2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) | 979-8-3503-7149-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/EMBC53108.2024.10782458

Optimal Transport-Based Network Alignment: Graph Classification of Small Molecule Structure-Activity Relationships in Biology

Mohammed Aburidi[1] and Roummel Marcia[1]

***Abstract*— This paper tackles the challenge of aligning graph representations of small biological molecules, crucial for understanding structure-function relationships in biological research. We focus on matching undirected and attributed graphs, addressing the limitations of current methodologies that overlook functional insights provided by node embeddings. Our approach considers graphs as probability distributions in a metric space, introducing a novel embedding scheme that accounts for immediate and secondary neighbors of nodes with continuous attributes. We formulated the graph matching problem as an optimal transport and we present an innovative graph kernel that is based on optimal transport to overcome limitations in traditional kernels, specifically addressing naive aggregation. Experimental results show that our approach outperforms state-of-the-art techniques in five of six common datasets, promising advancements in graph alignment methodologies for molecular analysis and computational biology.**

***Clinical relevance*— The porposed method advances drug discovery and development by employing a graph-based framework rooted in optimal transport theory. This approach facilitates enhanced classification of small molecules, notably proteins and enzymes, potentially revolutionizing therapeutic advancements for targeted treatments and drug design precision.**

I. INTRODUCTION

The accurate alignment and comparison of graphs representing small biological molecules, such as proteins and enzymes, hold paramount importance in understanding their structure-function relationships and facilitating robust classification [1]. Graph-based representations encapsulate intricate molecular structures, enabling nuanced analysis crucial for biological and pharmaceutical research. Aligning these graphs aids in discerning structural similarities, identifying conserved functional motifs, and elucidating evolutionary relationships among molecules. This alignment-driven classification approach not only enhances our comprehension of molecular characteristics but also empowers drug discovery, protein function prediction, and biological system modeling [2]. Understanding the significance of graph alignment underscores the foundational role of node embeddings, the initial step in graph matching methodologies.

The initial step in graph matching, pivotal to understanding graph alignment, centers on the task of learning node embeddings. This process aims to derive latent vector representations for each node within a graph, effectively encapsulating its topology and structural characteristics. Acquiring appropriate node embeddings holds immense value in graph matching as it facilitates the alignment of multiple graphs based on the metric structure associated with these embeddings. However, many existing graph matching methods predominantly rely on topological information, such as adjacency matrices, often neglecting the valuable functional insights offered by node embeddings [3]. Moreover, several proposed variations of graph matching techniques struggle to generalize effectively when confronted with graphs featuring high-dimensional continuous node attributes, necessitating intricate extensions to accommodate such complex cases [4].

Optimal transport (OT) theory has significantly influenced recent advancements in machine learning [5]–[11], primarily due to its ability to compare empirical distributions. In response to the limitations observed in existing approaches, this paper introduces a novel OT-based matching method. Our method leverages graph representations derived from an innovative graph embedding technique coupled with principles from optimal transport theory. Specifically, our approach focuses on enhancing node embeddings by considering both the first and second neighbors of the nodes, thereby improving the quality and comprehensiveness of the embeddings generated. Additionally, we introduce a novel formulation for graph kernels. Finally, we demonstrate the effectiveness of our proposed method through successful experimental results. Our method achieves higher accuracy in comparison to state-of- the-art methods. Notably, it outperforms existing methods in 5 out of 6 common datasets.

II. METHOD

A. Optimal transport background

Optimal transport [5] is a mathematical concept that defines the problem of finding the most efficient way of moving an object such as probability distribution from one configuration onto another (e.g., matching two probability distributions). Efficient here means a lower cost.

To mathematically represent the transportation problem, consider two sets of points denoted as $\mathcal{X} = \{x_i\}_{i=1}^N$ and $\mathcal{Y} = \{y_i\}_{i=1}^K$ representing the source and target samples, respectively. Let $\mathbf{p} \in \mathcal{H}_N$ and $\mathbf{q} \in \mathcal{H}_K$ represent two discretized distributions of interest, where $\mathcal{H}_N$ and $\mathcal{H}_K$ denote histograms with N and K bins respectively. The elements are constrained such that $\mathbf{p} \in \mathbb{R}_+^N$ and $\sum_i p_i = 1$, and $\mathbf{q} \in \mathbb{R}_+^K$ with $\sum_i q_i = 1$.

