



Science Arts & Métiers (SAM)

is an open access repository that collects the work of Arts et Métiers Institute of Technology researchers and makes it freely available over the web where possible.

This is an author-deposited version published in: <https://sam.ensam.eu>
Handle ID: <http://hdl.handle.net/10985/15078>

To cite this version :

Hadj-Ahmed BAY-AHMED, Delphine DARE-EMZIVAT, Abdel-Ouahab BOUDRAA - Graph Signals Classification Using Total Variation and Graph Energy Informations - In: 5th IEEE Global Conference on Signal and Information Processing, Canada, 2017-11 - 5th IEEE Global Conference on Signal and Information Processing - 2017

Any correspondence concerning this service should be sent to the repository

Administrator : scienceouverte@ensam.eu



Graph Signals Classification Using Total Variation and Graph Energy Informations

H. Bay Ahmed, D. Dare and A.O. Boudraa

IRENav, Ecole Navale/Arts-Metiers ParisTech, CC 600, 29240 Brest Cedex 9, France.

E-mail: (hadj_ahmed.bay_ahmed,dare,boudraa)@ecole-navale.fr

Abstract—In this work, we consider the problem of graph signals classification. We investigate the relevance of two attributes, namely the total variation (TV) and the graph energy (GE) for graph signals classification. The TV is a compact and informative attribute for efficient graph discrimination. The GE information is used to quantify the complexity of the graph structure which is a pertinent information. Based on these two attributes, three similarity measures are introduced. Key of these measures is their low complexity. The effectiveness of these similarity measures are illustrated on five data sets and the results compared to those of five kernel-based methods of the literature. We report results on computation runtime and classification accuracy on graph benchmark data sets. The obtained results confirm the effectiveness of the proposed methods in terms of CPU runtime and of classification accuracy. These findings also show the potential of TV and GE informations for graph signals classification.

I. INTRODUCTION

Discrete Signal Processing on Graphs (DSP_G) is an emerging topic that has drawn much interest recently in multiple disciplines [1]-[2]. Graphs are data structures well suited to represent complex relationships among high-dimensional data and can be used in many fields of engineering and science. The graph can be constructed using the data to capture the underlying geometry. With the growth of information and communication, data signals are generated at different rate from various sources, including social networks, biological, citation and physical networks. Nowadays, many data analysis and processing tasks have to handle large structured datasets, where the structure information carries important information about the nature of the observations. Unlike time series or images, these data have complex and irregular structures, which require new processing methods leading to the emerging field of DSP_G. Classical DSP can be viewed as a special case of DSP_G. More precisely, signal processing on graphs aims to extend the tools used for analysis, denoising, classification, machine learning and interpolation of traditional signals to signals defined on graphs. The aim of graph signals processing framework is to allow us to process given data while taking into consideration the underlying connectivity between the data points. Due to their ability to capture the complex relationships in high-dimensional datasets, graphs constitute a potential tool for data analysis such as classification, which is an important problem in machine learning and data mining [3]. The present paper deals with the classification of graph signals using similarity measures exploiting the total variation (TV) [4],[5], [6], and the Laplacian graph energy [7] informations.

The graph energy is a measure of the graph complexity and TV operator takes into account both data and graph structure. More precisely, TV of a graph signal measures how much the signal changes between node values and thus measures the change upon the graph structure. Based on these two attributes three similarity measures are introduced. These measures are of low complexity and easy to implement. The effectiveness of these similarity methods are illustrated on five data sets and the results compared to those of five kernel-based methods of the literature. Before graph signals classification, we need to ensure that the data are well mapped to the graph space and that the interpretation will be more realistic in the new space [8].

