

# Classification des Signaux sur Graphes par Mesures Spectrales Algébriques

H BAY AHMED, Abdel-Ouahab BOUDRAA, Delphine DARE, Yves PRÉAUX

IRENav, Ecole Navale/Arts-Métiers ParisTech, CC 600,29240 Brest Cedex 9, France.  
(hadj\_ahmed.bay\_ahmed,boudra,dare,yves.preaux)@ecole-navale.fr

**Résumé** – La notion de mesure de similarité est très importante dans de nombreux domaines tels que l'apprentissage statistique, la fouille de données ou les sciences cognitives. Dans cet article, nous nous intéressons à la similarité des signaux sur graphes et nous proposons deux nouvelles mesures de similarité spectrales, compactes et efficaces, basées sur la comparaison des spectres propres des graphes, appelées Covariance Spectrale (CS) et Similarité Spectrale Conjointe (SSC). Combinées à un noyau de diffusion sur graphe, ces nouvelles mesures ont permis d'obtenir des performances de classification excellentes sur des données moléculaires réelles, montrant ainsi la pertinence des valeurs propres pour la classification des signaux sur graphes. Les résultats sont comparés à ceux obtenus par les algorithmes k-NN et SVM appliqués sur des graphes projetés dans un espace vectoriel.

**Abstract** – The notion of similarity measure is very important in various domains such as machine learning, data mining or cognitive sciences. In this work, we are interested in the study of graph signals similarity for classification purpose. Two new spectral similarity measures based on the matching of eigen spectrums of graph signals, termed as spectral covariance (CS) and joint spectral similarity (SSC) are proposed. Combined with graph diffusion kernel, these new metrics give excellent classification performances on real molecular data thus, showing the pertinence of eigenvalues for the classification of signals on the graphs. Results are compared to those obtained with k-NN and SVM algorithms applied on embedded graphs.

## 1 Introduction

Le traitement du signal sur graphes est un sujet émergent qui a suscité ces dernières années beaucoup d'intérêt dans de multiples domaines [1],[2],[3],[4],[5]. Les graphes constituent des structures bien appropriées à la représentation des relations complexes pour les données multi-dimensionnelles. Le graphe peut être construit à partir des données pour capturer la géométrie sous-jacente. Le traitement du signal sur graphes vise à généraliser les outils pour l'analyse, le débruitage, la classification, la détection, la compression des signaux conventionnels à ceux définis sur graphes. L'objectif principal est de pouvoir traiter les données en prenant en considération les interactions existant entre les différentes entités. La mise sous forme de graphe des données multi-dimensionnelles a entraîné le développement d'un grand nombre d'algorithmes pour étudier les propriétés des graphes, particulièrement les algorithmes de comparaison de graphes pour mesurer leur similarité notamment dans l'étude des réseaux sociaux et les applications web [6], la détection des cyber attaques et des anomalies dans les réseaux informatiques [7], l'étude de l'évolution des connexions dans le cerveau humain [8], la classification des molécules biologiques [9]. La majorité des algorithmes vise à déterminer la similarité structurelle des graphes, c'est le cas entre autres des méthodes basées sur le coût de modification de graphe [10], l'isomorphisme inter-graphes [11], la marche aléatoire [12] ou bien la distance de Levenshtein [13]. La comparaison spectrale des graphes occupe aussi une place importante dans la

littérature, comme les travaux basés sur la cospectralité, l'excentricité spectrale [14], la distance de Dirac et la distance de Wasserstein [15]. Nous proposons dans ce travail de traiter le problème de classification des signaux sur graphes en utilisant des mesures spectrales de similarité. Deux nouvelles mesures de similarité, appelées la Covariance Spectrale (CS) et la Similarité Spectrale Conjointe (SSC) sont proposées. Elles utilisent conjointement les propriétés énergétiques et structurelles des graphes à travers leur spectres propres. Pour un meilleur apprentissage, un noyau approprié qui reflète au mieux la structure du graphe est nécessaire. L'intérêt de la fonction noyau est de transformer les algorithmes, développés initialement pour les données vectorielles, pour les adapter aux données structurées comme les graphes [16]. Un des noyaux les plus utilisés dans  $\mathbb{R}^N$  est le noyau de diffusion, défini positif et solution de l'équation de la chaleur [17]. Dans ce travail, les deux mesures CS et la SSC sont intégrées dans le noyau de diffusion afin d'améliorer son pouvoir discriminatoire. Des tests sont réalisés pour classifier des molécules chimiques réelles. Les résultats obtenus ont confirmé la pertinence de ces nouvelles mesures.