*This research is partially supported by NSF Grant IIS 1741490 and DMS 1840265.

[1] Mohammed Aburidi and Roummel Marcia are with the Department of Applied Mathematics, University of California Merced, Merced, California, USA.

(Corresponding author: Mohammed Aburidi, email: maburidi@ucmerced.edu)

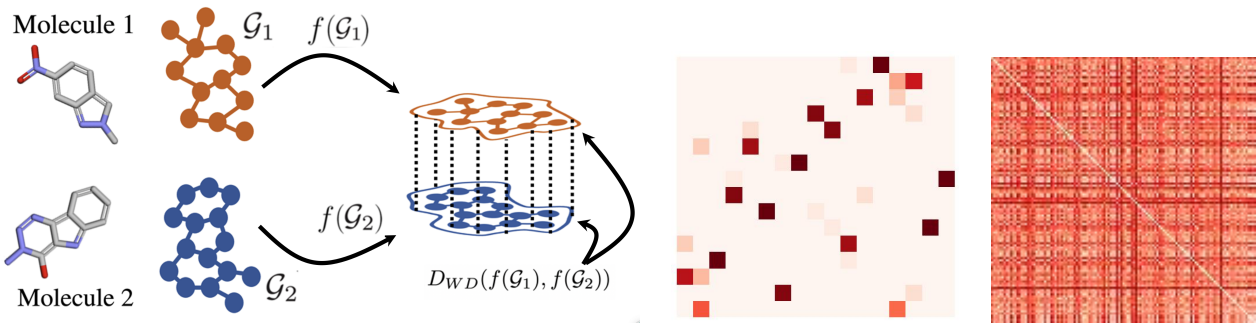

Fig. 1. (a) Graph matching scheme. First, graph embeddings are generated using the scheme $f(\cdot)$, then optimal transport is employed to align the two graphs and compute its similarity. (b) The transportation map between two graphs obtained from MUTAG dataset. (c) The similarity matrix of graphs from MUTAG dataset. Highlighted color means higher similarity. The diagonal is enforced to be zero to avoid trivial solution.

Consider the matrix $Q \in \mathbb{R}^{K\times N}$, which has non-negative entries $Q_{i,j}$ and which serves as a transportation plan or coupling matrix. This matrix describes how much mass p_j located at x_j is to be transported to match with the mass q_i situated at y_i. To ensure the validity of Q, it is necessary for the matrix to adhere to the subsequent two conditions:

$$\sum_{j=1}^{N} Q_{ij} = q_i \qquad \text{for all} \quad i \in \{1,..,K\},$$
$$\sum_{i=1}^{K} Q_{ij} = p_j \qquad \text{for all} \quad j \in \{1,..,N\}.$$

Optimal transport tackles the challenge of transporting $\mathbf{p}$ to $\mathbf{q}$ in an optimal manner. This is achieved by considering a cost matrix C_{ij} that's chosen to represent the Euclidean distance between x_i and y_j, defined as $C_{ij} = \|x_i - y_j\|_2$. The optimal transport is defined by the optimization problem

$$\begin{aligned} &\underset{Q\in\mathbb{R}^{K\times N}}{\text{minimize}} \quad \langle C, Q\rangle_F = \sum_{i=1}^{K}\sum_{j=1}^{N} C_{ij}Q_{ij} \\ &\text{subject to} \quad Q\mathbf{1}_N = \mathbf{q},\ Q^T\mathbf{1}_K = \mathbf{p},\ (Q)_{i,j} \geq 0 \end{aligned} \tag{1}$$

where the cumulative cost of the transportation plan is denoted by the Frobenius dot-product of two matrices $\langle \cdot, \cdot \rangle$ and where $\mathbf{1}_N$ and $\mathbf{1}_K$ represent vectors filled with ones, both of dimension N and K respectively. We establish the collection of all permissible couplings or transport plans, denoted as $\mathbf{Q}(\mathbf{p}, \mathbf{q})$, for histograms in the following manner:

$$\mathbf{Q}(\mathbf{p}, \mathbf{q}) = \{Q \in \mathbb{R}_+^{K\times N} \mid Q\mathbf{1}_N = \mathbf{q}, Q^T\mathbf{1}_K = \mathbf{p}\}.$$