II. BASICS OF GRAPH SIGNALS

A graph is defined with a finite non-empty set V of N points called vertices and with another set E of M unordered pairs of distinct points of V . A graph with N vertices and M edges is called a (N, M) graph. In other words, a graph is a set of vertices with a connectivity relation between its elements. A labelled graph with N vertices and M edges can be represented with an $(N \times N)$ adjacency matrix $(\mathbf{A} = [a_{ij}])$ where the element a_{ij} is equal to one if the vertex v_i is connected to the vertex v_j . If not, a_{ij} is zero. A graph signal is given by a mapping of a real (or complex) dataset to a graph structure and is noted as :

$$\begin{aligned} \mathbf{X} : V &\rightarrow \mathbb{R}, \\ v_n &\rightarrow X_n \end{aligned} \quad (1)$$

The dataset of the graph signal can be written as a vector:

$$\mathbf{X} = [X_0, X_1, \dots, X_{N-1}]^T \in \mathbb{R}^N \quad (2)$$

Each element X_n is mapped to a vertex v_n of a graph $G(V, \mathbf{A})$ completing the graph signal construction. In the following, a graph signal $G(X, \mathbf{A})$ is defined as the association of data supported by nodes and a graph structure defined by its adjacency matrix.

III. GRAPH TOTAL VARIATION

Classical TV of a discrete signal is defined as the accumulation of the differences between consecutive signal samples [2]:

$$\text{TV}(\mathbf{X}) = \sum_{k \in \mathbb{Z}} |X_k - X_{k-1}| \quad (3)$$

For a finite time series where $X_n = X_{n \bmod N}$, the TV is defined as:

$$\text{TV}(\mathbf{X}) = \sum_{n=0}^{N-1} |X_n - X_{n-1 \bmod N}| \quad (4)$$

TV, given by Eqs. (3)-(4), measures the variation of the signal in the time. Equation (3) calculates a cumulative difference of the signal over the time. TV compares the original signal \mathbf{X} to its shifted version. When signal variation is low, then the accumulation difference is also low. TV (3) can be written using a cyclic permutation matrix \mathbf{C} :

$$\text{TV}(\mathbf{X}) = \|\mathbf{X} - \mathbf{C}\mathbf{X}\|_1 \quad (5)$$

where $\|\cdot\|_1$ is the l_1 -norm. TV, given by (5), measures the cumulative difference between the signal data at each node and its neighbours. In DSP_G framework, relation (5) is generalized for any graph signal $G(\mathbf{X}, \mathbf{A})$ by considering the TV quantity as a measure of the difference between a graph signal and its shifted version:

Definition: the TV_G [9] of a graph signal $G(\mathbf{X}, \mathbf{A})$ is defined as

$$\text{TV}_G(\mathbf{X}) = \|\mathbf{X} - \mathbf{A}\mathbf{X}\|_q \quad (6)$$

where \mathbf{A} is the normalized adjacency matrix of the graph and $\mathbf{A}\mathbf{X}$ is the shifted version of the signal. The graph TV can be based on the l_q -norm.

IV. GRAPH SIGNALS SIMILARITY MEASURE BASED ON TV

TV characterizes the oscillations of the signal values upon the structure of the graph, where high variations give indications about high frequency components [2]. These frequencies depend directly on both signal values on each node and their respective degrees. A high local oscillation corresponds to abrupt variation of the signal value with respect to its neighbours. The more the neighbourhood is large the more important the oscillation is. Given the structure of the graph signal and its node values, the TV can be considered as a good and informative descriptor for graph discrimination. More precisely, TV_G is well suited for graph signals comparison. Let us consider two graph-signals $\{G_i(\mathbf{X}_i, \mathbf{A}_i), G_j(\mathbf{X}_j, \mathbf{A}_j)\} \subseteq \mathcal{G}$ where \mathbf{A}_i and \mathbf{X}_i are respectively the adjacency matrix and the vector of the scalar data supported on the graph nodes, with \mathcal{G} being the set of graph signals. Using relation (6), the TV of the k^{th} graph signal is given by :