## 2 Bases des Signaux sur Graphes

Le graphe est construit en utilisant un ensemble  $V = \{v_1, v_2, \dots, v_N\}$  de  $N$  éléments appelés nœuds et un ensemble  $E = V \times V$  de  $m$  paires d'éléments de  $V$ . Chaque paire  $E_k = \{v_i, v_j\}$  de  $E$

est une arête du graphe. Elle définit une relation d'adjacence entre les nœuds  $v_i$  et  $v_j$ . Un graphe est un ensemble de nœuds avec une fonction de connectivité entre ses éléments. Il peut être représenté par une matrice d'adjacence ( $\mathbf{A} = [a_{ij}]$ ) de taille  $(N \times N)$ . L'élément  $a_{ij}$  de la matrice est égale à un si le nœud  $v_i$  est connecté au nœud  $v_j$ , sinon,  $a_{ij}$  vaut zero. De par la symétrie des relations entre les nœuds d'un graphe non orienté, la matrice d'adjacence est également symétrique avec une diagonale à valeurs nulles. Une représentation alternative des graphes est donnée par la matrice laplacienne  $\mathbf{L} = \mathbf{D} - \mathbf{A}$ , avec  $\mathbf{D}$  une matrice diagonale où la  $i^{eme}$  entrée de la diagonale est le nombre de connexions du nœud  $v_i$ . Si le graphe est pondéré, le degré  $d_i$  est égal à la somme des poids  $w_{ij}$  des arêtes liées à  $v_i$ , et la matrice laplacienne prend la forme  $\mathbf{L} = \mathbf{D} - \mathbf{W}$ , où  $\mathbf{W}$  est la matrice des poids. La matrice laplacienne normalisée  $\mathcal{L}$  est définie par :

$$\mathcal{L} = \mathbf{I} - \mathbf{D}^{-\frac{1}{2}} \mathbf{W} \mathbf{D}^{-\frac{1}{2}} \quad (1)$$

où  $\mathbf{I} \in \mathbb{R}^{N \times N}$  est la matrice *identité* de taille  $(N, N)$ . Dans la suite, nous notons le graphe par  $\mathbf{G}(V, \mathbf{A}, \mathbf{L})$ , où  $\mathbf{A}$  et  $\mathbf{L}$  des matrices qui contiennent l'essentiel de l'information structurale du graphe. Un signal sur graphe est une projection d'un ensemble d'éléments  $\mathbf{X}$  complexes ou réels sur une structure de graphe particulier. Chaque élément  $X_i$  est projeté vers un nœud  $v_i$  du graphe. Ainsi le signal sur graphe est noté par  $\mathbf{G}(X, \mathbf{A}, \mathbf{L})$  avec les éléments de  $X$  comme nœuds.

### 3 Covariance Spectrale

Soit  $\mathcal{G}$  un ensemble de signaux sur graphes non orientés, et soient deux signaux sur graphes  $\mathbf{G}_1(X_1, \mathbf{A}_1, \mathbf{L}_1)$  et  $\mathbf{G}_2(X_2, \mathbf{A}_2, \mathbf{L}_2) \in \mathcal{G}$  avec  $\mathbf{A}_i$  la matrice d'adjacence et  $\mathbf{L}_i$  la matrice laplacienne. En diagonalisant la matrice d'adjacence/laplacienne, nous obtenons les spectres propres  $\lambda(\mathbf{G}_1)$ ,  $\mu(\mathbf{G}_2)$  des graphes avec  $N_1, N_2$  valeurs propres respectives ordonnées en ordre croissant. Si nous considérons que les spectres propres sont des variables aléatoires générées par des lois de probabilité  $p(\lambda), p(\mu)$  ayant chacune une variance finie ( $\sigma_\lambda^2, \sigma_\mu^2$ ), nous pouvons alors évaluer leur similarité par le biais de leur covariance. Les graphes sont similaires si leurs spectres respectifs sont fortement corrélés. La covariance spectrale (CS) entre  $\lambda$  et  $\mu$  pour  $l$  réalisations est donnée par :

$$\mathbf{CS}(\mathbf{G}_1, \mathbf{G}_2) = \widehat{\mathbf{Cov}}(\lambda, \mu) = \frac{1}{l-1} \sum_{i=1}^l (\lambda_i - \bar{\lambda})(\mu_i - \bar{\mu}) \quad (2)$$