To be more precise, particularly when the cost C takes the form of a distance matrix, the optimal transport corresponds to the Wasserstein distance over the space $\mathcal{H}_N \times \mathcal{H}_K$. This distance is defined as follows:

$$W(\mathbf{p}, \mathbf{q}) = \min_{Q\in\mathbf{Q}(\mathbf{p},\mathbf{q})} \langle C, Q\rangle_F \tag{2}$$

We used a simplex method [12] to solve (2), which involves moving iteratively from one feasible solution to another along the edges of the feasible region until an optimal solution is reached. The method utilizes a tableau representation, with each iteration selecting a pivot column and row to update the tableau. The algorithm continues until optimality conditions are met.

B. Wasserstein distance on graphs

The inadequacy of the current aggregation step in $\mathcal{R}$-convolution graph kernels, which might obscure significant differences in substructures due to averaging, has motivated us to develop a more detailed distance measure between structures and their constituents. Our approach is built upon the following sequential stages: (1) converting each graph into a collection of node embeddings, (2) quantifying the Wasserstein distance between every pair of graphs, and (3) generating a similarity matrix that will be employed within the learning algorithm. Fig. 1 represents the outlined steps. We first establish a proposed and novel embedding scheme and proceed to demonstrate the integration of embeddings within the framework of the Wasserstein distance.

The Weisfeiler–Lehman kernels: The Weisfeiler–Lehman graph kernels [13] offer an efficient solution for feature extraction on graphs characterized by discrete node labels. It is based on the Weisfeiler-Lehman test of isomorphism on graphs. This method transforms the initial graph into a sequence of graphs, where the node attributes of these graphs encapsulate both topological and label-related information. It looks and compares kernel comparing subtree-like patterns. Its process involves generating a sequence of ordered strings by amalgamating the labels of a node and its neighboring nodes. These strings are then hashed to generate compressed node labels, effectively capturing a broader context. As the algorithm progresses through iterations, these labels encompass progressively larger neighborhoods around each node, enabling the comparison of more extensive substructures.

Extension to continuous attributes: The Weisfeiler–Lehman kernels heavily rely on discrete node labels. If the graph data contains continuous or more complex attributes, the discretization process might lead to information loss. To overcome this limitation, we propose an embedding scheme for graphs characterized by continuous node attributes and weighted edges. The key concept

TABLE I

INFORMATION ABOUT 6 BENCHMARK DATASETS USED IN OUR EXPERIMENTS, INCLUDING THEIR SIZE USED IN A CLASSIFICATION TASK, AND A BRIEF DESCRIPTION. THE FIRST FOUR DATASETS ARE VECTOR ATTRIBUTED GRAPHS WHILE THE NEXT TWO CONTAIN GRAPHS WITH DISCRETE ATTRIBUTES.

Type	Dataset	Size	Description
Vector Attributes	**BZR**	405	Ligends classified into holding an active or inactive Benzodiazepine Receptor (BZR) site.
	COX-2	467	Cyclooxygenase-2 (COX-2) inhibitors classified into an active or inactive against human recombinant enzyme.
	ENZYMES	600	A set of protein tertiary structures categorized based on the type of reaction they catalyze.
	PROTEINS	1113	Two categories of proteins: enzymes (functional) and non-enzymes (non-functional).
Discrete Attributes	**MUTAG**	188	Whether a chemical compound is mutagenic aromatic and heteroaromatic nitro or non mutagenic.
	PTC-MR	344	Chemical compounds labeled based on their carcinogenicity in male rates (MR) rodents.

