable sizes. The weighting factor  $\alpha$  controls the significance of each distance and allows more importance to be given to the A-spectral distance or to the L-spectral distance.

## 5. Experimental Results

We illustrate the information conveyed by A and L matrices on real graphs (molecular data) and on conceptual ones obtained by mapping time series to the graph using *Visibility Graph* (VG) algorithm [29]. There are altogether nine benchmark real data sets used in our experiment.

### 5.1. Graphs from molecular data

Natural graphs are derived from databases built from chemical/biological data. MUTAG [30] is a data set of 188 mutagenic aromatic and heteroaromatic nitro compounds labeled according to whether or not they have mutagenic effect on the Gram-negative bacterium *Salmonella typhimurium*. NCI1 and NCI109 represent two data sets of chemical compounds screened for activity against non-small cell lung cancer and ovarian cancer cell lines [31]. ENZYMES is a data set of protein tertiary structures obtained from [32] consisting of 600 enzymes from BRENDA enzyme database [33] where the task is to assign each enzyme to one of the 6 top-level classes. D&D is a data set of 1178 protein structures [34], the classification task is to distinguish protein structures between enzymes and non-enzymes.

### 5.2. Time series data

Four problems are considered namely [35]: Computers is a data about home electricity consumption to help reduce the UK's carbon footprint, it concerns the use of computers in the houses, the purpose is to know if it is a Desktop or a Laptop computer. The ToeSegmentation data are derived from the CMU Graphics Lab Motion Capture database, walk motions in the database are classified by their nature, whether they are normal or abnormal. SonyAIBORobotSurface1 database is donated by Carnegie Mellon University. The robot Sony1 has roll/pitch/yaw accelerometers whose task is to detect the surface being walked on whether is cement or carpet. Lightning2 is a database that was built using FORTE satellite, that detects transient electromagnetic events associated with lightning. The aim is to classify power densities into two different categories of lightning.

### 5.3. Building Graphs from time series

To map time series into complex networks on the graph domain, the VG algorithm is used [29]. This algorithm defines connections between nodes and the weights of the graph from a time series. Every node of the graph corresponds, in the same order, to a sample from the series, and two nodes are connected if visibility exists between the corresponding samples. More formally, two arbitrary samples  $(t_a, y_a)$  and  $(t_b, y_b)$  will have visibility, and they are represented by two connected nodes in the associated graph, if any other sample  $(t_c, y_c)$  placed between them fulfills :

$$y_c < y_b + (y_a - y_b) \frac{t_b - t_c}{t_b - t_a} \quad (16)$$

We consider the edge weight between two nodes as the absolute value of the angle between the straight line that connects them and the horizontal axis. Formally it is written as :

$$w_{ab} = \left| \arctan \left( \frac{y_b - y_a}{t_b - t_a} \right) \right| \quad (17)$$

The weights  $w_{ab}$  are the entries of A associated to the VG graph.

### 5.4. Experimental configuration

As for Radial Basis Function (RBF) kernel, the spectral similarity measure is integrated in a classical exponential function, leading to the following distance between the graphs  $G_i, G_j$

$$K_{i,j} = \exp(-\gamma s(G_i, G_j)) \quad (18)$$

where  $K$  is the kernel matrix and  $s(G_i, G_j)$  can be JSS,  $\text{JSS}_A$  or  $\text{JSS}_L$  measure. The parameter  $\gamma$  is a smoothing factor. The kernel matrix is integrated on a support vector machine (SVM) algorithm. Next, 10-fold cross-validation approach is performed using 9 folds for training and 1 for testing. Data sets are randomly shuffled before partitioning and the whole experiment is repeated 10 times to avoid random effects of fold assignments. The kernel factor  $\gamma$  is set to 1. Average classification accuracies and their corresponding standard deviations are summarized in Tables 1 and 3. For molecular data, the measures JSS,  $\text{JSS}_A$  and  $\text{JSS}_L$  are compared to some kernels of the literature in terms of prediction accuracy and computation runtime. For time series data, only prediction accuracies are reported. Otherwise, the well-known graph kernels are tested: those based on walks, sub-trees and Weisfeiler-Lehman isomorphism. Else, the measures are compared to the fast geometric random walk kernel [36], counts common labelled walks and to  $p$ -random walk kernel that compares random walks up to length  $p$  in two graphs [37]. From sub-tree kernels, we chose Ramon Gärtner kernel [38] which compares all pairs of nodes from two graphs by iteratively comparing their neighbourhoods. From Weisfeiler-Lehman kernels, we picked up the Weisfeiler-Lehman edge kernel [39]: it counts matching pairs of edges with identically labeled endpoints in two graphs. For molecular data, computing set-up, accuracy and runtime values of the benchmark kernels are performed in [39]. We followed the same procedures to compute random walk and Weisfeiler-Lehman edge kernels for time series data, without Ramon Gärtner and  $p$ -random walk kernels because of runtime constraint. The time series