$$\widetilde{\text{TV}}_{G_k} = \frac{\text{TV}_{G_k}}{N_k} = \frac{\|\mathbf{X}_k - \mathbf{A}_k \mathbf{X}_k\|_q}{N_k} \quad (7)$$

where N_k is the number of the nodes of G_k . We define the difference between two graph signals, $G_i(\mathbf{X}_i, \mathbf{A}_i)$ and $G_j(\mathbf{X}_j, \mathbf{A}_j)$, as the difference of their respective mean TVG as :

$$\text{TVG}(G_i, G_j) = |\widetilde{\text{TV}}_{G_i} - \widetilde{\text{TV}}_{G_j}| \quad (8)$$

The smallest the difference is, the highest the similarity is. It is important to point out that the similarity between graph signals involves both the node values and the graph structure

(Eq. 7). Besides the TV measure, the complexity information of the graph structure is pertinent attribute for graph signals classification.

V. GRAPH COMPLEXITY

Due to its practical importance, graph complexity quantification has attracted significant attention in various domains such as pattern recognition, control theory or network analysis [10]. The measure of this complexity is important for different applications including embedding [11], classification [12] and filtering of image description hierarchies [13]. The graph complexity can be measured in different ways [10]. An attractive measure is the Laplacian graph energy (GE), introduced by Gutman and Zhou [7], which has shown promising results, for example, in image processing [13]. This information is exploited for similarity measure of graph signals.

A. Laplacian graph energy

The concept of GE has a chemical motivation and goes back to the work of Hückel on the approximate solution of the equation of Schrödinger of certain organic molecules [14]. Based on facts observed in molecular theory, Gutman [15] introduced the definition of the GE as follows:

$$E(G) = \sum_{k=1}^N |\mu_k| \quad (9)$$

where μ_k are the eigenvalues of the adjacency matrix of graph. This spectrum-based graph invariant has been largely studied in both chemical and mathematical literature. Instead of Eq. (9), a GE-like quantity defined in terms of Laplacian and preserving the main features of the original GE has been proposed by Gutman and Zhou [7]:

$$\text{LE}(G) = \sum_{k=1}^N \left| \omega_k - \frac{2M}{N} \right| \quad (10)$$

where ω_k are the eigenvalues of the Laplacian matrix of graph and $\frac{2M}{N}$ is the average vertex degree. Using this energy index two measures of similarity of graph signals are introduced.

B. Laplacian graph energy similarity measure

The first measure, GE, is based on the difference in energy between two graph signals G_i and G_j as follows

$$\text{GE}(G_i, G_j) = |\text{LE}(G_i) - \text{LE}(G_j)| \quad (11)$$

C. Joint Energy Total variation similarity measure

The second measure is a joint difference of GE and the TVG called JET

$$\text{JET}(G_i, G_j) = \lambda \times \text{GE}(G_i, G_j) + (1 - \lambda) \times \text{TVG}(G_i, G_j) \quad (12)$$

where $\lambda \in [0, 1]$ is a weighted parameter controlling the contributions of GE and TVG. The JET measure takes account both the complexity of the graph and the interaction between graph structure and data supported by nodes.

VI. GRAPH KERNEL BASED ON TVG, GE AND JET SIMILARITY MEASURES

In kernel strategies, patterns are represented by pairwise similarity function instead of individual feature vectors. Let us consider a pattern data set \mathcal{D} with M objects $\{d_1, d_2, \dots, d_M\} \subset \mathcal{P}$, where \mathcal{P} is the space of all possible patterns, and a similarity function $\mathcal{K} : \mathcal{D} \times \mathcal{D} \rightarrow \mathbb{R}$, known as *kernel function*. It associates to each pattern pair a kernel value $\mathcal{K}_{ij}(d_i, d_j)$. These values can replace inner products in learning algorithms and permit the accomplishment of the learning task in a higher implicit vector space. Kernels extract relevant information for classification insofar as some criteria are respected. That is, kernel values must be obtained from a symmetric and *positive definite* function [16] so that they are called *valid kernels or Mercer Kernels* [17]. Several graph kernels have been developed and tested in literature [18],[19] such as diffusion kernels, walk kernels, convolution kernels [20], and Weisfeiler-Lehman kernels [21].Inspiring from Radial Basis Function kernel [16], we built a graph signals kernel by the integration of TVG, GE and JET on an exponential function. We associate to each pair of graph signals $\mathbf{G}_i, \mathbf{G}_j$ the quantity:

