



HAL
open science

Learning from Diffusion-Weighted Magnetic Resonance Images using graph kernels

Sylvain Takerkart, Gottfried Berton, Nicole Malfait, François-Xavier Dupé

► **To cite this version:**

Sylvain Takerkart, Gottfried Berton, Nicole Malfait, François-Xavier Dupé. Learning from Diffusion-Weighted Magnetic Resonance Images using graph kernels. 11th IAP Workshop on Graph-based Representation, GbR2017, May 2017, Anacapri, Italy. hal-01484660

HAL Id: hal-01484660

<https://hal.science/hal-01484660v1>

Submitted on 7 Mar 2017

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

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

Learning from Diffusion-Weighted Magnetic Resonance Images using graph kernels

Sylvain Takerkart[†], Gottfried Berton[†], Nicole Malfait[†], and François-Xavier Dupé^{*}

[†]Aix Marseille Univ, CNRS, INT, Inst Neurosci Timone, Marseille, France

^{*} Aix Marseille Univ, CNRS, Centrale Marseille, LIF, Marseille, France
sylvain.takerkart@univ-amu.fr

Abstract. Diffusion-weighted magnetic resonance imaging (DWI) is a scanning procedure that allows inferring the anatomical connectivity of the brain non invasively. DWI can be used to segment the brain into a set of relevant sub-regions, yielding what is called a parcellation in the neuroimaging literature. In this paper, we introduce a generic framework that allows building predictive models using parcellations obtained on a single individual. It consists in constructing attributed region adjacency graphs to represent the parcellations and using suitable graph kernels to exploit the versatility of kernel methods. We demonstrate the relevance of this framework on real data, by showing that we can predict the age range of an individual from the connectivity structure of its corpus callosum, the main hub of connections between the left and right hemispheres of the brain. Furthermore, we study the behavior of different graph kernels for this task. This work opens new opportunities to identify DWI-based biomarkers of neurodegenerative and psychiatric diseases.

Keywords: region-adjacency graphs, graph kernel, magnetic resonance imaging, brain connectivity, brain parcellation

1 Introduction

In recent years, the use of neuroimaging-based predictive models has seen a fast development. In most cases, these machine learning models are designed in order to build diagnosis or prognosis tools to help clinicians deal with neurological or psychiatric disorders [23]. But they also prove valuable in the field of basic research aiming at a better understanding of the organization of the brain [26]. Amongst the different neuroimaging modalities used in this context, diffusion-weighted magnetic resonance imaging (DWI) is under-exploited, with only a limited number of published studies (see for instance [16, 7, 17, 9, 11]). Yet it is attracting a growing interest.

DWI is a magnetic resonance imaging procedure that uses the diffusion of water molecules to reveal the micro-architecture of a physiological tissue. The white matter of the brain (WM) is of particular interest for neuroscientists and can be studied using DWI. Indeed, the WM mostly contains myelinated axons,

the neuronal fibers, which are organized into large groups of axons called *bundles*. The structure of these bundles alters the free motion of the water molecules, and the resulting anisotropy can be captured using DWI. In the nervous system, the role of the axons is to transmit information between neurons, connecting different brain regions. Thanks to its ability to characterize the axon bundles of the WM, DWI has developed into a major tool to study *brain connectivity*. Furthermore, because a large number of brain disorders involve abnormal connectivity [6], DWI has allowed to gain unprecedented insights about these pathologies. It is therefore crucial to further design new DWI-based predictive models in order to identify non-invasive biomarkers of connectivity-based neurological syndromes.

One of the major questions to be addressed when designing predictive models based on DWI data is the representation of the data itself. A commonly used high-level summary representation of the diffusion data consists in segmenting the domain of interest – either the full brain or a given region of interest – into homogeneous sub-regions, often denoted as *parcels* in the neuroimaging community. The literature focuses on estimating parcellation models that are valid for a population, for two main reasons: first, because one of the main aims of neuroscience is to find invariants across individuals; secondly, because it is not possible to produce inference at the population level using parcellations estimated on data from a single individual using the statistical tools that are most commonly used in neuroimaging. However, because such parcellations simply are the results of a segmentation, the field of pattern recognition offers a wide range of methods which could help overcome this challenge. Our objective in this paper is therefore to demonstrate that pattern recognition tools can make it possible to build predictive models based on DWI-based parcellations estimated on a single individual, in order to make inference at the population level.