results are also compared to Linear and RBF kernels, applied directly on the samples of time series without converting them to graphs. As reported in Table 2, runtime in minutes and seconds of our method is measured using Anaconda2 4.1.1 Python 2.7.12 Lab installed on a PC with 3 GHz Intel 8-Core processor and 16GB RAM.

### 5.5. Results analysis

As expected, the obtained results confirm the unequal contribution of both A and L matrices and their overlapping in terms of graphs representation. The reported results in Tables 1 and 3, and in figure 6 show that the achieved classification accuracies of JSS measure are promising with respect to some state-of-the-art methods. The JSS based kernel performs well in the majority of datasets. In molecular data, it reaches second best accuracy in all datasets (MUTAG, NCI1, NCI109, ENZYMES and D&D) compared to the benchmark kernels. But in terms of CPU time, the JSS based kernel is faster than other kernels in many cases, such as in MUTAG, ENZYMES and mostly in D&D, where JSS kernel performs almost as well as Weisfeiler-Lehman edge kernel (75.75 Vs 77.95), while execution time is 600 times faster. In Table 3, the JSS based kernel provides good results applied to time series classification problems. Especially in the Lightning2, ToeSegmentation1 and Computers databases, it performs better than other graph kernels, and even better than kernels applied directly to time series. The representation of time series as VG allowed a clear improvement of the classification accuracy, approaching 12% for ToeSegmentation1, 8% for Lightning2 and 6% for Computers. The reported results, on different data sets with varying complexity and heterogeneity, in terms of classification accuracy and computational cost, demonstrate the effectiveness and the interest of the proposed JSS measure. This spectral similarity generalizes the spectral distance between graphs based on purely A ( $JSS_A$ ) or L ( $JSS_L$ ) matrix. As reported in Tables 1 and 3, and figure 6, JSS is well sensitive to these graph properties and allows to effectively handle them. Figure 6 shows that the JSS-based method highly outperforms the method purely based on  $JSS_A$  or  $JSS_L$ . A careful examination of figure 6 shows that for both  $\alpha = 0$  and  $\alpha = 1$ , corresponding respectively to  $JSS_L$  and  $JSS_A$ , neither of  $JSS_A$  nor  $JSS_L$  is able to perform better than JSS. However, overall,  $JSS_A$  achieves better results than  $JSS_L$  and this can be attributed, as expected, to the less sensitivity of A to structural changes, compared to L matrix, but more efficient for graph discrimination. According to these results, the weighting parameter lies in  $]0; 1[$ , showing that L and A are complementary and thus carry different information about the underlying graph. For all data sets, we have the parameter  $\alpha \neq 0.5$ , which indicates that  $JSS_L$  and  $JSS_A$  are unequally contributed and highlights that the A and L representation matrices recover different structures of the graph. Also, these results show that JSS effectively captures information conveyed by A and L matrices. The fact that the assigned weighting parameter  $\alpha$  varies from one data set to another emphasizes that each graph has its own structure and also that both A and L matrices convey different information.