$$\mathbf{K}_{i,j} = \sum_{n=0}^{\infty} \frac{1}{n!} (-\gamma s(\mathbf{G}_i, \mathbf{G}_j))^n = \exp(-\gamma s(\mathbf{G}_i, \mathbf{G}_j)) \quad (13)$$

where \mathbf{K} is the kernel matrix and $s(\mathbf{G}_i, \mathbf{G}_j)$ is one of similarity measures proposed above (TVG, GE and JET), γ is a smoothing factor. Using the kernel (13), graph signals are classified in a hidden infinite dimensional Euclidean space instead of one dimensional raw space.

VII. EXPERIMENTAL RESULTS

A. Graph Signals Data sets

To show the effectiveness of the proposed similarity measures, five data sets are used, all concerning chemical/biological compounds. MUTAG [22] 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 balanced subsets of data sets of chemical compounds screened for activity against non-small cell lung cancer and ovarian cancer cell lines [23]. ENZYMES is a data set of protein tertiary structures obtained from [24] consisting of 600 enzymes from BRENDA enzyme database [25]. In this data set, 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 [26]. Each protein is represented by a graph, in which the nodes are amino acids and two nodes are connected by an edge if they are separated by less than 6 Angström. The classification task is to distinguish protein structures between enzymes and non-enzymes. Note that all these graphs are graph signals with labelled nodes.

B. Experimental configuration

The kernel used (Eq. 13) is integrated on Support Vector Machine algorithm and 10-fold cross-validation strategy 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 γ is set to 1. The best performance of the JET measure, in term of classification accuracy, is obtained by a suitable selection of λ . We find the best value by varying this parameter from 0 to 1 and picking out the value where the classification accuracy is maximal. This parameter is not known a-priori and has to be determined based on the data set used. For TVG and JET measures, the *norm 2* ($q = 2$) is used to calculate TV operator. Average classification accuracies and the associated standard deviations are summarized in Table I. Performances of kernel based on TVG, GE and JET measures are compared to different kernels of the literature in terms of prediction accuracy and computation runtime on graph signals benchmark data sets. The most known graph kernels are tested: those based on walks, Weisfeiler-Lehman isomorphism and limited-size subgraphs. Our similarity measures are compared to the fast geometric random walk kernel proposed by Vishwanathan et al. [27], counts common labelled walks and to p -random walk kernel that compares random walks up to length p in two graphs (a special case of random walk kernels: Kashima et al. [28], Gärtner et al. [18]). In the case of limited-size subgraphs family, we compare with an extension of the graphlet kernel proposed by Shervashidze et al. [29] that counts common induced labelled connected subgraphs of size 3. From Weisfeiler-Lehman kernels, we picked up the Weisfeiler-Lehman edge kernel [21],[29]: it counts matching pairs of edges with identically labeled endpoints (incident nodes) in two graphs. Computing set-up, accuracy and runtime values of the benchmark kernels are performed and reported by Shervashidze et al. in [21]. As reported in Table II, runtime in minutes and seconds of our methods are measured in Anaconda2 4.1.1 Python 2.7.12 Lab installed on a PC with 3 GHz Intel 8-Core processor and 16GB RAM.

