Graph Kernels

State-of-the-Art and Future Challenges

Other titles in Foundations and Trends® in Machine Learning

Computational Optimal Transport
Gabriel Peyre and Marco Cuturi
ISBN: 978-1-68083-550-2

An Introduction to Deep Reinforcement Learning
Vincent Francois-Lavet, Peter Henderson, Riashat Islam, Marc G. Bellemare and Joelle Pineau
ISBN: 978-1-68083-538-0

An Introduction to Wishart Matrix Moments
Adrian N. Bishop, Pierre Del Moral and Angele Niclas
ISBN: 978-1-68083-506-9

A Tutorial on Thompson Sampling
Daniel J. Russo, Benjamin Van Roy, Abbas Kazerouni, Ian Osband and Zheng Wen
ISBN: 978-1-68083-470-3

Graph Kernels

State-of-the-Art and Future Challenges

Karsten Borgwardt
Elisabetta Ghisu
Felipe Llinares-López
Leslie O'Bray
Bastian Rieck

Machine Learning and Computational Biology Lab
Department of Biosystems Science and Engineering
ETH Zurich
Basel, Switzerland
& Swiss Institute of Bioinformatics

now
the essence of knowledge
Boston — Delft

Foundations and Trends® in Machine Learning

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

K. Borgwardt, E. Ghisu, F. Llinares-López, L. O'Bray, B. Rieck. *Graph Kernels.* Foundations and Trends® in Machine Learning, vol. 13, no. 5-6, pp. 531–712, 2020.

ISBN: 978-1-68083-771-1

Foundations and Trends® in Machine Learning
Volume 13, Issue 5-6, 2020

Editorial Board

Editorial Scope

Topics

Foundations and Trends® in Machine Learning publishes survey and tutorial articles in the following topics:

- Adaptive control and signal processing
- Applications and case studies
- Behavioral, cognitive and neural learning
- Bayesian learning
- Classification and prediction
- Clustering
- Data mining
- Dimensionality reduction
- Evaluation
- Game theoretic learning
- Graphical models
- Independent component analysis
- Inductive logic programming
- Kernel methods
- Markov chain Monte Carlo
- Model choice
- Nonparametric methods
- Online learning
- Optimization
- Reinforcement learning
- Relational learning
- Robustness
- Spectral methods
- Statistical learning theory
- Variational inference
- Visualization

Information for Librarians

Foundations and Trends® in Machine Learning, 2020, Volume 13, 6 issues. ISSN paper version 1935-8237. ISSN online version 1935-8245. Also available as a combined paper and online subscription.

Contents

Graph Kernels

Karsten Borgwardt[1], Elisabetta Ghisu[1], Felipe Llinares-López[1], Leslie O'Bray[1] and Bastian Rieck[1]

[1]*Machine Learning and Computational Biology Lab, D-BSSE, ETH Zurich*
Swiss Institute of Bioinformatics
Basel, Switzerland
{karsten.borgwardt, elisabetta.ghisu, felipe.llinares, leslie.obray, bastian.rieck}@bsse.ethz.ch

ABSTRACT

Graph-structured data are an integral part of many application domains, including chemoinformatics, computational biology, neuroimaging, and social network analysis. Over the last two decades, numerous *graph kernels*, *i.e.* kernel functions between graphs, have been proposed to solve the problem of assessing the similarity between graphs, thereby making it possible to perform predictions in both classification and regression settings. This manuscript provides a review of existing graph kernels, their applications, software plus data resources, and an empirical comparison of state-of-the-art graph kernels.

Karsten Borgwardt, Elisabetta Ghisu, Felipe Llinares-López, Leslie O'Bray and Bastian Rieck (2020), Graph Kernels, Foundations and Trends® in Machine Learning: Vol. 13, No. 5-6, pp 531–712. DOI: 10.1561/2200000076.

1

Introduction

Among the data structures commonly used in machine learning, graphs are arguably one of the most general. Graphs allow modelling complex objects as a collection of entities (nodes) and of relationships between such entities (edges), each of which can be annotated by metadata such as categorical or vectorial node and edge features. Many ubiquitous data types can be understood as particular cases of graphs, including unstructured vectorial data as well as structured data types such as time series, images, volumetric data, point clouds or bags of entities, to name a few. Most importantly, numerous applications benefit from the extra flexibility that graph-based representations provide.

In *chemoinformatics*, graphs have been used extensively to represent molecular compounds (Trinajstic, 2018), with nodes corresponding to atoms, edges to chemical bonds, and node and edge features encoding known chemical properties of each atom and bond in the molecule. Machine learning approaches operating on such graph-based representations of molecules are becoming increasingly successful in learning to predict complex molecular properties from large annotated data sets (Azencott *et al.*, 2007; Duvenaud *et al.*, 2015; Gilmer *et al.*, 2017; Wu *et al.*, 2018), offering a promising set of tools for drug discovery (Vamathevan

et al., 2019). In *computational biology*, graphs have likewise risen to prominence due to their ability to describe multi-faceted interactions between (biological) entities. Examples of crucial importance include, but are not limited to, protein-protein interaction networks (Szklarczyk *et al.*, 2018), co-expression networks (Zhang and Horvath, 2005; Lonsdale *et al.*, 2013), metabolic pathways (Kanehisa and Goto, 2000), gene regulatory networks (Karlebach and Shamir, 2008), gene-phenotype association networks (Goh *et al.*, 2007), protein structures (Borgwardt *et al.*, 2005) and phylogenetic networks (Huson and Bryant, 2005). Graphs also play a key role in other application domains in the life sciences, such as *neuroscience*, where they are commonly used to concisely represent the brain connectivity patterns of different individuals (He and Evans, 2010), or *clinical machine learning*, where they have been employed to describe and exploit relationships between medical concepts by means of ontologies (Choi *et al.*, 2017) and knowledge graphs (Rotmensch *et al.*, 2017). In recent years, social network analysis has become a research field on its own, generating ever-larger graph data sets (Scott, 2011; Wasserman and Faust, 1994) and spanning a wide range of applications such as viral marketing (Leskovec *et al.*, 2007), community detection (Du *et al.*, 2007), influence estimation (Du *et al.*, 2013) or fake news detection (Tschiatschek *et al.*, 2018).