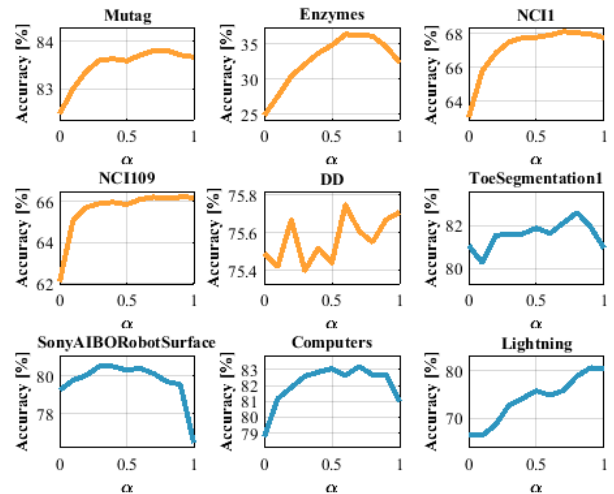


Fig. 6: Accuracy variation of JSS based kernel as function of  $\alpha$ .

## 6. Conclusion

The representation of graphs using matrices plays an important role in graph spectral theory and in many other applications dealing with graphs. In this paper a new joint spectral similarity (JSS) measure for graphs classification is introduced. We have shown that both adjacency and Laplacian matrices carry different structures information of the underlying graph. The adjacency matrix characterizes the topological graph complexity in terms of connections between nodes and their intensities, and also underscores the local cohesiveness of nodes. These properties explain why the good classification accuracies achieved by JSS measure are more attributed to adjacency matrix ( $\alpha > 0.5$ ). Through VN entropy, it is easy to see that Laplacian matrix brings out changes in node degrees information. Furthermore, this matrix is well suited to recover information about clusters of the graph and thus capture its inherent structure. The obtained results highlight the fact that JSS combines both advantages of Laplacian and adjacency matrices. Also, these findings confirm that these matrices contribute unequally and emphasize the fact that they represent differently information about structures of the underlying graph. Additionally, these results show the interest of the VG approach for classification of time series. As a result of this work, we hope to have increased the awareness about the importance of the properly choice of the representation matrix for graph spectral analysis purposes. Even the JSS measure handles cospectral graphs with respect to both A and L, it can be extended to the case of graphs that are cospectral in regard to a large class of graph representation matrices. At last, the optimal value of  $\alpha$  is in general not known and is determined only through experimentation. As future work we plan to develop a strategy to find automatically the optimal value of  $\alpha$ .

## References

- [1] E. R. Van Dam, W. H Haemers, Which graphs are determined by their spectrum?, *Linear Algebra and its applications* 373 (2003) 241–272.
- [2] I. Jovanović, Z. Stanić, Spectral distances of graphs based on their different matrix representations, *FILOMAT* 28 (4) (2014) 723–734.

Table 1: Classification accuracy on molecular data. Ramon and Gärtner [38],  $p$ -random walk [37], Random walk [36], Weisfeiler-Lehman edge [39] are some state of the art kernels.

Method/Data	MUTAG	NCI1	NCI109	ENZYMES	D&D
JSS <sub>A</sub>	83.67 ±0.22	67.74 ±0.14	66.21 ±0.28	32.35 ±1.15	75.71 ±0.14
JSS <sub>L</sub>	82.48 ±0.36	63.07 ±0.25	62.07 ±0.26	24.86 ±0.93	75.49 ±0.24
JSS	83.81 ±0.48	68.10 ±0.16	66.22 ±0.13	36.35 ±0.90	75.75 ±0.19
[38]	85.72 ±0.49	61.86 ±0.27	61.67 ±0.21	13.35 ±0.87	57.27 ±0.07
[37]	79.19 ±1.09	58.66 ±0.28	58.36 ±0.94	27.67 ±0.95	66.64 ±0.83
[36]	80.72 ±0.38	64.34 ±0.27	63.51 ±0.18	21.68 ±0.94	71.70 ±0.47
[39]	81.06 ±1.95	84.37 ±0.30	84.49 ±0.20	53.17 ±2.04	77.95 ±0.70

Table 2: CPU runtime for kernel computation on molecular data. Ramon and Gärtner [38],  $p$ -random walk [37], Random walk [36], Weisfeiler-Lehman edge [39].

Method/Data	MUTAG	NCI1	NCI109	ENZYMES	D&D
JSS <sub>A</sub>	0.32''	2'29''	2'27''	3''	3'24''
JSS <sub>L</sub>	0.88''	2'28''	2'30''	3.50''	3'34''
JSS	1.30''	4'14''	4'19''	5.88''	6'53''
[38]	40'6''	81 days	81 days	38 days	103 days
[37]	4'42''	5 days	5 days	10'	4 days
[36]	12''	9 days	9 days	12'19''	48 days
[39]	3''	1'5''	58''	11''	3 days

Table 3: Classification accuracy on time series data sets. Random walk [36], Weisfeiler-Lehman edge [39] are some state of the art kernels.