C. Evaluation

Overall, as shown in Table I, On NCI1, NCI109, ENZYMES and D&D Weisfeiler-Lehman edge kernels reach the highest accuracy but perform less than JET kernel on MUTAG. As shown in Table I, on MUTAG, NCI1, NCI109, and D&D, the TVG, GE, JET based kernels reach good accuracy and are competitive with other kernels. On MUTAG, the GE and JET based kernels give the second best accuracy and perform better than random walk and Weisfeiler-Lehman edge kernels. On NCI1, NCI109, ENZYMES, and D&D, the JET kernel reaches the third best accuracy and performs better than random walk kernels. In term of computation runtime, on all data sets the JET, TVG and GE kernels are more faster than all other kernels and particularly compared to Ramon and Gärtner kernel. As shown in Table II, our similarity measures outperform state-of-the-art graph kernels in term of computation runtime. The best computational time over all the data sets and all the considered methods is provided by the TVG kernel. This result highlights the low complexity of

TABLE I
CLASSIFICATION ACCURACY (\pm STANDARD DEVIATION) ON GRAPH SIGNALS BENCHMARK DATA SETS.

Methode/Dataset	MUTAG	NCI1	NCI109	ENZYMES	D&D
Graph Signal Total Variation TVG	77.88 (± 1.43)	55.91 (± 0.38)	56.62 (± 0.10)	20.36 (± 0.91)	75 (± 0.10)
Graph Signal Energy GE	82.34 (± 1.25)	61.81 (± 0.40)	62.22 (± 0.35)	23.33 (± 1.44)	64.45 (± 0.61)
Energy and Total Variation JET	83.51 (± 1.14)	64.43 (± 0.21)	64.88 (± 0.16)	31 (± 0.53)	75 (± 0.04)
Ramon and Gärtner	85.72 (± 0.49)	61.86 (± 0.27)	61.67 (± 0.21)	13.35 (± 0.87)	57.27 (± 0.07)
p -random walk	79.19 (± 1.09)	58.66 (± 0.28)	58.36 (± 0.94)	27.67 (± 0.95)	66.64 (± 0.83)
Random walk	80.72 (± 0.38)	64.34 (± 0.27)	63.51 (± 0.18)	21.68 (± 0.94)	71.70 (± 0.47)
Graphlet count	75.61 (± 0.49)	66.00 (± 0.07)	66.59 (± 0.08)	32.70 (± 1.20)	78.59 (± 0.12)
Weisfeiler-Lehman edge	81.06 (± 1.95)	84.37 (± 0.30)	84.49 (± 0.20)	53.17 (± 2.04)	77.95 (± 0.70)

TABLE II
CPU RUNTIME FOR KERNEL COMPUTATION ON GRAPH SIGNALS CLASSIFICATION BENCHMARK DATA SETS.

Methode/Dataset	MUTAG	NCI1	NCI109	ENZYMES	D&D
Graph Signal Total Variation TVG	0.036''	17''	17.4''	0.4''	1.4''
Graph Signal Energy GE	0.12''	20.6''	21.8''	1''	2'19''
Energy and Total Variation JET	0.18''	51''	51.2''	1.6''	2'21''
Ramon and Gärtner	40'6''	81 days	81 days	38 days	103 days
p -random walk	4'42''	5 days	5 days	10'	4 days
Random walk	12''	9 days	9 days	12'19''	48 days
Graphlet count	3''	1'27''	1'27''	25''	30'21''
Weisfeiler-Lehman edge	3''	1'5''	58''	11''	3 days

this kernel. Note that the JET kernel is faster ($\times 2000$) than Weisfeiler-Lehman edge kernel in D&D data set with almost the same accuracy (75% vs 78%). By combining TVG and GE similarity measures, we improve their individual performance about 7% for ENZYMES, 3% for NCI1/NCI109 and 1% for MUTAG. While for D&D, no improvement is obtained, the best accuracy is reached by TVG alone. Figure 1 shows the behaviour of prediction performance of JET kernel according to the weighting factor λ . We note that the maximum accuracy is obtained between the limit values of λ ($\lambda \in [0, 1]$), which confirms that with the contribution of both TVG and GE measures, higher accuracy rates can be reached.