The framework we propose starts by the construction of attributed graphs from such individual parcellations and then relies on kernel methods [21], using appropriate graph kernels, to build a predictive model. We demonstrate the effectiveness of this framework on a real DWI dataset, by showing, for the first time, that we can estimate the age category of a subject from the connectivity structure of his/her corpus callosum, the main hub of connections between the two hemispheres of the brain. Furthermore, we study the influence of the choice of the kernel on the performance of the model, by comparing the expressivity of several classical graph kernels. In Section 2, we describe the processing workflow necessary to obtain a parcellation from the raw DWI data, the processes used to construct attributed graphs and the different graph kernels. Then, in Section 3, we provide a full description of the data itself, the neuroscientific question addressed in our experiments and the obtained results. Finally, we discuss these results and future research directions in Section 4.

2 Methods

2.1 Constructing DWI-based parcellations

The processing of DWI data involves a complex analysis workflow to obtain a parcellation from the raw data, which is a set of three-dimensional volumes that each provides, at every brain voxel, a measurement of the amount of water diffusion in a given direction – the set of directions being a priori chosen to isotropically span the entire sphere. After correction of the image distortions and compensation of the between-volume displacements, we first estimate the fibre orientation distribution at each voxel. We use the so-called *ball-and-stick* model, which assumes that the orientation diffusion function includes an isotropic component – the ball – and one or several directions that follows the neuronal fibers passing through this voxel – the stick(s). This model is fit at each voxel using a sampling-based Bayesian technique [2]. Secondly, a probabilistic *tractography* algorithm allows to estimate the probability of connection $p(m, n)$ between two brain voxels / regions m and n [2]. For the domain \mathcal{D} to be parcellated and a set of target brain regions \mathcal{T} , we can thus compute the *connectivity matrix* $C = [p(m, n)]_{m \in \mathcal{T}, n \in \mathcal{D}} \in \mathbb{R}^{M \times N}$, where M and N are respectively the number of regions in \mathcal{T} and the number of voxels in \mathcal{D} . Note that the vector $[p(m, n)]_{m \in \mathcal{T}}$ is called the tractogram of n . The segmentation of \mathcal{D} into parcels is obtained using a clustering of the columns of C . Here, we use Ward’s hierarchical clustering, with an added spatial constraint, to identify contiguous clusters of voxels that have similar tractograms. Each of these clusters forms a parcel v_i , i.e a sub-region with a homogeneous connectivity structure, as estimated using DWI data. This workflow is illustrated on Fig. 1.

2.2 Graph construction

Given a parcellation V of the domain of interest \mathcal{D} , i.e a set of R sub-regions / parcels $\{v_i\}_{i=1}^R$ that fully cover \mathcal{D} and do not intersect, we would like to obtain an adequate representation of the connectivity structure within \mathcal{D} . We use region adjacency graphs to encode the topographical character of the information contained in the parcellation. We therefore define a node of the graph at each parcel $v_i \in V$. The set of edges $E \subset V \times V$ is given by the spatial adjacency of the parcels, i.e $e = v_i v_j \in E \iff$ parcels v_i and v_j are spatially adjacent. We then add two vectorial attributes to each graph node: i) the average tractogram of all voxels of the parcel, and ii) the coordinates of the barycenter of the parcel, in a coordinate system that is comparable across participants. We note $\varphi_1 : V \rightarrow \mathbb{R}^M$ and $\varphi_2 : V \rightarrow \mathbb{R}^3$ the functions respectively giving the tractogram and coordinate attributes of a node, and $\varphi : V \rightarrow \mathbb{R}^M \times \mathbb{R}^3$ so that $\varphi(v) = (\varphi_1(v), \varphi_2(v))$. The connectivity structure of a given brain region is therefore represented by the attributed graph defined as $G = (V, E, \varphi)$.