$\mathbf{CS}(\mathbf{G}_1, \mathbf{G}_2)$  est le degré de similarité des graphes  $\mathbf{G}_1(X_1, \mathbf{A}_1, \mathbf{L}_1)$ ,  $\mathbf{G}_2(X_2, \mathbf{A}_2, \mathbf{L}_2)$  avec  $\bar{\lambda}, \bar{\mu}$  les moyennes statistiques et  $l = \min(N_1, N_2)$ . La version normalisée est donnée par :

$$\mathbf{CS}_{\text{Norm}} = |\mathbf{CS}(\mathbf{G}_1, \mathbf{G}_2)| / \mathbf{CS}_{\text{max}} \quad (3)$$

où  $\mathbf{CS}_{\text{max}} = \max_{\mathbf{G}_1, \mathbf{G}_2 \in \mathcal{G}} |\mathbf{CS}(\mathbf{G}_1, \mathbf{G}_2)|$ ,  $\mathbf{CS}_{\text{Norm}} \in [0, 1]$

## 4 Similarité Spectrale Conjointe

Soient deux signaux sur graphes  $\mathbf{G}_1(X_1, \mathbf{A}_1, \mathbf{L}_1)$  et  $\mathbf{G}_2(X_2, \mathbf{A}_2, \mathbf{L}_2) \in \mathcal{G}$  avec  $\lambda_{1i}, \lambda_{2i}$  les spectres propres de leur matrices laplaciennes et  $\omega_{1i}, \omega_{2i}$  celles de leur matrices d'adjacence. La similarité spectrale conjointe (SSC) vise à exploiter conjointement les propriétés des spectres propres des matrices d'adjacence et laplaciennes afin, de discriminer au mieux les graphes à comparer. Elle est définie comme suit :

$$\mathbf{SSC}(\mathbf{G}_1, \mathbf{G}_2) = \alpha \mathbf{SSC}_{\mathbf{L}}(\mathbf{G}_1, \mathbf{G}_2) + (1-\alpha) \mathbf{SSC}_{\mathbf{A}}(\mathbf{G}_1, \mathbf{G}_2) \quad (4)$$

avec  $\alpha \in [0, 1]$  un facteur de pondération et :

$$\mathbf{SSC}_{\mathbf{L}}(\mathbf{G}_1, \mathbf{G}_2) = \sum_{i=1}^k (\lambda_{1i} - \lambda_{2i})^2 \quad (5)$$

$$\mathbf{SSC}_{\mathbf{A}}(\mathbf{G}_1, \mathbf{G}_2) = \sum_{i=1}^l (\omega_{1i} - \omega_{2i})^2 \quad (6)$$

$$k = \arg \min_{j \in \{1,2\}} \left( \sum_{i=1}^k \lambda_{ji} \geq \rho_L \right), l = \arg \min_{j \in \{1,2\}} \left( \sum_{i=1}^l \omega_{ji} \geq \rho_A \right) \quad (7)$$

Nous gardons les premières  $\{k, l\}$  valeurs propres qui contiennent les pourcentages  $\{\rho_L, \rho_A\}$  des énergies totales des spectres. Les seuils d'énergie sont choisis selon le nombre de valeurs propres à prendre en considération. La mesure (4) est normalisée comme suit :

$$\mathbf{SSC}_{\text{Norm}} = \mathbf{SSC}(\mathbf{G}_1, \mathbf{G}_2) / \mathbf{SSC}_{\text{max}} \quad (8)$$

où  $\mathbf{SSC}_{\text{max}} = \max_{\mathbf{G}_1, \mathbf{G}_2 \in \mathcal{G}} (\mathbf{SSC}(\mathbf{G}_1, \mathbf{G}_2))$ ,  $\mathbf{SSC}_{\text{Norm}} \in [0, 1]$

L'utilisation de la distance maximale pour la normalisation est valide seulement dans le cas de graphes de tailles proches pour éviter un grand écart du maximum par rapport à la moyenne. Dans le cas de la CS, nous prenons en compte les  $l$  premières valeurs propres des spectres des signaux sur graphes, ce qui diminue la plage de variation de la distance. Même chose pour la SSC en utilisant cette fois ci les seuils d'énergie  $\rho_A$  et  $\rho_L$ . Dans ce travail, les vecteurs de données  $X_1$  et  $X_2$  n'interviennent pas directement dans le calcul des distances CS et SSC, mais ils pourront être pris en compte dans le calcul des poids de la matrice d'adjacence.