The great representational power of graph-structured data is however a source of important challenges for method development. Graphs are intrinsically discrete objects, containing a combinatorial number of substructures. As a result, even seemingly simple questions, such as determining whether two graphs are identical (*graph isomorphism*) or whether one graph is contained in another graph (*subgraph isomorphism*), are remarkably hard to solve in practice. In particular, no polynomial time algorithm is known for the former question, while the latter question is known to be NP-complete. Machine learning methods operating on graphs must therefore grapple with the need to balance computational tractability with the ability to leverage as much of the information conveyed by each graph as possible.

To this end, a popular family of early approaches, many of which were motivated by chemoinformatics (Todeschini and Consonni, 2008), aim to embed graphs into fixed-dimensional vectorial representations by

computing a set of hand-engineered features (also known as *topological descriptors*). However, designing these features often proved to be a daunting task, requiring substantial application-specific prior knowledge and potentially depending on which statistical learning algorithm was to be subsequently used to learn from the resulting vectorial representations. Moreover, the amount of topological information captured by these representations was not only limited by the need to maintain computational tractability, but also often in practice by the desire to obtain a parsimonious representation of low-to-moderate dimensionality.

Crucially, the popularisation of *kernel methods* in machine learning (Schölkopf and Smola, 2002) provided a principled way to ameliorate all of the aforementioned limitations. Put briefly, kernel methods represent objects by implicitly embedding them as elements of a reproducing kernel Hilbert space by means of a *positive definite kernel*, which explicitly quantifies the similarity between any pair of objects and is mathematically equivalent to the inner product between the corresponding embeddings. This allows kernel methods to lift the wealth of existing statistical approaches based on linear models for vectorial data to other settings, such as non-linear modelling of vectorial data or, as is the case for graphs, modelling of structured data for which a vectorial representation might not be available or might be too high-dimensional to use explicitly. Moreover, they accomplish this while allowing to control the capacity of the underlying model via regularisation (Hofmann *et al.*, 2008).

These aspects make kernel methods a great fit for machine learning on graph-structured data, as evidenced by the almost two decades of fruitful research on *graph kernels*[1] which we review in this manuscript. Existing graph kernels mainly differ in (i) the type of substructures they use to define the positive definitive kernel function that measures the similarity between two graphs, and (ii) the underlying algorithm used to efficiently evaluate this function. In this quest to construct increasingly informative and more computationally efficient approaches to quantify

[1] In this monograph, by *graph kernel* we refer to a kernel function between two *graphs*. Notice that the term *graph kernels* sometimes is also used to refer to the different subject of kernel functions between *nodes* of a single graph (*e.g.* Kondor and Lafferty (2002).)

the similarity of graphs, research on graph kernels has led to algorithms for supervised learning (Kriege *et al.*, 2020), dimensionality reduction (which can then be used to visualise graphs in a lower dimensional space) (Lee and Verleysen, 2012), and clustering (Aggarwal and Wang, 2010). Moreover, in doing so, the literature on graph kernels has produced a great amount of empirical results characterising the usefulness of different representations for graph-structured data in distinct application domains, which we exhaustively gather, reproduce and analyse. These experimental observations might not only pave the way to the development of novel graph kernels, but might also be of further use in the emerging field of graph neural networks, many of which can be understood as natural extensions of certain graph kernels in the context of representation learning (Xu *et al.*, 2019).

Before proceeding, we would like to mention two other recent graph kernel surveys and highlight how our review is different. Kriege *et al.* (2020) provide an excellent narrative overview of existing graph kernels; we additionally provide an in-depth description of the kernels. Their review is a good starting point for a researcher looking to understand the landscape at a high-level or looking for a reference on which graph kernel paper to read. Nikolentzos *et al.* (2019) provide more details about the kernels discussed; we additionally provide a conceptual categorisation of the kernels. Unlike these two reviews, our survey discusses trends and emerging topics in the field. Hence our review contributes to the literature in that it provides an in-depth description, categorisation and empirical comparison of graph kernels and gives a detailed outlook to the future of the field.

This review is divided into two parts: the first part focuses on the theoretical description of common graph kernels. After a short general introduction to graph theory and kernels in Chapter 2, we provide a detailed description, typology, and analysis of relevant graph kernels in Chapter 3. We take care to expose relations between different kernels and briefly comment on their applicability to certain types of data. The second part in Chapter 4 focuses on a large-scale empirical evaluation of graph kernels, as well as a description of desirable properties and requirements for benchmark data sets. We conclude our review with

an outline of future trends and open challenges for graph kernels in Chapter 5.

References

Aggarwal, C. and H. Wang. 2010. "A Survey of Clustering Algorithms for Graph Data". In: *Managing and Mining Graph Data*. Vol. 40. Springer. 275–301.

Aiolli, F., M. Donini, N. Navarin, and A. Sperduti. 2015. "Multiple Graph-Kernel Learning". In: *IEEE Symposium Series on Computational Intelligence*. 1607–1614.

Alon, N., M. Krivelevich, and B. Sudakov. 1998. "Finding a Large Hidden Clique in a Random Graph". *Random Structures & Algorithms*. 13(3–4): 457–466.

Aronszajn, N. 1950. "Theory of Reproducing Kernels". *Transactions of the American Mathematical Society*. 68(3): 337–404.

Arora, S. and B. Barak. 2009. *Computational Complexity: A Modern Approach*. Cambridge, United Kingdom: Cambridge University Press.