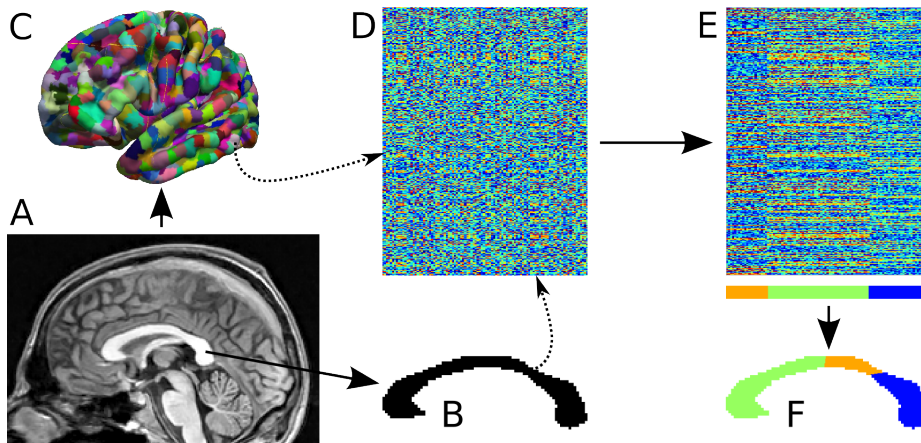


Fig. 1. Processing workflow of the DWI data. A: a structural MRI depicting the shape of the brain. B: extraction of the domain \mathcal{D} to be parcellated (here, the corpus callosum). C: definition of the set of target regions \mathcal{T} for the tractography (here, a set of anatomically defined regions of the cortex). D: the connectivity matrix C contains the probability of connection of any given voxel of \mathcal{D} to any of the target regions. E: the re-ordered connectivity matrix after clustering of the tractograms / columns. F: the parcellation V , where voxels are colored with the corresponding cluster label.

2.3 Learning from graphs

Learning from graphs is a difficult task and graph kernels, which provide an indirect projection of a graph onto a Hilbert space, are popular for that matter because they render accessible a vast array of machine learning methods. Several strategies can be employed to design graph kernels, such as instantiating an R-convolution kernel [8] (see e.g the random walk kernel [10, 27], kernels over sets of paths [24], trees [13] or graphlets [22]), exploiting spectral graph decompositions [28] or the concept of graph edit distance [18]. While most recent works focus on improving kernel scalability for unlabeled or weakly-labeled graphs (see e.g [22, 12, 5, 1]), we here need kernels that can easily accommodate vector-valued attributes. We describe below the kernels that we will use for our experiments.

Walk-based R-Convolution graph kernels. Given a graph G , we denote V its set of vertices, $E \subset V \times V$ its set of edges and $\varphi : V \rightarrow \mathbb{R}^L$ the vector-valued function giving the attributes of the nodes. A walk w is defined as a sequence of adjacent nodes in the graph, and we suppose that a positive definite walk kernel, denoted K_{walk} , is available.

The first kernel we will use is the classical random walk graph kernel [10]. For two graphs G_1 and G_2 , it is defined as

$$K_{random}(G_1, G_2) = \sum_{w_1 \in G_1} \sum_{w_2 \in G_2} K_{walk}(w_1, w_2) p(w_1|G_1) p(w_2|G_2), \quad (1)$$

where $p(w|G)$ is the probability of walk w in G . It compares the density of walks between graphs, thus taking into account both local and global information.

Our second kernel is the bag of paths kernel [24, 4] – paths are walks with unique nodes. Let P_1 and P_2 be bags of paths respectively associated with G_1 and G_2 . We denote by $|\cdot|$ the number of paths inside a bag. The mean bag of paths kernel is constructed by averaging the walk kernel over all couples of paths from each bag:

$$K_{mean}(P_1, P_2) = \frac{1}{|P_1|} \frac{1}{|P_2|} \sum_{h \in P_1} \sum_{h' \in P_2} K_{walk}(h, h'). \quad (2)$$

In practice, we use bags of paths of constant size, i.e $P_{\mathcal{P}}(G) = \{w \in G \mid |w| \in \mathcal{P}\}$.