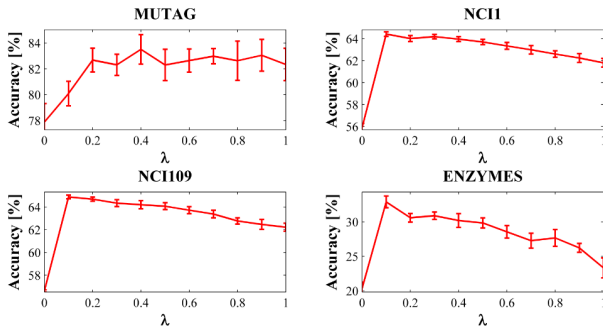


Fig. 1. Accuracy variation of JET based kernel in function of λ .

VIII. CONCLUSION

In this work, new graph-signals similarity measures based on graph TV and Laplacian GE are proposed. These measures

are not limited to unweighed graphs. TVG measure is an informative and efficient descriptor for graph signals discrimination. It quantifies the interaction between the node values and the supporting structure of the graph. GE information is a pertinent characteristic that measures the complexity degree of the graph signal. It takes into account both connections distribution of the network and its density. These two measures integrated in an exponential kernel showed competitive performance on binary and multiclass graph signals classification. TVG and GE are combined into a new and relevant joint measure, JET. TVG, GE and JET measures are illustrated on five data sets and the results compared to those of five kernel-based methods of the literature. Our measures yield competitive accuracy levels on all considered data sets and outperform state-of-the-art graph kernels in term of computation runtime. Results of JET measure show the benefits of hybrid approaches on discriminating graph signals without significant increase of complexity. This work opens up new perspectives and encourages us to carry out more studies about joint similarity measures and graph signals energy distribution. We aim to optimize JET measure and study further its behavior in other special graph signals datasets. We plan to develop a strategy to select automatically the best λ value and to test TVG, GE, JET measures according to different values of kernel parameter γ . Finally, it would be interesting to investigate the effect of different l_q -norms ($q \neq 2$) on the obtained accuracy levels.