Azencott, C.-A., A. Ksikes, S. J. Swamidass, J. H. Chen, L. Ralaivola, and P. Baldi. 2007. "One- to Four-Dimensional Kernels for Virtual Screening and the Prediction of Physical, Chemical, and Biological Properties". *Journal of Chemical Information and Modeling*. 47(3): 965–974.

Babai, L. and L. Kucera. 1979. "Canonical Labelling of Graphs in Linear Average Time". In: *20th Annual Symposium on Foundations of Computer Science*. 39–46.

Babai, L. 2015. "Graph Isomorphism in Quasipolynomial Time". *arXiv e-prints*. arXiv: 1512.03547 [cs.DS].

Bach, F. R. 2008. "Graph Kernels between Point Clouds". In: *Proceedings of the 25th International Conference on Machine Learning*. 25–32.

Bagnall, A., J. Lines, A. Bostrom, J. Large, and E. Keogh. 2017. "The Great Time Series Classification Bake Off: a Review and Experimental Evaluation of Recent Algorithmic Advances". *Data Mining and Knowledge Discovery*. 31(3): 606–660.

Bai, L., E. R. Hancock, A. Torsello, and L. Rossi. 2013. "A Quantum Jensen–Shannon Graph Kernel Using the Continuous-time Quantum Walk". In: *Graph-Based Representations in Pattern Recognition*. Ed. by W. G. Kropatsch, N. M. Artner, Y. Haxhimusa, and X. Jiang. Heidelberg, Germany: Springer. 121–131.

Bai, L., L. Rossi, H. Bunke, and E. R. Hancock. 2014. "Attributed Graph Kernels Using the Jensen–Tsallis q-Differences". In: *Machine Learning and Knowledge Discovery in Databases*. Ed. by T. Calders, F. Esposito, E. Hüllermeier, and R. Meo. Heidelberg, Germany: Springer. 99–114.

Bai, L., Z. Zhang, P. Ren, L. Rossi, and E. R. Hancock. 2015. "An Edge-based Matching Kernel through Discrete-time Quantum Walks". In: *Image Analysis and Processing – ICIAP*. Ed. by V. Murino and E. Puppo. Cham, Switzerland: Springer. 27–38.

Bai, Y., H. Ding, S. Bian, T. Chen, Y. Sun, and W. Wang. 2018. "Graph Edit Distance Computation via Graph Neural Networks". *arXiv e-prints*. arXiv: 1808.05689 [cs.LG].

Berger, F., P. Gritzmann, and S. de Vries. 2009. "Minimum Cycle Bases and Their Applications". In: *Algorithmics of Large and Complex Networks: Design, Analysis, and Simulation*. Ed. by J. Lerner, D. Wagner, and K. A. Zweig. Heidelberg, Germany: Springer. 34–49.

Bernstein, M., V. De Silva, J. C. Langford, and J. B. Tenenbaum. 2000. "Graph Approximations to Geodesics on Embedded Manifolds". *Tech. rep.* Stanford University.

Blizard, W. D. 1988. "Multiset Theory". *Notre Dame Journal of Formal Logic*. 30(1): 36–66.

Bloom, B. H. 1970. "Space/Time Trade-offs in Hash Coding with Allowable Errors". *Communications of the ACM*. 13(7): 422–426.

Borg, I. and P. J. F. Groenen. 2005. *Modern Multidimensional Scaling. Theory and Applications*. 2nd ed. New York, NY, USA: Springer.

Borgwardt, K. M., C. S. Ong, S. Schönauer, S. Vishwanathan, A. J. Smola, and H.-P. Kriegel. 2005. "Protein Function Prediction via Graph Kernels". *Bioinformatics*. 21(suppl_1): i47–i56.

Borgwardt, K. and H.-P. Kriegel. 2005. "Shortest-Path Kernels on Graphs". In: *Proceedings of the Fifth IEEE International Conference on Data Mining*. Washington, DC, USA: IEEE Computer Society. 74–81.

Boser, B. E., I. M. Guyon, and V. N. Vapnik. 1992. "A Training Algorithm for Optimal Margin Classifiers". In: *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*. ACM. New York, NY, USA. 144–152.

Brouwer, A. E. and W. H. Haemers. 2012. *Spectra of Graphs*. New York, NY, USA: Springer.

Bunke, H. and K. Riesen. 2007. "A Family of Novel Graph Kernels for Structural Pattern Recognition". In: *Progress in Pattern Recognition, Image Analysis and Applications*. Ed. by L. Rueda, D. Mery, and J. Kittler. Heidelberg, Germany: Springer. 20–31.

Calvo, B. and G. Santafé. 2016. "`scmamp`: Statistical Comparison of Multiple Algorithms in Multiple Problems". *The R Journal*. 8(1): 248–256.

Ceroni, A., F. Costa, and P. Frasconi. 2007. "Classification of Small Molecules by Two- and Three-dimensional Decomposition Kernels". *Bioinformatics*. 23(16): 2038–2045.

Choi, E., M. T. Bahadori, L. Song, W. F. Stewart, and J. Sun. 2017. "`GRAM`: Graph-based Attention Model for Healthcare Representation Learning". In: *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM. New York, NY, USA. 787–795.

Choi, E., M. W. Dusenberry, G. Flores, Z. Xu, Y. Li, Y. Xue, and A. M. Dai. 2019. "Learning Graphical Structure of Electronic Health Records with Transformer for Predictive Healthcare". In: *ICML Workshop on Learning and Reasoning with Graph-Structured Data*.

Chung, F. R. K. 1997. *Spectral Graph Theory*. Providence, RI, USA: American Mathematical Society.

Correa, C. D. and P. Lindstrom. 2011. "Towards Robust Topology of Sparsely Sampled Data". *IEEE Transactions on Visualization and Computer Graphics*. 17(12): 1852–1861.

Costa, F. and K. De Grave. 2010. "Fast Neighborhood Subgraph Pairwise Distance Kernel". In: *Proceedings of the 26th International Conference on Machine Learning*. 255–262.