At the core of these graph kernels, the walk kernel K_{walk} measures the similarity between two walks. Clearly most of the expressivity of the whole kernel relies on how we compare walks with K_{walk} . Here, we use the classical kernel proposed by [10], where we only compare walks of the same size by making a direct alignment of both nodes and edges. The considered walk kernel writes as follow for two walks $h = v_1 v_2 \dots v_{|h|}$ and $h' = v'_1 v'_2 \dots v'_{|h'|}$,

$$K_{walk}(h, h') = \begin{cases} \prod_{i=1}^{|h|} K_{node}(\varphi(v_i), \varphi(v'_i)) & \text{if } |h| = |h'| \\ 0 & \text{otherwise} \end{cases}, \quad (3)$$

where K_{node} is a kernel on the attributes of the nodes, usually defined with a combination of Gaussian kernels. Given our definition $\varphi = (\varphi_1, \varphi_2)$ given in 2.2, we use the following node kernel,

$$K_{node}(\varphi(v_1), \varphi(v_2)) = \exp\left(-\frac{\|\varphi_1(v_1) - \varphi_1(v_2)\|^2}{2\sigma_1^2}\right) \exp\left(-\frac{\|\varphi_2(v_1) - \varphi_2(v_2)\|^2}{2\sigma_2^2}\right), \quad (4)$$

where $\sigma_1 > 0$ and $\sigma_2 > 0$ are two hyper-parameters.

Graph edit distance kernels. Graph edit distances [14, 20] are an attractive way of comparing graphs as they provide a set of editions (e.g. node/edge substitution, suppression, addition...) with its cost. In [18], the authors proposed to build a graph kernel from such distances by applying them inside a Gaussian kernel. Let d_E be a graph edit distance, the corresponding kernel is defined by:

$$K_{edit}(G_1, G_2) = \exp\left(\frac{-d_E(G_1, G_2)^2}{2\sigma_E^2}\right), \quad (5)$$

where $\sigma_E > 0$ is the hyper-parameter of the kernel.

In order to avoid the computational burden associated with computing d_E , we use the approximation proposed in [20], which relies on the Munkres assignment algorithm and requires comparing nodes attributes. For this, given that $\varphi = (\varphi_1, \varphi_2)$, we use a combination of euclidean norms:

$$d_{node}(v_1, v_2) = \|\varphi_1(v_1) - \varphi_1(v_2)\| + \gamma \|\varphi_2(v_1) - \varphi_2(v_2)\|, \quad (6)$$

where γ is an equilibrium parameter.

3 Experiments and results

The objective of our experiments is to demonstrate that our framework makes it possible to perform predictions from an individual’s DWI-based parcellation. In the following, we present the addressed prediction task and its neuroscientific motivation, the dataset used in our experiments, the experimental procedure and the quantitative results that we obtained.

3.1 Aging trajectory of the corpus callosum

The brain is an organ that continuously evolves throughout the lifespan. In particular, numerous markers of aging can be identified using neuroimaging techniques, such as for instance the reduction of the global volume of grey matter in the brain. Numerous brain pathologies – such as Alzheimer’s disease – induce some alterations compared to the normal aging process. Establishing the aging trajectory of specific brain features in a healthy population can therefore allow to use a deviation from this normal trajectory as a potential marker of a disease, which opens possibilities to design diagnosis and/or prognosis tools. In a predictive framework, establishing an aging trajectory comes down to obtaining a model that can guess the age of the subject from a given set of brain features.

Amongst the particular parts of the brain that have been identified to display age-related changes, the corpus callosum (CC), the largest commissure of white matter connecting the two hemispheres of the brain, is of particular interest because its integrity is known to be altered in several neurodegenerative diseases, with for instance an abnormally low CC size [25]. However, the age-related changes of the spatial organisation of the CC structural connectivity have not been studied until now. It is well known that the connectivity structure of the CC is topographically organized: among others, the fibers that pass through its most anterior part – called the *genu* – project to the anterior part of the cortex – the frontal lobe, while fibers passing through the posterior CC – the *splenium* – project to the back of the brain – the occipital lobe. In the present paper, we use DWI data to investigate whether the spatial organization of the CC connectivity changes with age.

3.2 Data and experiments

Our data comes from the enhanced Nathan Kline Institute-Rockland Sample¹. A small sub-sample of the participants enrolled in this initiative have been scanned using structural and diffusion-weighted MRI, allowing us to work with data from 65 participants aged 36 to 77 year-old. We analyzed the structural MRI (MPRAGE sequence, voxel size: 1 mm, volume size: 256 × 256 × 176, acquisition time: 6 mn, see Fig. 1A) using the *freesurfer* software suite² to identify the CC (Fig. 1B) and define a set of 1000 target regions for the tractography (Fig. 1C).