## 5 Noyau de Diffusion pour Graphes

Dans les approches à noyaux, les patterns sont représentés par une fonction de similarité mutuelle au lieu des vecteurs d'attributs individuels. Ils permettent l'entraînement de la machine d'apprentissage dans un espace de grande dimension pour améliorer la probabilité de séparation des classes. Les valeurs du noyau sont obtenues à partir d'une fonction symétrique, définie positive [18]. Ce *noyau valide* est appelé *noyau de Mercer* [19]. Dans la littérature, nous trouvons plusieurs types de noyaux valides [20],[21]. Parmi eux il y a le noyau de diffusion qui a été étudié pour la première fois par Kondor et Laferty [22]. Ils ont utilisé la fonction exponentielle pour générer

une famille de noyaux sur graphes, en s’inspirant de la solution de l’équation de la chaleur. L’idée de base est d’utiliser une mesure de similarité symétrique ( $\text{SSC}_{\text{Norm}}$ ,  $\text{CS}_{\text{Norm}}$ ) pour construire une matrice de noyau  $\mathbf{K}$  définie positive et valide [22]. Considérons un ensemble de graphes  $\{g_1, g_2, \dots, g_M\} \subseteq \mathcal{G}$ , un facteur de pondération  $0 < \beta < 1$  et des mesures de similarité  $\mathcal{G} \times \mathcal{G} \rightarrow \mathbb{R}$ . La matrice  $\mathbf{S} = (s_{ij})_{M \times M}$  des similarités mutuelles entre les graphes de taille  $(M \times M)$  peut être convertie en une matrice de noyau définie positive  $\mathbf{K} = (\kappa_{ij})_{M \times M}$  en utilisant le *noyau exponentiel de diffusion* [22] comme suit :

$$\mathbf{K} = \sum_{p=0}^{\infty} \frac{1}{p!} \beta^p \mathbf{S}^p = \exp(\beta \mathbf{S}) \quad (9)$$

Nous intégrons ce noyau dans l’algorithme d’apprentissage SVM pour la classification des graphes.

## 6 Bases de Données des Signaux

### 6.1 Graphes de Mutagénicité

Comme décrit dans [19], la mutagénicité est la capacité d’une molécule chimique à entraîner des mutations dans l’ADN. Une molécule mutagène interagit avec l’ADN conduisant à sa modification par l’ajout ou la suppression de sections spécifiques. Ainsi, les molécules mutagènes présentent un grand risque de toxicité, spécialement pour les humains. La base de données Chemical Carcinogenicity Research Information System (CCRIS) [23] contient 7000 molécules. Dans ce travail, les données utilisées ont été préparées originalement par les auteurs dans [19]. La base de données est composée de deux classes : molécules mutagènes et non-mutagènes. Afin de les convertir en signaux sur graphes, nous associons à chaque nœud un scalaire qui correspond à la masse atomique de la composante chimique par rapport à son isotope le plus proche, les poids  $W_{ij}$  des arêtes sont considérés comme une force d’attraction fictive entre les atomes individuels et ils sont calculés par la formule suivante :

$$W_{ij} = \begin{cases} (\text{valence} \times |m_i - m_j|)^2 & \text{if } m_i \neq m_j \\ (\text{valence})^2 & \text{if } m_i = m_j \end{cases} \quad (10)$$

où  $W_{ij}$  est le poids de l’arrête entre les nœuds  $i$  et  $j$  et les quantités  $m_i$  et  $m_j$  sont leurs masses atomiques respectives. Plus la différence des masses atomiques est importante, plus l’interaction est forte. Les connexions ayant une liaison double voient leurs forces d’attraction multipliées par deux. Il est important de souligner que la stratégie de pondération proposée (Eq. 10) est conceptuelle et non réelle. Notre ensemble d’entraînement est composé seulement de 200 graphes (100 mutagènes, 100 non-mutagènes), afin de réduire les problèmes liés à la complexité et de tester la pertinence de nos approches dans le cas d’un ensemble d’entraînement petit. La taille de l’ensemble de test est de 2337 graphes.