REFERENCES

- [1] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 83–98, 2013.
- [2] A. Sandryhaila and J. M. Moura, "Discrete signal processing on graphs: Frequency analysis," *IEEE Trans. Signal Processing*, vol. 62, no. 12, pp. 3042–3054, 2014.
- [3] A. Sandryhaila and J. M. Moura, "Classification via regularization on graphs," in *GlobalSIP*, pp. 495–498, 2013.
- [4] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1–4, pp. 259–268, 1992.
- [5] T. F. Chan, S. Osher, and J. Shen, "The digital tv filter and nonlinear denoising," *IEEE Transactions on Image processing*, vol. 10, no. 2, pp. 231–241, 2001.
- [6] F. Mahmood, N. Shahid, U. Skoglund, and P. Vandergheynst, "Compressed sensing and adaptive graph total variation for tomographic reconstructions," *arXiv preprint arXiv:1610.00893*, 2016.
- [7] I. Gutman and B. Zhou, "Laplacian energy of a graph," *Linear Algebra and its applications*, vol. 414, no. 1, pp. 29–37, 2006.
- [8] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: A general framework for dimensionality reduction," *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, no. 1, pp. 40–51, 2007.
- [9] S. Chen, A. Sandryhaila, J. M. Moura, and J. Kovačević, "Signal recovery on graphs: Variation minimization," *IEEE Transactions on Signal Processing*, vol. 63, no. 17, pp. 4609–4624, 2015.
- [10] F. Escolano, E. R. Hancock, and M. A. Lozano, "Polytopal graph complexity, matrix permanents, and embedding," in *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*, pp. 237–246, Springer, 2008.
- [11] A. Robles-Kelly and E. R. Hancock, "A riemannian approach to graph embedding," *Pattern Recognition*, vol. 40, no. 3, pp. 1042–1056, 2007.
- [12] A. Shokoufandeh, S. J. Dickinson, K. Siddiqi, and S. W. Zucker, "Indexing using a spectral encoding of topological structure," in *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, vol. 2, pp. 491–497, IEEE, 1999.
- [13] Y.-Z. Song, P. Arbelaez, P. Hall, C. Li, and A. Balikai, "Finding semantic structures in image hierarchies using laplacian graph energy," *Computer vision—ECCV 2010*, pp. 694–707, 2010.
- [14] I. Gutman and O. E. Polansky, *Mathematical concepts in organic chemistry*. Springer Science & Business Media, 2012.
- [15] I. Gutman, S. Z. Firoozabadi, J. A. de la Pena, and J. Rada, "On the energy of regular graphs," *MATCH Commun. Math. Comput. Chem*, vol. 57, pp. 435–442, 2007.
- [16] B. Schölkopf and A. J. Smola, *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2002.
- [17] K. Riesen and H. Bunke, *Graph classification and clustering based on vector space embedding*, vol. 77. World Scientific, 2010.
- [18] T. Gärtner, "A survey of kernels for structured data," *ACM SIGKDD Explorations Newsletter*, vol. 5, no. 1, pp. 49–58, 2003.
- [19] T. Gärtner, P. Flach, and S. Wrobel, "On graph kernels: Hardness results and efficient alternatives," in *Learning Theory and Kernel Machines*, pp. 129–143, Springer, 2003.
- [20] D. Haussler, "Convolution kernels on discrete structures," tech. rep., Citeseer, 1999.
- [21] N. Shervashidze, P. Schweitzer, E. J. v. Leeuwen, K. Mehlhorn, and K. M. Borgwardt, "Weisfeiler-lehman graph kernels," *Journal of Machine Learning Research*, vol. 12, no. Sep, pp. 2539–2561, 2011.
- [22] A. K. Debnath, R. L. Lopez de Compadre, G. Debnath, A. J. Shusterman, and C. Hansch, "Structure-activity relationship of mutagenic aromatic and heteroaromatic nitro compounds. correlation with molecular orbital energies and hydrophobicity," *Journal of medicinal chemistry*, vol. 34, no. 2, pp. 786–797, 1991.
- [23] N. Wale and G. Karypis, "Acyclic subgraph based descriptor spaces for chemical compound retrieval and classification," tech. rep., DTIC Document, 2006.
- [24] K. M. Borgwardt, C. S. Ong, S. Schönauer, S. Vishwanathan, A. J. Smola, and H.-P. Kriegel, "Protein function prediction via graph kernels," *Bioinformatics*, vol. 21, no. suppl 1, pp. i47–i56, 2005.
- [25] I. Schomburg, A. Chang, C. Ebeling, M. Gremse, C. Heldt, G. Huhn, and D. Schomburg, "Brenda, the enzyme database: updates and major new developments," *Nucleic acids research*, vol. 32, no. suppl 1, pp. D431–D433, 2004.
- [26] P. D. Dobson and A. J. Doig, "Distinguishing enzyme structures from non-enzymes without alignments," *Journal of molecular biology*, vol. 330, no. 4, pp. 771–783, 2003.
- [27] S. V. N. Vishwanathan, N. N. Schraudolph, R. Kondor, and K. M. Borgwardt, "Graph kernels," *Journal of Machine Learning Research*, vol. 11, no. Apr, pp. 1201–1242, 2010.
- [28] H. Kashima, K. Tsuda, and A. Inokuchi, "Marginalized kernels between labeled graphs," in *ICML*, vol. 3, pp. 321–328, 2003.
- [29] N. Shervashidze and K. M. Borgwardt, "Fast subtree kernels on graphs," in *NIPS*, pp. 1660–1668, 2009.