¹ http://fcon_1000.projects.nitrc.org/indi/enhanced/

² <https://surfer.nmr.mgh.harvard.edu/>

The DWI data (EPI sequence, 137 directions, voxel size: 2 mm, volume size: $106 \times 90 \times 64$, acquisition time: 6 mn) was pre-processed with the FSL software suite³. The rest of the experiments were conducted using in-house software (see Fig. 2 for an illustration of the resulting parcellations and graphs).

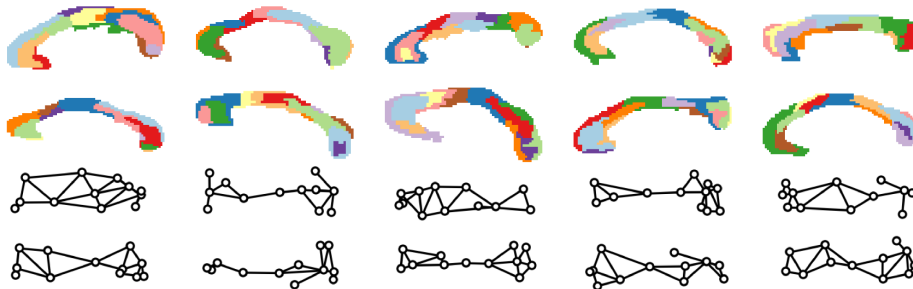


Fig. 2. Top: 10 examples of parcellations of the corpus callosum (with $R = 12$ parcels). Bottom: the resulting graphs, displayed in a local rectangular coordinate system.

We separated the 65 participants into two groups, the 32 oldest and the 33 youngest, to set up a binary classification problem. Our experiments therefore asked whether it is possible to predict the age group of an individual from the graph G of his/her CC, which would imply, if successful, that the spatial organization of the CC connectivity does indeed change with aging. We used a Support Vector Machine to perform this classification task, using five kernels amongst the ones described above: i) K_{random} (using $\sigma_1 = \sigma_2 = 0.5$, as with all the following walk-based kernels); ii) K_{mean} with $\mathcal{P} \in \{2\}$ (i.e walks with 2 nodes), that we will denote K_{mean_2} ; iii) K_{mean} with $\mathcal{P} \in \{3\}$, hereafter K_{mean_3} ; iv) K_{mean} with $\mathcal{P} \in \{2, 3\}$, hereafter K_{mean_23} ; v) K_{edit} (with $\gamma = 1000$ and $\sigma_E = 1$). We assessed the quality of the predictions with the mean classification accuracy obtained using a cross-validation scheme that included 500 randomly drawn balanced data splits, each with 56 and 9 participants respectively in the training and test set. Given the small size of the dataset, this procedure ensures obtaining an unbiased estimate of the mean accuracy.

We conducted two sets of experiments. First, in order to study the influence of the number of graph nodes R , we computed the classification accuracy when R is kept constant – taken in $\{4, 6, 8, \dots, 68, 70\}$ – for all splits of the cross-validation. Secondly, we performed a model order selection, choosing R within the same range, separately for each split, using an inner cross-validation within the training set. The selected value was used to fit the model on the full training set and test it on the left-out data.

³ <http://fsl.fmrib.ox.ac.uk>

3.3 Results

The results of the experiments are shown on Fig.3. All the classification results are above chance level (0.5), establishing the two main outcomes of the paper: from a methodological point of view, this demonstrates the validity of our framework and therefore the feasibility of using individual parcellations to build predictive models; from an applicative angle, this shows for the first time that the connectivity structure of the corpus callosum changes with aging. The classification scores are not very high (maximum: 0.68), which is common with neuroimaging data, for which the signal to noise ratio of the brain signatures of interest is usually very weak and the sample size very limited.

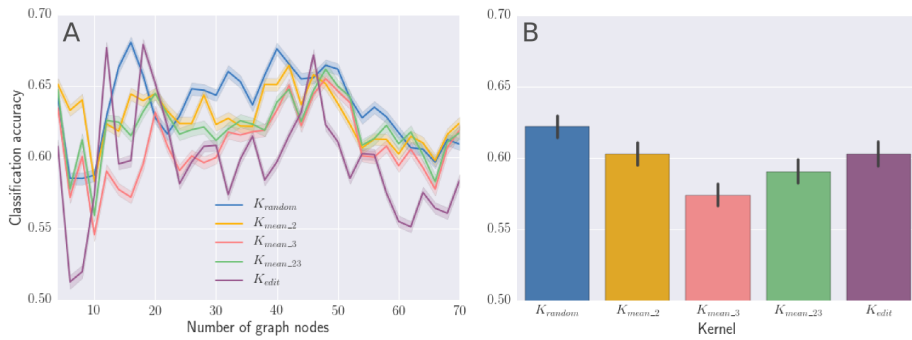


Fig. 3. Mean classification accuracy (\pm standard error) for five kernels. A: in function of the number R of graph nodes (kept constant for all folds of the cross-validation). Right: when R is selected through an inner cross-validation.