of our approach revolves around formulating a dual-tier explicit propagation approach that harnesses and enhances the existing node features through an averaging process spanning both the immediate and secondary neighborhoods. While similar methodologies have been indirectly explored in the context of calculating node-level features, they are either dependent on extra hashing procedures to handle the continuous attributes [14], [15] or they only consider the first immediate neighborhood [16]. Furthermore, we can readily incorporate edge weights by factoring them into the computation of the average within each neighborhood. Let's assume we have a continuous attribute denoted as $a^0(v) = a(v)$ for every node $v \in G$. In this scenario, our recursive embedding scheme is defined by

$$a^{m+1}(v) = \frac{1}{3}\Bigg(a^m(v) + \frac{1}{\deg(v)} \sum_{u \in \mathcal{N}(v)} \alpha((v,u)) \cdot a^m(u) + \frac{1}{\deg(v)} \sum_{u \in \mathcal{N}(v)} \frac{1}{\deg(u)} \sum_{\substack{w \in \mathcal{N}(u) \\ w \neq v}} \alpha((u,w)) \cdot a^m(w)\Bigg),$$

where α is an edge weight, and $\mathcal{N}(u)$ represents the neighborhood of node u, which includes the neighbors of u. In cases where edge weights are absent, we assign α a value of 1. The inclusion of the $1/3$ factor is to maintain consistent scale of features throughout the iterations. The ability of our proposed scheme to incorporate edge weights renders it suitable for diverse graph types, eliminating the need for an additional hashing procedure. Furthermore, it bears similarity to the propagation strategies employed in numerous graph neural networks, a technique that has demonstrated efficacy in node classification tasks involving large datasets [17].

Wasserstein approach for attributed graph alignment: In this paper, we focus on comparing attributed graphs that hold continuous attributes as well as graphs with discrete attributes on the vertices. In a more formal formulation, we consider undirected labeled graphs, which can be represented as tuples in the following structure $\mathcal{G}(\mathcal{V}, \mathcal{E}, f)$ where $(\mathcal{V}, \mathcal{E})$ are the set of vertices and edges of the graph. f is a labeling function that assigns each vertex $v_i \in \mathcal{V}$ with a feature $a_i = f(v_i)$ in some feature metric space. We suggest enhancing the aforementioned graph by introducing a histogram, intended to convey the relative significance of the vertices within the graph. To achieve this, assuming the graph comprises N vertices, we allocate those vertices with weights h_i. Through this procedure, our graph will take the form $\mathcal{G}(\mathcal{V}, \mathcal{E}, f, h_\mathcal{G})$, where $h_\mathcal{G}$ is a function that associates a weight to each vertex, such that $h_i = h_\mathcal{G}(v_i)$. This definition permits the graph to be depicted as a probability measure with complete support across the feature space. When the weights are all equal $h_i = \frac{1}{N}$, every vertex holds identical relative significance.

We aim at defining a matching distance between two graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ with N and K vertices, respectively. The two graphs are described respectively by their probability measure $h_{\mathcal{G}_1}$ and $h_{\mathcal{G}_2}$. We introduce $\mathbf{Q}(h_{\mathcal{G}_1}, h_{\mathcal{G}_2})$ as the set of all permissible couplings between $h_{\mathcal{G}_1}$ and $h_{\mathcal{G}_2}$, as follows:

$$\mathbf{Q}(h_{\mathcal{G}_1}, h_{\mathcal{G}_2}) = \{Q \in \mathbb{R}_+^{K \times N} \mid Q\mathbf{1}_N = h_{\mathcal{G}_2}, Q^T\mathbf{1}_K = h_{\mathcal{G}_1}\}$$

In this regard, the matrix Q depicts a probabilistic matching of nodes between the two graphs. We also consider the cost matrix $C \in \mathbb{R}^{K \times N}$ that stands for the distance between the features (i.e., embedding). This optimization problem can be expressed as

$$\begin{aligned} &\underset{Q \in \mathbb{R}_+^{K \times N}}{\text{minimize}} \quad \langle C, Q \rangle_F \\ &\text{subject to} \quad Q\mathbf{1}_N = h_{\mathcal{G}_2},\ Q^T\mathbf{1}_K = h_{\mathcal{G}_1},\ (Q)_{i,j} \geq 0. \end{aligned} \tag{3}$$

After generating the node embeddings using the graph embedding scheme, we proceed to assess the pairwise Wasserstein distance between the graphs. This begins with the computation of ground distances for every node pair. In the case of categorical (discrete) node features, we employ the normalized Hamming distance, which equals 1 when two vectors share no common features and 0 when the vectors are identical. For continuous node features, on the other hand, we employ the Euclidean distance. Subsequently, we insert the ground distance into (3) and calculate the Wasserstein distance utilizing a network simplex method [18].