Curado, M., F. Escolano, E. R. Hancock, F. Nourbakhsh, and M. Pelillo. 2015. "Similarity Analysis from Limiting Quantum Walks". In: *Similarity-Based Pattern Recognition – SIMBAD*. Ed. by A. Feragen, M. Pelillo, and M. Loog. Cham, Switzerland: Springer. 38–53.

Dai, H., B. Dai, and L. Song. 2016. "Discriminative Embeddings of Latent Variable Models for Structured Data". In: *International Conference on Machine Learning*. 2702–2711.

Datar, M., N. Immorlica, P. Indyk, and V. S. Mirrokni. 2004. "Locality-Sensitive Hashing Scheme Based on p-stable Distributions". In: *Proceedings of the 20th Annual Symposium on Computational Geometry*. New York, NY, USA: ACM. 253–262.

Demšar, J. 2006. "Statistical Comparisons of Classifiers Over Multiple Data Sets". *Journal of Machine Learning Research*. 7: 1–30.

Dexiong Chen, L. J. and J. Mairal. 2020. "Convolutional Kernel Networks for Graph-Structured Data". In: *Proceedings of the 37th International Conference on Machine Learning (ICML)*. 2483–2453.

Dijkstra, E. W. 1959. "A Note on Two Problems in Connexion with Graphs". *Numerische Mathematik*. 1(1): 269–271.

Du, N., L. Song, M. G. Rodriguez, and H. Zha. 2013. "Scalable Influence Estimation in Continuous-time Diffusion Networks". In: *Advances in Neural information Processing Systems 26*. Ed. by C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger. 3147–3155.

Du, N., B. Wu, X. Pei, B. Wang, and L. Xu. 2007. "Community Detection in Large-Scale Social Networks". In: *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis*. ACM. New York, NY, USA. 16–25.

Du, S. S., K. Hou, B. Póczos, R. Salakhutdinov, R. Wang, and K. Xu. 2019. "Graph Neural Tangent Kernel: Fusing Graph Neural Networks with Graph Kernels". In: *Advances in Neural Information Processing Systems*. 5723–5733.

Duvenaud, D. K., D. Maclaurin, J. Iparraguirre, R. Bombarell, T. Hirzel, A. Aspuru-Guzik, and R. P. Adams. 2015. "Convolutional Networks on Graphs for Learning Molecular Fingerprints". In: *Advances in Neural Information Processing Systems 28*. 2224–2232.

Dwivedi, V. P., C. K. Joshi, T. Laurent, Y. Bengio, and X. Bresson. 2020. "Benchmarking Graph Neural Networks". *arXiv e-prints*. arXiv: 2003.00982.

Expert, P., L.-D. Lord, M. L. Kringelbach, and G. Petri. 2019. "Editorial: Topological Neuroscience". *Network Neuroscience*. 3(3): 653–655.

Feragen, A., N. Kasenburg, J. Petersen, M. de Bruijne, and K. Borgwardt. 2013. "Scalable Kernels for Graphs with Continuous Attributes". In: *Advances in Neural Information Processing Systems 26*. Ed. by C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger. 216–224.

Floyd, R. W. 1962. "Algorithm 97: Shortest Path". *Communications of the ACM*. 5(6): 345.

Fröhlich, H., J. K. Wegner, F. Sieker, and A. Zell. 2005. "Optimal Assignment Kernels for Attributed Molecular Graphs". In: *Proceedings of the 22nd International Conference on Machine Learning*. ACM. New York, NY, USA. 225–232.

Gabriel, K. R. and R. R. Sokal. 1969. "A New Statistical Approach to Geographic Variation Analysis". *Systematic Biology*. 18(3): 259–278.

Gärtner, T., P. Flach, and S. Wrobel. 2003. "On Graph Kernels: Hardness Results and Efficient Alternatives". In: *Learning Theory and Kernel Machines*. Ed. by B. Schölkopf and M. K. Warmuth. Heidelberg, Germany: Springer. 129–143.

Gaüzère, B., L. Brun, and D. Villemin. 2012. "Two New Graphs Kernels in Chemoinformatics". *Pattern Recognition Letters*. 33(15): 2038–2047. Graph-Based Representations in Pattern Recognition. ISSN: 0167-8655.

Gaüzère, B., P.-A. Grenier, L. Brun, and D. Villemin. 2015. "Treelet Kernel Incorporating Cyclic, Stereo and Inter-Pattern Information in Chemoinformatics". *Pattern Recognition*. 48(2): 356–367. ISSN: 0031-3203.

Gilmer, J., S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl. 2017. "Neural Message Passing for Quantum Chemistry". In: *Proceedings of the 34th International Conference on Machine Learning*. Ed. by D. Precup and Y. W. Teh. *Proceedings of Machine Learning Research*. PMLR. 1263–1272.

Gkirtzou, K., J. Honorio, D. Samaras, R. Goldstein, and M. Blaschko. 2013. "fMRI Analysis with Sparse Weisfeiler–Lehman Graph Statistics". In: *Machine Learning in Medical Imaging*. Ed. by G. Wu, D. Zhang, D. Shen, P. Yan, K. Suzuki, and F. Wang. Cham, Switzerland: Springer. 90–97.

Glem, R., A. Bender, C. Hasselgren, L. Carlsson, S. Boyer, and J. Smith. 2006. "Circular Fingerprints: Flexible Molecular Descriptors with Applications from Physical Chemistry to ADME". *IDrugs : The Investigational Drugs Journal*. 9: 199–204.

Goh, K.-I., M. E. Cusick, D. Valle, B. Childs, M. Vidal, and A.-L. Barabási. 2007. "The Human Disease Network". *Proceedings of the National Academy of Sciences*. 104(21): 8685–8690.

Graf, A. B. A. and S. Borer. 2001. "Normalization in Support Vector Machines". In: *Pattern Recognition*. Ed. by B. Radig and S. Florczyk. Heidelberg, Germany: Springer. 277–282. ISBN: 978-3-540-45404-5.