Method/Data	ToeSegmentation1	SonyAIBORobotSurface1	Computers	Lightning2
JSS <sub>A</sub>	80.94 ±0.82	76.50 ±0.41	80.97 ±0.67	80.52 ±1.76
JSS <sub>L</sub>	81.08 ±0.52	79.01 ±0.46	78.76 ±0.42	66.63 ±1.07
JSS	82.61 ±0.79	80.52 ±0.34	83.21 ±0.60	80.68 ±1.71
[36]	59.03 ±0.64	55.70 ±0.48	60.50 ±0.01	60.30 ±2.30
[39]	61.08 ±1.68	56.46 ±1.01	76.22 ±1.02	49.15 ±2.65
Linear-SVM	54.10 ±2.05	97.43 ±0.23	53.42 ±1.67	59.54 ±3.21
RBF-SVM	71.11 ±1.64	99.08 ±0.07	56.28 ±0.65	72.99 ±1.06

- [3] R.K. Fan Chung, Spectral graph theory, AMS, 1996.
- [4] J. H. Koolen, H. Yu, The distance-regular graphs such that all of its second largest local eigenvalues are at most one, arXiv preprint arXiv:1102.4292.
- [5] D. Conte, P. Foggia, C. Sansone, M. Vento, Thirty years of graph matching in pattern recognition, *Int. J. Patt. Recogn. Art. Intell.* 18 (3) (2004) 265–298.
- [6] J. Kunegis, Exploiting the structure of bipartite graphs for algebraic and spectral graph theory applications, *Internet Math.* 11 (3) (2015) 201–321.
- [7] D. Shuman, S. Narang, K. Sunil, P. Frossard, A. Ortega, P. Vandergheynst, The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains, *IEEE Sig. Proc. Mag.* 30 (3) (2013) 83–98.
- [8] A. Sandryhaila, J. Moura, Discrete signal processing on graphs: Frequency analysis, *IEEE Trans. Sig. Proc.* 62 (12) (2014) 3042–3054.
- [9] C.D. Godsil, B.D. McKay, Constructing cospectral graphs, *Aequationes Mathematicae* 25 (1) (1982) 257–268.
- [10] L. Bai, L. Rossi, Z. Zhang, E. R. Hancock, An aligned subtree kernel for weighted graphs, in: *International Conference on Machine Learning*, 2015, pp. 30–39.
- [11] N. M. Kriege, P. L. Giscard, R. C. Wilson, On valid optimal assignment kernels and applications to graph classification, in: *Advances in Neural Information Processing Systems*, 2016, pp. 1623–1631.
- [12] L. Bai, E. R. Hancock, P. Ren, Jensen-shannon graph kernel using information functionals, in: *Pattern Recognition (ICPR)*, 2012 21st International Conference on, IEEE, 2012, pp. 2877–2880.
- [13] C. Morris, K. Kersting, P. Mutzel, Glocalized weisfeiler-lehman graph kernels: Global-local feature maps of graphs, in: *Data Mining (ICDM)*, 2017 IEEE International Conference on, IEEE, 2017, pp. 327–336.
- [14] L. Bai, E. R. Hancock, Fast depth-based subgraph kernels for unattributed graphs, *Pattern Recognition* 50 (2016) 233–245.
- [15] K. Anand, G. Bianconi, S. Severini, Shannon and Von Neumann entropy of random networks with heterogeneous expected degree, *Phys. Rev. E* 83 (3) (2011) 036109.
- [16] J. Von Neumann, *Mathematical foundations of quantum mechanics*, no. 2, Princeton university press, 1955.
- [17] L. Han, F. Escolano, E.R. Hancock, R.C. Wilson, Graph characterizations from von neumann entropy, *Patt. Recogn. Lett.* 33 (15) (2012) 1958–1967.
- [18] S. Braunstein, S. Ghosh, S. Severini, The Laplacian of a graph as a density matrix: a basic combinatorial approach to separability of mixed states, *Ann. Combinatorics* 10 (3) (2006) 291–317.
- [19] F. Passerini, S. Severini, Quantifying complexity in networks: the Von Neumann entropy, *Int. J. Agent Technol. Syst.* 1 (4) (2009) 58–67.
- [20] L. Bai, L. Rossi, L. Cui, Z. Zhang, P. Ren, X. Bai, E. R. Hancock, Quantum kernels for unattributed graphs using discrete-time quantum walks, *Pattern Recognition Letters* 87 (2017) 96–103.
- [21] L. Bai, L. Rossi, A. Torsello, E. R. Hancock, A quantum jensen-shannon graph kernel for unattributed graphs, *Pattern Recognition* 48 (2) (2015) 344–355.
- [22] X. Chen, Perturbation theory of von neumann entropy, *Chinese Physics B* 19 (4) (2010) 1–7.
- [23] M. Fiedler, Algebraic connectivity of graphs, *Czechoslovak Mathematical J.* 23 (2) (1973) 298–305.
- [24] I. Gutman, S. Firoozabadi, J.A. de la Pena, J. Rada, On the energy of regular graphs, *Comm. Math. Comput. Chem.* 57 (2007) 435–442.
- [25] I. Gutman, B. Zhou, Laplacian energy of a graph, *Lin. Algebra Appl.* 414 (1) (2006) 29–37.
- [26] A.E. Brouwer, W.H. Haemers, *Spectra of graphs*, Springer Science &