III. NUMERICAL EXPERIMENTS

We use the Python implementation, Python Optimal Transport package, to solve the optimization problem and compute the Wasserstein distance between two graphs [19]. Our code will be released upon publication.

TABLE II

AVERAGE CLASSIFICATION ACCURACY ON THE GRAPH DATASETS WITH CONTINUOUS ATTRIBUTES. THE BEST RESULT FOR EACH DATASET IS HIGHLIGHTED IN **BOLD**.

	BZR	**COX2**	**ENZYMES**	**PROTEIN**
OTGK (Ours)	**85.82±1.5**	**79.72±2.15**	**74.01±6.76**	76.75±3.04
WWL	84.42±2.03	78.29±0.47	73.25±0.86	**77.91±0.88**
WD	84.01±3.02	75.2±2.31	72.4 ± 2.45	63.97±3.26
HOPPER	84.15±5.26	79.57±3.46	45.33±4.00	71.96±3.22
PROPAK	79.51±5.02	77.66±3.95	71.67±5.63	61.34±4.38
HGK-WLK	78.59±0.63	78.13±0.45	63.04±0.65	75.93±0.17
HGK-SP	76.42±0.72	72.57±1.18	66.36±0.37	75.78±0.17
PATCHY-SAN	82.20±4.23	71.91±3.40	27.33±4.16	71.79±3.39
VH-C	74.82±2.13	48.51±0.63	47.15±0.79	60.79±0.12

TABLE III

AVERAGE CLASSIFICATION ACCURACY ON THE GRAPH DATASETS WITH DISCRETE ATTRIBUTES. THE BEST RESULT FOR EACH DATASET IS HIGHLIGHTED IN **BOLD**.

	MUTAG	**PTC-MR**
OTGK (Ours)	**92.1±1.2**	**66.8±3.4**
GK K=3	82.4±8.4	56.4±8.0
RWK	79.4±8.1	55.0±7.3
SPK	82.9±8.1	60.0±7.3
WLK	86.2±8.4	62.8±7.2
PATCHY-SAN	83.0±10.8	55.3±8.2

Data sets: We evaluate our framework on 6 commonly used benchmark datasets, and these datasets are categorized into 2 distinct groups. BZR, COX2 [20], PROTEINS, ENZYMES [21], are vector attributed graphs. MUTAG [22] and PTC-MR [4] contain graphs featuring discrete attributes originating from small molecules. All data are available in [23]. Datasets used in this study are described in Table I.

Experimental setup: Regarding the feature distance matrix C between node features, for graphs possessing discrete attributes, we adopt a Weisfeiler-Lehman (WL) labeling approach, achieved by concatenating the labels of neighboring nodes. For graphs featuring continuous attributes, we implement our suggested embedding scheme outlined in Sec. II, which involves assigning weights to the attributes of neighboring nodes as well as those of second-level neighbors, and subsequently calculating their average.

Derived from the graph Wasserstein distance, a similarity matrix denoted as M_{WD} can be formulated for integration into a learning algorithm. We run two tasks and show the results of the classification task. For the classification task, we run a support vector machine using the indefinite kernel matrix $e^{-\lambda M_{\text{WD}}}$, which is an instance of a Laplacian kernel and seen as a noisy observation of the true positive semidefinite kernel [24]. We name our kernel **Optimal Transport-based Graph Kernel (OTGK)**. We assess classification accuracy through comparison with state-of-the-art graph kernel methods.