### 6.2 Graphes de Protéines

Présentée dans [19], cette base de données contient des Protéines (enzymes) qui appartiennent à six classes (EC 1, EC 2, EC 3,

EC 4, EC 5 et EC 6). Les protéines sont converties en graphes par la représentation des éléments de leur structure secondaire sous forme de nœuds et d’arêtes. Les nœuds sont étiquetés selon leur type (helix, sheet, ou loop) et la séquence de leur acide aminé (e.g. TFKEVVRLT). Chaque nœud est connecté par des arêtes à ses trois voisins les plus proches dans l’espace. Les arêtes sont étiquetées par leur type et leur distance en Angströms. Pour obtenir les signaux sur graphes, nous associons à chaque nœud la somme des masses des atomes qui composent sa séquence. Il y a 600 protéines au total, 100 par classe. Nous utilisons un ensemble d’entraînement de taille (200) et un ensemble de test de taille (400). La classification de ces données consiste à déterminer la classe de l’enzyme.

## 7 Résultats

Les nouvelles mesures ont été intégrées dans le noyau de diffusion pour graphes et testées pour résoudre les problèmes de classification décrits ci-dessus. Pour les deux bases de données, nous avons utilisé des ensembles d’entraînement de plus petite taille (200 éléments) par rapport à ceux exploités dans [19] pour réduire le temps et la complexité du calcul. Pour procéder à la comparaison, la taille des ensembles de test est identique à celle employée par Riesen et Bunke [24] soit 2337 éléments pour la mutagénicité et 400 pour les protéines. Nous reportons dans la Table I, les résultats de deux approches proposées dans [19] [24] (k-NN et SVM sur graphes encastés) et les résultats de nos approches correspondant aux paramètres  $\beta = 0.1$ ,  $\alpha = 0.7$ ,  $\rho_L = \rho_A = 0.9$ . Ces résultats montrent que les méthodes basées sur les mesures CS et SSC conduisent à des taux de classification de 100% pour la mutagénicité, quelle que soit la matrice utilisée. En utilisant que 8% des données d’entraînement d’origine, nous avons amélioré le taux de classification de 28%. Les résultats obtenus ont été validés en fixant les valeurs des paramètres  $\beta, \alpha, \rho_L, \rho_A$  expérimentalement. Une stratégie pour l’optimisation du choix de ces paramètres est nécessaire.

TABLE 1 – Taux de classification relatifs aux Bases de Données Mutagénicité et Protéines.

Algorithme/Bases de Données	Mutagénicité	Protéines
k-NN/ Graphes Encastés	67%	95%
SVM/ Graphes Encastés	72%	97.3%
SVM-CS Adjacence	<b>100%</b>	<b>98.9%</b>
SVM-CS Laplacienne	<b>100%</b>	<b>98.9%</b>
SVM-CS Laplacienne Normalisée	<b>100%</b>	<b>98.9%</b>
SVM-SSC Adjacence et Laplacienne	<b>98.0%</b>	<b>98.49%</b>
SVM-SSC Adjacence et Laplacienne Normalisée	<b>100%</b>	<b>98.9%</b>

Dans le cas du problème multi-classes des protéines, nous avons utilisé l’approche One-vs-Rest dans l’algorithme d’apprentis-

sage SVM. Les résultats sont très bons. Le classificateur a atteint 98.9% de bonnes réponses en utilisant les deux métriques **CS** et **SSC**, soit une amélioration de 1.3%. Dans l'ensemble les résultats obtenus sont très encourageants et montrent l'intérêt de construire des mesures de similarité, pour la classification des signaux graphes, basées sur les valeurs propres issues des matrices d'adjacence et laplacienne.

## 8 Conclusion

Dans ce travail, deux nouvelles mesures de similarité **CS** et **SSC** des signaux sur graphes ont été proposées. Elles représentent deux descripteurs optimisés, informatifs et efficaces pour la discrimination des graphes de petite taille. La **CS** et la **SSC** ont été testées dans le cadre d'une classification binaire et multiclassées à partir d'un jeu de données d'entraînement réduit. Les méthodes ont été appliquées sur des données moléculaires pour la prédiction de la mutagénicité et la sélection des protéines. Les résultats obtenus sont très prometteurs avec un faible taux d'erreurs de classification. Une projection simple et efficace des données moléculaires en des signaux sur graphes a été également proposée. Cette projection a amélioré nettement le pouvoir discriminatoire du classificateur. Nos résultats montrent la pertinence des valeurs propres pour la classification des signaux sur graphes. La non implication actuelle des données supportées par les nœuds dans le calcul des distances peut être évitée en proposant une pondération des arêtes dépendant justement de ces données. Nous comptons aussi étudier l'évolution des performances de classification en fonction des différents paramètres intervenant dans les mesures de similarité (**CS**,**SSC**) et dans le noyau de diffusion, et comparer nos approches avec d'autres distances spectrales définies pour les signaux sur graphes.