Greene, D. and P. Cunningham. 2006. "Practical Solutions to the Problem of Diagonal Dominance in Kernel Document Clustering". In: *Proceedings of the 23rd International Conference on Machine Learning (ICML)*. New York, NY, USA: Association for Computing Machinery. 377–384. DOI: 10.1145/1143844.1143892.

Hamilton, W. L., R. Ying, and J. Leskovec. 2017a. "Inductive Representation Learning on Large Graphs". In: *Advances in Neural Information Processing Systems 30*. 1024–1034.

Hamilton, W. L., R. Ying, and J. Leskovec. 2017b. "Representation Learning on Graphs: Methods and Applications". *arXiv e-prints*. arXiv: 1709.05584 [cs.SI].

Haussler, D. 1999. "Convolution Kernels on Discrete Structures". *Tech. rep.* University of California at Santa Cruz.

He, Y. and A. Evans. 2010. "Graph Theoretical Modeling of Brain Connectivity". *Current Opinion in Neurology*. 23(4): 341–350.

Hido, S. and H. Kashima. 2009. "A Linear-Time Graph Kernel". In: *Proceedings of the Ninth IEEE International Conference on Data Mining*. Washington, DC, USA: IEEE Computer Society. 179–188.

Hoerl, A. E. and R. W. Kennard. 1970. "Ridge Regression: Biased Estimation for Nonorthogonal Problems". *Technometrics*. 12(1): 55–67.

Hofer, C., F. Graf, B. Rieck, M. Niethammer, and R. Kwitt. 2020. "Graph Filtration Learning". In: *Proceedings of the 37th International Conference on Machine Learning*. Ed. by H. D.-l. III and A. Singh. Vol. 119. *Proceedings of Machine Learning Research*. PMLR. 4314–4323.

Hofer, C., R. Kwitt, M. Niethammer, and A. Uhl. 2017. "Deep Learning with Topological Signatures". In: *Advances in Neural Information Processing Systems 30 (NeurIPS)*. Ed. by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett. Red Hook, NY, USA: Curran Associates, Inc. 1634–1644.

Hofmann, T., B. Schölkopf, and A. J. Smola. 2008. "Kernel Methods in Machine Learning". *The Annals of Statistics*. 36(3): 1171–1220.

Horváth, T., T. Gärtner, and S. Wrobel. 2004. "Cyclic Pattern Kernels for Predictive Graph Mining". In: *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: Association for Computing Machinery. 158–167.

Hotelling, H. 1933. "Analysis of a Complex of Statistical Variables into Principal Components". *Journal of Educational Psychology*. 24(6): 417.

Hsieh, C.-J., S. Si, and I. S. Dhillon. 2014. "Fast Prediction for Large-Scale Kernel Machines". In: *Advances in Neural Information Processing Systems 27*. Ed. by Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger. Curran Associates, Inc. 3689–3697.

Hu, W., M. Fey, M. Zitnik, Y. Dong, H. Ren, B. Liu, M. Catasta, and J. Leskovec. 2020. "Open Graph Benchmark: Datasets for Machine Learning on Graphs". arXiv: 2005.00687.

Huson, D. H. and D. Bryant. 2005. "Application of Phylogenetic Networks in Evolutionary Studies". *Molecular Biology and Evolution*. 23(2): 254–267.

Jaromczyk, J. W. and G. T. Toussaint. 1992. "Relative Neighborhood Graphs and Their Relatives". *Proceedings of the IEEE*. 80(9): 1502–1517.

Jebara, T. and R. Kondor. 2003. "Bhattacharyya and Expected Likelihood Kernels". In: *Learning Theory and Kernel Machines*. Ed. by B. Schölkopf and M. K. Warmuth. Heidelberg, Germany: Springer. 57–71.

Jebara, T., R. Kondor, and A. Howard. 2004. "Probability Product Kernels". *Journal of Machine Learning Research*. 5: 819–844.

Kanehisa, M. and S. Goto. 2000. "KEGG: Kyoto Encyclopedia of Genes and Genomes". *Nucleic Acids Research*. 28(1): 27–30.

Kanungo, T., D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu. 2002. "An Efficient k-Means Clustering Algorithm: Analysis and Implementation". *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 24(7): 881–892.

Karlebach, G. and R. Shamir. 2008. "Modelling and Analysis of Gene Regulatory Networks". *Nature Reviews Molecular Cell Biology*. 9(10): 770.

Karp, R. M. 1972. "Reducibility among Combinatorial Problems". In: *Complexity of Computer Computations*. Ed. by R. E. Miller, J. W. Thatcher, and J. D. Bohlinger. Boston, MA, USA: Springer. 85–103.

Kashima, H., K. Tsuda, and A. Inokuchi. 2003. "Marginalized Kernels between Labeled Graphs". In: *Proceedings of the 20th International Conference on Machine Learning*. 321–328.

Kataoka, T. and A. Inokuchi. 2016. "Hadamard Code Graph Kernels for Classifying Graphs". In: *Proceedings of the 5th International Conference on Pattern Recognition Applications and Methods (ICPRAM)*. 24–32.

Kersting, K., N. M. Kriege, C. Morris, P. Mutzel, and M. Neumann. 2016. "Benchmark Data Sets for Graph Kernels". URL: http://graphkernels.cs.tu-dortmund.de.

Kipf, T. N. and M. Welling. 2017. "Semi-Supervised Classification with Graph Convolutional Networks". In: *International Conference on Learning Representations (ICLR)*.

Köbler, J. and O. Verbitsky. 2008. "From Invariants to Canonization in Parallel". In: *Computer Science – Theory and Applications*. Ed. by E. A. Hirsch, A. A. Razborov, A. Semenov, and A. Slissenko. Heidelberg, Germany: Springer. 216–227. ISBN: 978-3-540-79709-8.