For vector attributed graphs, we compared our proposed method with the Wasserstein Weisfeiler–Lehman (WWL) kernel that averages only over the direct neighbors [16] and without embedding the features (WD). Furthermore, we compared with the graph HOPPER kernel [27] and the propagation kernel (PROPAK) [15], we also conducted a comparison with two variations of the hash graph kernel, namely HGK-SP and HGK-WL [14]. We also compare our method with the PATCHY-SAN framework for convolutional neural networks on graphs (PATCHY-SAN) [28]. In addition, we compared to a continuous vertex histogram (VH-C), which is defined as an RBF kernel between the aggregate of graph node embeddings.

For the categorical case, we compared with the random walk kernel (RWK) [25], the graphlet count kernel (GK) [26], with size of the graphlets K=3, the Weisfeler-Lehman kernel (WLK) [13], and the shortest path kernel (SPK) [21].

Results and discussion: In our experiments, we use 10-fold cross-validation, and we repeat each cross-validation split 10 times and report the average accuracy. For the vector attributed graphs, the average accuracy values presented in Table II indicate that OTGK stands out as a leading-edge technique, achieving the top performance on 3 out of the 4 datasets. Furthermore, the our proposed method comes in close second for the fourth dataset, PROTEIN. For the discrete labeled graphs, Table III shows that our method outperforms all competitive methods. The consistency of other competitive methods is lower, as they demonstrate strong performances on specific datasets only. We also observe that our approach utilizing WL attributes achieves better performance compared to the WL kernel method, highlighting the advantage of employing an optimal transport-based distance instead of a kernel-based similarity. Furthermore, our approach surpasses convolutional neural networks in graph-related tasks by a substantial margin, which is particularly noteworthy.

IV. CONCLUSIONS

In this work, we introduce a novel learning method based on optimal transport, aimed at achieving attributed graph matching and the process of learning node embeddings within a single framework. We show that learning features from first- and second-level neighbors is beneficial to the

objective of obtaining higher accuracy performance in various matching tasks. Our experiments demonstrate that our proposed method outperforms the state-of-the-art approaches for graph classification in both continuous node attribute scenarios and categorical settings, showcasing its potential to contribute significantly to the field of graph alignment methodologies. The ability to align and compare graphs with high-dimensional continuous node attributes positions our approach as a promising solution for tackling complex cases in biological research.

REFERENCES

[1] P. Yanardag and S. V. N. Vishwanathan, 'Deep graph kernels', in Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2015, pp. 1365–1374.

[2] R. Singh, J. Xu, and B. Berger, 'Global alignment of multiple protein interaction networks with application to functional orthology detection', Proceedings of the National Academy of Sciences, vol. 105, no. 35, pp. 12763–12768, 2008.

[3] O. Kuchaiev, T. Milenković, V. Memišević, W. Hayes, and N. Pržulj, 'Topological network alignment uncovers biological function and phylogeny', J. of the Royal Society Interface, vol. 7, no. 50, pp. 1341–1354, 2010.

[4] N. M. Kriege, P.-L. Giscard, and R. Wilson, 'On valid optimal assignment kernels and applications to graph classification', Advances in Neural Information Processing Systems, vol. 29, 2016.

[5] F. Santambrogio, Optimal Transport for Applied Mathematicians. Birkhäuser, 2015.

[6] M. Aburidi and R. F. Marcia, 'Optimal Transport-Based Graph Kernels for Drug Property Prediction', Scientific reports, 2024.

[7] M. Aburidi and R. Marcia, 'Wasserstein Distance-Based Graph Kernel for Enhancing Drug Safety and Efficacy Prediction *', in 2024 IEEE First International Conference on Artificial Intelligence for Medicine, Health and Care (AIMHC), 2024, pp. 113–119.

[8] M. Aburidi and R. Marcia, 'Adversarial Attack and Training for Graph Convolutional Networks using Focal Loss-Projected Momentum', in 2024 3rd IEEE International Conference on Computing and Machine Intelligence (ICMI), 2024.