## Références

- [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 Sig. Proc. Mag.*, vol. 30, no. 3, pp. 83–98, 2013.
- [2] A. Sandryhaila and J. M. Moura, "Discrete signal processing on graphs : Frequency analysis," *IEEE Trans. Sig. Proc.*, vol. 62, no. 12, pp. 3042–3054, 2014.
- [3] N. Tremblay, "Réseaux et signal : des outils de traitement du signal pour l'analyse des réseaux," Ph.D. dissertation, ENS Lyon, 2014.
- [4] B. Girault, "Signal processing on graphs : Contributions to an emerging field," Ph.D. dissertation, ENS Lyon, 2015.
- [5] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 129–150, 2011.
- [6] Y. Shi and M. Macy, "Measuring structural similarity in large online networks," *Soc. Sc. Res.*, vol. 59, pp. 97–106, 2016.
- [7] G. S. Bopche and B. M. Mehtre, "Graph similarity metrics for assessing temporal changes in attack surface of dynamic networks," *Computers & Security*, vol. 64, pp. 16–43, 2017.
- [8] A. Mheich, M. Hassan, V. Gripon, M. Khalil, C. Berrou, O. Dufor, and F. Wendling, "A novel algorithm for measuring graph similarity : application to brain networks," in *Neural Eng., 2015 7th Int. IEEE/EMBS Conf.* IEEE, 2015, pp. 1068–1071.
- [9] P. D. Dobson and A. J. Doig, "Distinguishing enzyme structures from non-enzymes without alignments," *J. Mol. Biol.*, vol. 330, no. 4, pp. 771–783, 2003.
- [10] X. Gao, B. Xiao, D. Tao, and X. Li, "A survey of graph edit distance," *Patt. Anal. App.*, vol. 13, no. 1, pp. 113–129, 2010.
- [11] J. R. Ullmann, "An algorithm for subgraph isomorphism," *J. ACM*, vol. 23, no. 1, pp. 31–42, 1976.
- [12] F. Fouss, A. Pirotte, J.-M. Renders, and M. Saerens, "Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation," *IEEE Trans. Knowledge and Data Eng.*, vol. 19, no. 3, 2007.
- [13] B. Cao, Y. Li, and J. Yin, "Measuring similarity between graphs based on the levenshtein distance," *Appl. Math*, vol. 7, no. 1L, pp. 169–175, 2013.
- [14] I. Jovanović and Z. Stanić, "Spectral distances of graphs," *Linear Algebra and its Applications*, vol. 436, no. 5, pp. 1425–1435, 2012.
- [15] J. Gu, B. Hua, and S. Liu, "Spectral distances on graphs," *Discrete Applied Mathematics*, vol. 190, pp. 56–74, 2015.
- [16] F. Jäkel, B. Schölkopf, and F. A. Wichmann, "Similarity, kernels, and the triangle inequality," *J. Math. Psychology*, vol. 52, no. 5, pp. 297–303, 2008.
- [17] R. Kondor and J.-P. Vert, "Diffusion kernels," *kernel methods in computational biology*, pp. 171–192, 2004.
- [18] B. Schölkopf and A. J. Smola, *Learning with kernels : support vector machines, regularization, optimization, and beyond*. MIT press, 2002.
- [19] K. Riesen and H. Bunke, *Graph classification and clustering based on vector space embedding*. World Scientific, 2010, vol. 77.
- [20] T. Gärtner, "A survey of kernels for structured data," *ACM SIGKDD Explorations Newsletter*, vol. 5, no. 1, pp. 49–58, 2003.
- [21] N. Shervashidze, P. Schweitzer, E. J. v. Leeuwen, K. Mehlhorn, and K. M. Borgwardt, "Weisfeiler-lehman graph kernels," *J. Mach. Learning Research*, vol. 12, no. Sep, pp. 2539–2561, 2011.
- [22] R. I. Kondor and J. Lafferty, "Diffusion kernels on graphs and other discrete input spaces," in *ICML*, vol. 2, 2002, pp. 315–322.
- [23] T. Cameron, J. Stump, and L. Schofield, "Chemical carcinogenesis research information system (ccris) data bank, 1981-june 1986 (1988 version). data file," National Cancer Cancer Inst., Bethesda, MD (USA), Tech. Rep., 1986.
- [24] K. Riesen and H. Bunke, "Iam graph database repository for graph based pattern recognition and machine learning," *Structural, Syntactic, and Statistical Pattern Recognition*, pp. 287–297, 2008.