Fig.3A shows that the performance of the model strongly depends on the number R of graph nodes. Interestingly, the peak performances are not obtained for the same value of R for all kernels: the accuracy peaks for the random walk and edit distance kernels between $R = 12$ and 18 , while the mean kernels perform best around $R = 44$. Fig.3B shows that when performing model order selection, K_{random} outperforms the other kernels, followed by K_{mean_2} and K_{edit} .

4 Discussion and future work

In this paper, we have introduced a new framework that allows designing predictive models from individual DWI-based parcellations. To the best of our knowledge, this is the first of its kind.

The behavior of the different kernels we used is interesting since they are sensitive to various types of information. The fact that K_{mean_2} outperforms K_{mean_3} and K_{mean_23} indicates that most of the information lies at the local level, and more precisely at the level of pairs of adjacent nodes. Adding an extra

node in the paths seems to introduce noise rather than information, which could be expected from the noisy character of the parcellation process illustrated by the variability of the graphs on Fig. 2. But the higher performances of K_{random} , which can catch both local and global features, suggests that some extra information might lie at a more global scale. Also note that K_{edit} is more unstable than the other kernels, which can be caused by its non positive-definiteness [19] or by a failure of the Munkres-based assignment on our noisy parcellations.

This opens several directions for future work. First, the graph construction should benefit from concatenating different parcellations into multi-scale hierarchical graphs. Indeed this will provide a more robust representation of the data and render more explicit the combination of local and global information. Furthermore, because these graphs have a geometric embedding, the use of combinatorial maps and pyramids [3] could also be beneficial. Secondly, improvements should come from the graph kernel itself. We believe that the level of expressivity of the kernel on this data should increase by incorporating ideas from recent work such as [5] or using more structured base elements such as graphlets [22].

From an applicative perspective, we have demonstrated that the connectivity structure of the corpus callosum changes with age. However, additional work is needed to understand the nature of these modifications. Furthermore, in order to apply this framework on clinical data and identify biomarkers of a given neurological disorder, two main lines of work lie ahead of us. First, we will have to assess more finely the aging trajectory – of the corpus callosum or any other brain region – using a regression model that would offer a direct prediction of the age of the subject (similarly to [16]). Second, we will have to improve the performances far beyond the classification accuracies obtained here. This will require working with much larger datasets and therefore using more scalable kernels such as introduced in [15].

References

1. Bai, L., Rossi, L., Zhang, Z., Hancock, E.R.: An aligned subtree kernel for weighted graphs. In: ICML. pp. 30–39 (2015)
2. Behrens, T., Berg, H.J., Jbabdi, S., Rushworth, M., Woolrich, M.: Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? *NeuroImage* 34(1), 144–155 (Jan 2007)
3. Brun, L., Kropatsch, W.: Introduction to combinatorial pyramids. In: Digital and image geometry. pp. 108–128. Springer (2001)
4. Dupé, F.X., Brun, L.: Tree covering within a graph kernel framework for shape classification. In: ICIAP. pp. 278–287. Springer (2009)
5. Feragen, A., Kasenburg, N., Petersen, J., de Bruijne, M., Borgwardt, K.: Scalable kernels for graphs with continuous attributes. In: NIPS. pp. 216–224 (2013)
6. Fornito, A., Zalesky, A., Breakspear, M.: The connectomics of brain disorders. *Nature Reviews Neuroscience* 16(3), 159–172 (Feb 2015)
7. Ghanbari, Y., Smith, A.R., Schultz, R.T., Verma, R.: Identifying group discriminative and age regressive sub-networks from DTI-based connectivity via a unified framework of non-negative matrix factorization and graph embedding. *Medical Image Analysis* 18(8), 1337–1348 (Dec 2014)