[9] M. Aburidi and R. Marcia, 'Topological Adversarial Attacks on Graph Neural Networks via Projected Meta Learning', in IEEE International Conference on Evolving and Adaptive Intelligent Systems 2024 (IEEE EAIS 2024), 2024.

[10] M. Aburidi and R. Marcia, 'Optimal Transport and Contrastive-Based Clustering for Annotation-Free Tissue Analysis in Histopathology Images', in 2023 International Conference on Machine Learning and Applications (ICMLA), 2023, pp. 301–307.

[11] M. Aburidi and R. Marcia, 'CLOT: Contrastive Learning-Driven and Optimal Transport-Based Training for Simultaneous Clustering', in 2023 IEEE International Conference on Image Processing (ICIP), 2023, pp. 1515–1519.

[12] G. B. Dantzig, 'Programming of interdependent activities: I. General discussion', Econometrica: Journal of the Econometric Society, vol. 16, no. 3, pp. 203–210, 1947.

[13] N. Shervashidze and K. Borgwardt, 'Fast subtree kernels on graphs', Advances in Neural Information Processing Systems, vol. 22, 2009.

[14] C. Morris, N. M. Kriege, K. Kersting, and P. Mutzel, 'Faster kernels for graphs with continuous attributes via hashing', in 2016 IEEE 16th International Conference on Data Mining (ICDM), 2016, pp. 1095–1100.

[15] M. Neumann, R. Garnett, C. Bauckhage, and K. Kersting, 'Propagation kernels: efficient graph kernels from propagated information', Machine learning, vol. 102, pp. 209–245, 2016.

[16] M. Togninalli, E. Ghisu, F. Llinares-López, B. Rieck, and K. Borgwardt, 'Wasserstein Weisfeiler-Lehman graph kernels', Advances in Neural Information Processing Systems, vol. 32, 2019.

[17] J. Klicpera, A. Bojchevski, and S. Günnemann, 'Combining neural networks with personalized pagerank for classification on graphs', in International Conference on Learning Representations, 2019.

[18] GG. Peyré, M. Cuturi, and Others, 'Computational optimal transport', Center for Research in Economics and Statistics Working Papers, no. 2017–86, 2017.

[19] R. Flamary et al., 'Pot: Python optimal transport', The Journal of Machine Learning Research, vol. 22, no. 1, pp. 3571–3578, 2021.

[20] J. J. Sutherland, L. A. O'brien, and D. F. Weaver, 'Spline-fitting with a genetic algorithm: A method for developing classification structure-activity relationships', Journal of Chemical Information and Computer Sciences, vol. 43, no. 6, pp. 1906–1915, 2003.

[21] K. M. Borgwardt and H.-P. Kriegel, 'Shortest-path kernels on graphs', in Fifth IEEE International Conference on Data Mining (ICDM'05), 2005, pp. 8-pp.

[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] K. Kersting, N. M. Kriege, C. Morris, P. Mutzel, and M. Neumann, 'Benchmark data sets for graph kernels', 2016.

[24] R. Luss and A. d'Aspremont, 'Support vector machine classification with indefinite kernels', Advances in Neural Information Processing Systems, vol. 20, 2007.

[25] T. Gärtner, P. Flach, and S. Wrobel, 'On graph kernels: Hardness results and efficient alternatives', in Learning Theory and Kernel Machines: 16th Annual Conference on Learning Theory and 7th Kernel Workshop, COLT/Kernel 2003, Washington, DC, USA, August 24-27, 2003. Proceedings, 2003, pp. 129–143.

[26] N. Shervashidze, S. V. N. Vishwanathan, T. Petri, K. Mehlhorn, and K. Borgwardt, 'Efficient graphlet kernels for large graph comparison', in Artificial Intelligence and Statistics, 2009, pp. 488–495.

[27] A. Feragen, N. Kasenburg, J. Petersen, M. de Bruijne, and K. Borgwardt, 'Scalable kernels for graphs with continuous attributes', Advances in Neural Information Processing Systems, vol. 26, 2013.

[28] M. Niepert, M. Ahmed, and K. Kutzkov, 'Learning convolutional neural networks for graphs', in International Conference on Machine Learning, 2016, pp. 2014–2023.