Koller, D. and N. Friedman. 2009. *Probabilistic Graphical Models: Principles and Techniques*. Cambridge, MA, USA: MIT press.

Kondor, R. I. and J. Lafferty. 2002. "Diffusion Kernels on Graphs and Other Discrete Structures". In: *Proceedings of the 19th International Conference on Machine Learning*. Vol. 2002. 315–322.

Kondor, R. and H. Pan. 2016. "The Multiscale Laplacian Graph Kernel". In: *Advances in Neural Information Processing Systems 29*. Ed. by D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett. 2990–2998.

Kondor, R., N. Shervashidze, and K. M. Borgwardt. 2009. "The Graphlet Spectrum". In: *Proceedings of the 26th Annual International Conference on Machine Learning*. ACM. New York, NY, USA. 529–536.

Kriege, N. M., P.-L. Giscard, and R. C. Wilson. 2016. "On Valid Optimal Assignment Kernels and Applications to Graph Classification". In: *Advances in Neural Processing Systems 29*. 1623–1631.

Kriege, N. M., F. D. Johansson, and C. Morris. 2020. "A Survey on Graph Kernels". *Applied Network Science*. 5(1): 6.

Kriege, N. M., M. Neumann, C. Morris, K. Kersting, and P. Mutzel. 2019. "A Unifying View of Explicit and Implicit Feature Maps of Graph Kernels". *Data Mining and Knowledge Discovery*. 33(6): 1505–1547.

Kriege, N. and P. Mutzel. 2012. "Subgraph Matching Kernels for Attributed Graphs". *Proceedings of the 29th International Conference on Machine Learning*: 291–298.

Ktena, S. I., S. Parisot, E. Ferrante, M. Rajchl, M. Lee, B. Glocker, and D. Rueckert. 2018. "Metric Learning with Spectral Graph Convolutions on Brain Connectivity Networks". *NeuroImage*. 169: 431–442.

Lanckriet, G. R. G., N. Cristianini, P. Bartlett, L. E. Ghaoui, and M. I. Jordan. 2004. "Learning the Kernel Matrix with Semidefinite Programming". *Journal of Machine Learning Research*. 5: 27–72.

Lee, J. A. and M. Verleysen. 2012. "Graph-Based Dimensionality Reduction". In: *Image Processing and Analysis with Graphs: Theory and Practice*. 351–382.

Leskovec, J., L. A. Adamic, and B. A. Huberman. 2007. "The Dynamics of Viral Marketing". *ACM Transactions on the Web*. 1(1): 5:1–5:39.

Levi, G. 1973. "A Note on the Derivation of Maximal Common Subgraphs of Two Directed or Undirected Graphs". *Calcolo*. 9(4): 341.

Lloyd, S. 1982. "Least Squares Quantization in PCM". *IEEE Transactions on Information Theory*. 28(2): 129–137.

Lonsdale, J., J. Thomas, M. Salvatore, R. Phillips, E. Lo, S. Shad, R. Hasz, G. Walters, F. Garcia, N. Young, *et al.* 2013. "The Genotype-Tissue Expression (GTEx) project". *Nature Genetics*. 45(6): 580–585.

Mahé, P., N. Ueda, T. Akutsu, J.-L. Perret, and J.-P. Vert. 2004. "Extensions of Marginalized Graph Kernels". In: *Proceedings of the 21st Conference on Machine learning*. New York, NY, USA: ACM. 70–78.

Mahé, P. and J.-P. Vert. 2009. "Graph Kernels Based on Tree Patterns for Molecules". *Machine Learning*. 75(1): 3–35.

Menchetti, S., F. Costa, and P. Frasconi. 2005. "Weighted Decomposition Kernels". In: *Proceedings of the 22nd International Conference on Machine Learning*. New York, NY, USA: ACM. 585–592.

Mikolov, T., K. Chen, G. Corrado, and J. Dean. 2013. "Efficient Estimation of Word Representations in Vector Space". *arXiv e-prints*. arXiv: 1301.3781 [cs.CL].

Morris, C., N. M. Kriege, F. Bause, K. Kersting, P. Mutzel, and M. Neumann. 2020. "TUDataset: A Collection of Benchmark Datasets for Learning with Graphs". In: *ICML 2020 Workshop on Graph Representation Learning and Beyond (GRL+ 2020)*. URL: www.graphlearning.io.

Morris, C., N. M. Kriege, K. Kersting, and P. Mutzel. 2016. "Faster Kernels for Graphs with Continuous Attributes via Hashing". In: *Proceedings of the 16th IEEE International Conference on Data Mining*. 1095–1100.

Müllner, D. 2011. "Modern Hierarchical, Agglomerative Clustering Algorithms". *arXiv e-prints*. arXiv: 1109.2378 [stat.ML].

Neumann, M., R. Garnett, C. Bauckhage, and K. Kersting. 2016. "Propagation Kernels: Efficient Graph Kernels from Propagated Information". *Machine Learning*. 102(2): 209–245.

Nikolentzos, G., P. Meladianos, S. Limnios, and M. Vazirgiannis. 2018a. "A Degeneracy Framework for Graph Similarity". In: *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*. International Joint Conferences on Artificial Intelligence Organization. 2595–2601.

Nikolentzos, G., P. Meladianos, A. J.-P. Tixier, K. Skianis, and M. Vazirgiannis. 2018b. "Kernel Graph Convolutional Neural Networks". In: *Artificial Neural Networks and Machine Learning – ICANN 2018*. Ed. by V. Kůrková, Y. Manolopoulos, B. Hammer, L. Iliadis, and I. Maglogiannis. Cham, Switzerland: Springer. 22–32.

Nikolentzos, G., G. Siglidis, and M. Vazirgiannis. 2019. "Graph Kernels: A Survey". arXiv: 1904.12218 [stat.ML].

Nikolentzos, G. and M. Vazirgiannis. 2018. "Message Passing Graph Kernels". *arXiv e-prints*. arXiv: 1808.02510 [stat.ML].