8. Haussler, D.: Convolution kernels on discrete structures. Tech. Rep. UCSC-CRL-99-10, Department of Computer Science, Univ. of California at Santa Cruz (1999)
9. Kamiya, K., Amemiya, S., Suzuki, Y., Kunii, N., Kawai, K., Mori, H., Kunimatsu, A., Saito, N., Aoki, S., Ohtomo, K.: Machine Learning of DTI Structural Brain Connectomes for Lateralization of Temporal Lobe Epilepsy. *Magnetic Resonance in Medical Sciences* 15(1), 121–129 (2016)
10. Kashima, H., Tsuda, K., Inokuchi, A.: Marginalized kernel between labeled graphs. In: *ICML*. pp. 321–328 (2003)
11. Kawahara, J., Brown, C.J., Miller, S.P., Booth, B.G., Chau, V., Grunau, R.E., Zwicker, J.G., Hamarneh, G.: BrainNetCNN: Convolutional neural networks for brain networks; towards predicting neurodevelopment. *NeuroImage* (Sep 2016)
12. Kriege, N., Mutzel, P.: Subgraph matching kernels for attributed graphs. In: *ICML*. pp. 1015–1022 (2012)
13. Mahé, P., Vert, J.P.: Graph kernels based on tree patterns for molecules. *Machine learning* 75(1), 3–35 (2009)
14. Messmer, B.T., Bunke, H.: A new algorithm for error-tolerant subgraph isomorphism detection. *IEEE PAMI* 20(5), 493–504 (1998)
15. Morris, C., Kriege, N.M., Kersting, K., Mutzel, P.: Faster Kernels for Graphs with Continuous Attributes via Hashing. In: *ICDM* (2016)
16. Mwangi, B., Hasan, K.M., Soares, J.C.: Prediction of individual subject’s age across the human lifespan using diffusion tensor imaging: A machine learning approach. *NeuroImage* 75, 58–67 (Jul 2013)
17. Mwangi, B., Wu, M.J., Bauer, I.E., Modi, H., Zeni, C.P., Zunta-Soares, G.B., Hasan, K.M., Soares, J.C.: Predictive classification of pediatric bipolar disorder using atlas-based diffusion weighted imaging and support vector machines. *Psychiatry Research: Neuroimaging* 234(2), 265–271 (Nov 2015)
18. Neuhaus, M., Bunke, H.: Bridging the Gap Between Graph Edit Distance and Kernel Machines. *World Scientific* (2007)
19. Pełkalska, E., Haasdonk, B.: Kernel discriminant analysis for positive definite and indefinite kernels. *IEEE PAMI* 31(6), 1017–1032 (2009)
20. Riesen, K., Bunke, H.: Approximate graph edit distance computation by means of bipartite graph matching. *Image and Vision computing* 27(7), 950–959 (2009)
21. Scholkopf, B., Smola, A.J.: *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, MA, USA (2001)
22. Shervashidze, N., Vishwanathan, S., Petri, T., Mehlhorn, K., Borgwardt, K.M.: Efficient graphlet kernels for large graph comparison. In: *AISTATS*. vol. 5, pp. 488–495 (2009)
23. Siuly, S., Zhang, Y.: Medical Big Data: Neurological Diseases Diagnosis Through Medical Data Analysis. *Data Science and Engineering* 1(2), 54–64 (Jun 2016)
24. Suard, F., Rakotomamonjy, A., Bensrhair, A.: Kernel on bag of paths for measuring similarity of shapes. In: *ESANN*. pp. 355–360 (2007)
25. Van Schependom, J., Jain, S., Cambron, M., Vanbinst, A.M., De Mey, J., Smeets, D., Nagels, G.: Reliability of measuring regional callosal atrophy in neurodegenerative diseases. *NeuroImage: Clinical* 12, 825–831 (2016)
26. Varoquaux, G., Thirion, B.: How machine learning is shaping cognitive neuroimaging. *GigaScience* 3(1), 28 (2014)
27. Vishwanathan, S.V.N., Schraudolph, N.N., Kondor, R., Borgwardt, K.M.: Graph kernels. *Journal of Machine Learning Research* 11(Apr), 1201–1242 (2010)
28. Wilson, R.C., Zhu, P.: A study of graph spectra for comparing graphs and trees. *Pattern Recognition* 41(9), 2833–2841 (2008)