- Business Media, 2011.
- [27] A. Schwenk, Almost all trees are cospectral, *New directions in the theory of graphs* (1973) 275–307.
  - [28] R. Merris, Large families of laplacian isospectral graphs, *Linear and Multilinear Algebra* 43 (1-3) (1997) 201–205.
  - [29] L. Lacasa, B. Luque, F. Ballesteros, J. Luque, J. Nuno, From time series to complex networks: The visibility graph, *PNAS* 105 (13) (2008) 4972–4975.
  - [30] A. Debnath, R.L. Lopez de Compadre, G. Debnath, A. Shusterman, C. Hansch, Structure-activity relationship of mutagenic aromatic and heteroaromatic nitro compounds. correlation with molecular orbital energies and hydrophobicity, *J. Med. Chem.* 34 (2) (1991) 786–797.
  - [31] N. Wale, G. Karypis, Acyclic subgraph based descriptor spaces for chemical compound retrieval and classification, *Tech. rep., DTIC Document* (2006).
  - [32] K. Borgwardt, C. Ong, S. Schönauer, S.V.N. Vishwanathan, A. Smola, H-P. Kriegel, Protein function prediction via graph kernels, *Bioinformatics* 21 (1) (2005) 47–56.
  - [33] I. Schomburg, A. Chang, C. Ebeling, M. Gremse, C. Heldt, G. Huhn, D. Schomburg, BRENDA, the enzyme database: updates and major new developments, *Nucleic Acids Research* 32 (1) (2004) 431–433.
  - [34] P. Dobson, A. Doig, Distinguishing enzyme structures from non-enzymes without alignments, *J. Mol. Biol.* 330 (4) 771–783.
  - [35] A. Bagnall, E. Keogh, J. Lines, A. Bostrom, J. Large, <http://timeseriesclassification.com/dataset.php>, accessed: 2018-07-17.
  - [36] S.V.N. Vishwanathan, N.N. Schraudolph, R. Kondor, K.M. Borgwardt, Graph kernels, *J. Mach. Learn. Res.* 11 (2010) 1201–1242.
  - [37] H. Kashima, K. Tsuda, A. Inokuchi, Marginalized kernels between labeled graphs, in: *ICML*, Vol. 3, 2003, pp. 321–328.
  - [38] T. Gärtner, P. Flach, S. Wrobel, On graph kernels: Hardness results and efficient alternatives, in: *Learning Theory and Kernel Machines*, 2003, pp. 129–143.
  - [39] N. Shervashidze, P. Schweitzer, E.J. Van Mehlhorn, K. Mehlhorn, M. Borgwardt, Weisfeiler-lehman graph kernels, *J. Mach. Learn. Res.* 12 (2011) 2539–2561.