Nyström, E. J. 1930. "Über die praktische Auflösung von Integralgleichungen mit Anwendungen auf Randwertaufgaben". *Acta Mathematica*. 54: 185–204.

Ó Searcóid, M. 2007. *Metric Spaces*. London, England: Springer.

Orsini, F., P. Frasconi, and L. De Raedt. 2015. "Graph Invariant Kernels". In: *Proceedings of the 24th International Conference on Artificial Intelligence*. Palo Alto, CA, USA: AAAI Press. 3756–3762.

Paulevé, L., H. Jégou, and L. Amsaleg. 2010. "Locality Sensitive Hashing: A Comparison of Hash Function Types and Querying Mechanisms". *Pattern Recognition Letters*. 31(11): 1348–1358.

Pearson, K. 1901. "LIII. On Lines and Planes of Closest Fit to Systems of Points in Space". *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*. 2(11): 559–572.

Pržulj, N., D. G. Corneil, and I. Jurisica. 2004. "Modeling Interactome: Scale-free or Geometric?" *Bioinformatics*. 20(18): 3508–3515.

Pržulj, N., D. G. Corneil, and I. Jurisica. 2006. "Efficient Estimation of Graphlet Frequency Distributions in Protein–Protein Interaction Networks". *Bioinformatics*. 22(8): 974–980.

Ralaivola, L., S. J. Swamidass, H. Saigo, and P. Baldi. 2005. "Graph Kernels for Chemical Informatics". *Neural Networks*. 18(8): 1093–1110.

Ramon, J. and T. Gärtner. 2003. "Expressivity Versus Efficiency of Graph Kernels". In: *Proceedings of the 1st International Workshop on Mining Graphs, Trees and Sequences*. 65–74.

Read, R. C. and D. G. Corneil. 1977. "The Graph Isomorphism Disease". *Journal of Graph Theory*. 1(4): 339–363.

Rieck, B., C. Bock, and K. Borgwardt. 2019. "A Persistent Weisfeiler–Lehman Procedure for Graph Classification". In: *Proceedings of the 36th International Conference on Machine Learning*. Ed. by K. Chaudhuri and R. Salakhutdinov. Vol. 97. *Proceedings of Machine Learning Research*. PMLR. 5448–5458.

Riesen, K. 2015. *Structural Pattern Recognition with Graph Edit Distance*. Cham, Switzerland: Springer.

Rotmensch, M., Y. Halpern, A. Tlimat, S. Horng, and D. Sontag. 2017. "Learning a Health Knowledge Graph from Electronic Medical Records". *Scientific Reports*. 7(1): 5994.

Schölkopf, B. and A. J. Smola. 2002. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press.

Schütt, K. T., F. Arbabzadah, S. Chmiela, K. R. Müller, and A. Tkatchenko. 2017. "Quantum-Chemical Insights from Deep Tensor Neural Networks". *Nature Communications*. 8.

Scott, J. 2011. "Social Network Analysis: Developments, Advances, and Prospects". *Social Network Analysis and Mining*. 1(1): 21–26.

Shawe-Taylor, J. and N. Cristianini. 2004. *Kernel Methods for Pattern Analysis*. Cambridge, United Kingdom: Cambridge University Press.

Shervashidze, N., P. Schweitzer, E. J. van Leeuwen, K. Mehlhorn, and K. M. Borgwardt. 2011. "Weisfeiler–Lehman Graph Kernels". *Journal of Machine Learning Research*. (12): 2539–2561.

Shervashidze, N. and K. Borgwardt. 2009. "Fast Subtree Kernels on Graphs". In: *Advances in Neural Information Processing Systems 22*. 1660–1668.

Shervashidze, N., S. Vishwanathan, T. Petri, K. Mehlhorn, and K. Borgwardt. 2009. "Efficient Graphlet Kernels for Large Graph Comparison". In: *Proceedings of the 12th International Conference on Artificial Intelligence and Statistics*. 488–495.

Siglidis, G., G. Nikolentzos, S. Limnios, C. Giatsidis, K. Skianis, and M. Vazirgiannis. 2020. "GraKeL: A Graph Kernel Library in Python". *Journal of Machine Learning Research*. 21(54): 1–5.

Silva, V. de and G. Carlsson. 2004. "Topological Estimation Using Witness Complexes". In: *Symposium on Point-Based Graphics*. Ed. by M. Gross, H. Pfister, M. Alexa, and S. Rusinkiewicz. The Eurographics Association. 157–166.

Smola, A. J. and I. Kondor. 2003. "Kernels and Regularization on Graphs". In: *Learning Theory and Kernel Machines*. Ed. by B. Schölkopf and M. K. Warmuth. Heidelberg, Germany: Springer. 144–158.

Sonnenburg, S., G. Rätsch, C. Schäfer, and B. Schölkopf. 2006. "Large Scale Multiple Kernel Learning". *Journal of Machine Learning Research*. 7: 1531–1565.

Sugiyama, M. and K. Borgwardt. 2015. "Halting in Random Walk Kernels". In: *Advances in Neural Information Processing Systems 28*. Ed. by C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett. 1639–1647.

Sugiyama, M., M. E. Ghisu, F. Llinares-López, and K. Borgwardt. 2017. "`graphkernels`: R and Python Packages for Graph Comparison". *Bioinformatics*. 34(3): 530–532.

Szklarczyk, D., A. L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, M. Simonovic, N. T. Doncheva, J. H. Morris, P. Bork, *et al.* 2018. "STRING v11: Protein–Protein Association Networks with Increased Coverage, Supporting Functional Discovery in Genome-Wide Experimental Datasets". *Nucleic Acids Research*. 47(D1): D607-bibrangedash D613.

Tan, P.-N., M. Steinbach, A. Karpatne, and V. Kumar. 2019. *Introduction to Data Mining*. 2nd ed. London, United Kingdom: Pearson.

Todeschini, R. and V. Consonni. 2008. *Handbook of Molecular Descriptors*. Vol. 11. Weinheim, Germany: Wiley-VCH.

Togninalli, M., E. Ghisu, F. Llinares-López, B. Rieck, and K. Borgwardt. 2019. "Wasserstein Weisfeiler–Lehman Graph Kernels". In: *Advances in Neural Information Processing Systems 32 (NeurIPS)*. Ed. by H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett. Curran Associates, Inc. 6436–6446.

Trinajstic, N. 2018. *Chemical Graph Theory*. 2nd ed. Boca Raton, FL, USA: CRC Press.

Tschiatschek, S., A. Singla, M. Gomez Rodriguez, A. Merchant, and A. Krause. 2018. "Fake News Detection in Social Networks via Crowd Signals". In: *Companion Proceedings of the The Web Conference 2018*. 517–524.

Vamathevan, J., D. Clark, P. Czodrowski, I. Dunham, E. Ferran, G. Lee, B. Li, A. Madabhushi, P. Shah, M. Spitzer, *et al.* 2019. "Applications of Machine Learning in Drug Discovery and Development". *Nature Reviews Drug Discovery*. 18: 463–477.

Vega-Pons, S. and P. Avesani. 2013. "Brain Decoding via Graph Kernels". In: *Proceedings of the 2013 International Workshop on Pattern Recognition in Neuroimaging. PRNI '13*. IEEE Computer Society. 136–139.

Vert, J.-P. 2008. "The Optimal Assignment Kernel Is Not Positive Definite". *arXiv e-prints*. arXiv: 0801.4061 [cs.LG].

Vert, J.-P., K. Tsuda, and B. Schölkopf. 2004. "A Primer on Kernel Methods". In: *Kernel Methods in Computational Biology*. Ed. by B. Schölkopf, K. Tsuda, and J.-P. Vert. Cambridge, MA, USA: MIT Press. 35–70.

Villani, C. 2008. *Optimal Transport: Old and New*. Heidelberg, Germany: Springer.

Vishwanathan, S. V. N., K. M. Borgwardt, and N. N. Schraudolph. 2006. "Fast Computation of Graph Kernels". In: *Advances in Neural Information Processing Systems 19*. 1449–1456.

Vishwanathan, S., N. N. Schraudolph, R. Kondor, and K. M. Borgwardt. 2010. "Graph Kernels". *Journal of Machine Learning Research*. 11(40): 1201–1242.

Wang, J., X. Zuo, and Y. He. 2010. "Graph-Based Network Analysis of Resting-State Functional MRI". *Frontiers in Systems Neuroscience*. 4: 16:1–16:14.

Warshall, S. 1962. "A Theorem on Boolean Matrices". *Journal of the ACM*. 9(1): 11–12.

Wasserman, S. and K. Faust. 1994. *Social Network Analysis: Methods and Applications*. Cambridge, UK: Cambridge University Press.

Weisfeiler, B. and A. A. Lehman. 1968. "A Reduction of a Graph to a Canonical Form and an Algebra Arising During This Reduction". *Nauchno-Technicheskaya Informatsia*. 2(9): 12–16.

Wiener, H. 1947. "Structural Determination of Paraffin Boiling Points". *Journal of the American Chemical Society*. 69(1): 17–20. ISSN: 0002-7863.

Williams, C. K. I. and M. Seeger. 2001. "Using the Nyström Method to Speed Up Kernel Machines". In: *Advances in Neural Information Processing Systems 13*. Ed. by T. K. Leen, T. G. Dietterich, and V. Tresp. MIT Press. 682–688.

Wilson, R. C. and P. Zhu. 2008. "A Study of Graph Spectra for Comparing Graphs and Trees". *Pattern Recognition*. 41(9): 2833–2841.

Wu, Z., B. Ramsundar, E. N. Feinberg, J. Gomes, C. Geniesse, A. S. Pappu, K. Leswing, and V. Pande. 2018. "MoleculeNet: a Benchmark for Molecular Machine Learning". *Chemical Science*. 9(2): 513–530.

Wu, Z., S. Pan, F. Chen, G. Long, C. Zhang, and S. Y. Philip. 2020. "A Comprehensive Survey on Graph Neural Networks". *IEEE Transactions on Neural Networks and Learning Systems*.

Xu, K., W. Hu, J. Leskovec, and S. Jegelka. 2019. "How Powerful are Graph Neural Networks?" In: *International Conference on Learning Representations (ICLR)*.

Yanardag, P. and S. Vishwanathan. 2015. "Deep Graph Kernels". In: *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: ACM. 1365–1374.

Yang, Y. 1999. "An Evaluation of Statistical Approaches to Text Categorization". *Information Retrieval*. 1(1): 69–90.

Ying, R., J. You, C. Morris, X. Ren, W. L. Hamilton, and J. Leskovec. 2018. "Hierarchical Graph Representation Learning with Differentiable Pooling". In: *Advances in Neural Information Processing Systems 32*. 4805–4815.

Zhang, B. and S. Horvath. 2005. "A General Framework for Weighted Gene Co-expression Network Analysis". *Statistical Applications in Genetics and Molecular Biology*. 4(1): 17:1–17:43.

Zhang, Z., M. Wang, Y. Xiang, Y. Huang, and A. Nehorai. 2018. "RetGK: Graph Kernels Based on Return Probabilities of Random Walks". In: *Advances in Neural Information Processing Systems 32*. 3964–3974.

Zhao, Q. and Y. Wang. 2019. "Learning Metrics for Persistence-Based Summaries and Applications for Graph Classification". In: *Advances in Neural Information Processing Systems 32 (NeurIPS)*. Ed. by H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett. Curran Associates, Inc. 9855–9866.

Zhou, J., G. Cui, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun. 2018. "Graph Neural Networks: A Review of Methods and Applications". *arXiv e-prints*. arXiv: 1812.08434 [cs.LG].