

Supervised Classification of Structural Brain Networks Reveals Gender Differences

Benjamin Chiêm^{1*}, Frédéric Crevecoeur^{1,2}, Jean-Charles Delvenne¹

¹Institute of Information and Communication Technologies, Electronics and Applied Mathematics

²Institute of Neuroscience

Université catholique de Louvain, Louvain-la-Neuve, Belgium

Corresponding e-mail: *benjamin.chiem@uclouvain.be

Abstract—The human brain anatomical organization can be modeled as a complex network, called a connectome. This representation enables the use of powerful mathematical tools to detect individual differences in brain structure. In this paper, we propose three methods based on Support Vector Machines, to classify connectomes according to a given label. We demonstrate the efficiency of our methods compared to existing approaches through experiments on a dataset built from publicly available MRI data, and for the classification between male and female brains. We further interpret the classification outcome in regards to one particular wiring feature: the inter-hemispheric ratio.

I. INTRODUCTION

The human brain is composed of approximately 85 billion neurons, each of them forming thousands synapses. This highly complex organ is essential to our cognition and behaviour. Thanks to the impressive progress in neuroimaging techniques made in last years, we are now able to explore the large-scale organization of brain circuitry. Recently, much interest has been granted to an innovative vision of human brain connectivity, thereby opening a new domain called connectomics [1].

A connectome is a mathematical modeling of the structural or functional connections between brain regions as a complex network. Connectomics takes advantage of the powerful tools of modern network theory to analyse and understand our brain structure and function [1]. Remarkably, it has been shown that connectomes exhibit several interesting topological features such as small-worldness or rich-club organization, thus assessing that the topology of brain networks differs from that of random networks. This has important implications in the study of mental disorders and other neuropathologies [2].

In particular, researchers have been interested in the ability to detect early signs of these pathologies. For this purpose, several studies have applied supervised classification methods in order to identify brain differences related to some condition. Among others, Craddock and colleagues showed the ability of Support Vector Machines (SVMs) to discriminate the functional connectivity patterns of depressive subjects from that of healthy controls [3]. Jamal et al. applied similar techniques along with discriminant analysis in the context of autism spectrum disorder [4]. Furthermore, Supekar et al. also applied SVMs to demonstrate that brain functional connectivity evolves with ageing [5].

These studies mostly consider resting-state *functional* connectomes, i.e. brain networks representing the statistical inter-dependence between the electrical activity of brain regions. In this paper, we are interested in studying individual differences in the *structural* connectome. Because the structure is a constraint to the function, efficiently extracting information from the structure is expected to provide much insight. Towards this goal, we study the performance of distinct kernel-based classification techniques relying on SVMs. Considering the linear kernel, we first present a feature-based method reported in the literature, followed by a new method using the weighted edges as features. Afterwards, we propose two novel graph kernels whose efficiency is compared with existing approaches. Our strategies take advantage of a particular property of connectomes, known as the perfect node correspondence property.

In section II, we first present the processing pipeline used to generate structural connectomes starting from Magnetic Resonance Imaging (MRI) data. Then, we explain the principles underlying the classification strategies that we want to compare. In section III, we describe our testing procedure. As a case of study, we then present the results of the classification between male and female brain networks, and we show that the proposed methods discriminate connectomes more efficiently than approaches reported in the literature. Afterwards, we show how the best performing method allows us to interpret the origin of these "black-box" results. Eventually, in section IV, we discuss our results and propose perspectives for future research.

II. MATERIALS AND METHODS

A. Construction of structural connectomes

In order to build our own dataset of structural connectomes, structural and diffusion MRI data of 91 healthy men and 113 healthy women, aged from 22 to 37 years, have been retrieved from the Human Connectome Project database [6]. Then, the definition of nodes and edges was performed using the MRTrx3 toolbox [7], and according to the following two steps.

1) *Structural MRI processing*: Starting from T1-weighted MRI images, 34 cortical Regions of Interest (ROIs) per hemisphere were defined according to the automated atlas introduced by Desikan et al. [8], and 8 subcortical ROIs

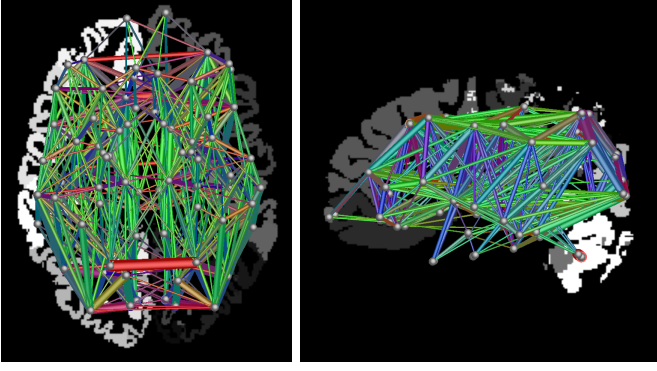


Fig. 1. Structural connectome: abstract representation of the anatomical structure of the human brain as a complex network. Gray spheres represent the ROIs. The color of the undirected edges corresponds to their spatial orientation. Left: axial view. Right: sagittal view.

per hemisphere (including cerebellum) were segmented following the procedure described by Patenaude et al. in [9]. It is interesting to notice that, following this procedure, the connectome of each of the 204 individuals will have 84 vertices. Moreover, each vertex corresponds to the same anatomical location in every connectome. This will be referred to as the perfect node correspondence property.

2) *Diffusion MRI processing and tractography*: Diffusion imaging allows us to non-invasively identify anatomical pathways inside the cerebral cortex by measuring the diffusivity of water molecules inside the brain. In white matter, this diffusivity is highly constrained by the direction of nerve fibers. Therefore, the measurement of the diffusivity along several directions allows us to deduce a privileged direction at a given voxel of the image.

Tractography is the process allowing the reconstruction of axonal bundles (also known as streamlines) from voxel to voxel. Here, we use the tractography algorithm of Tournier, Calamante and Connelly [10]. Unfortunately, in some cases, this procedure is known to introduce some biases that can affect the final result. Therefore, a filtering step is often needed after the tractography process in order to remove these biases and provide more biologically meaningful results. In our case, for each individual, we reconstructed six million streamlines and then filtered the result using the procedure described in [11], to remove spurious tracts and reduce the number of reconstructed streamlines to about four million.

Finally, the weight of an edge linking two ROIs (or vertices) was defined as the number of streamlines shared by these two ROIs. The final result is thus a connected, weighted, undirected network, as depicted in Figure 1.

B. Supervised classification

Let us consider a set of data points. Each of them is described by a vector of features. These points are labeled according to the class they belong to (C_1 or C_2). For the sake of simplicity, we will only consider two classes (binary classification) in this work. The supervised classification task

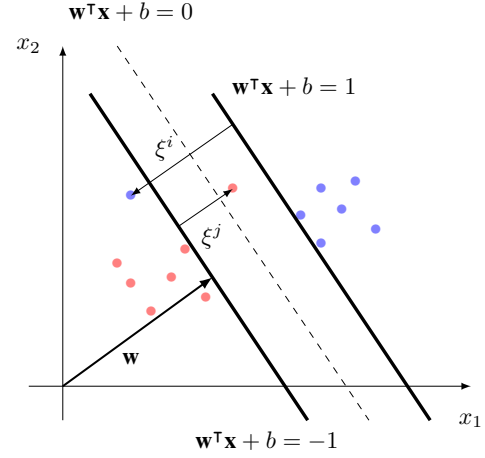


Fig. 2. The C -SVM classification algorithm seeks the maximal margin hyperplane separating two classes of data points C_1 (red) and C_2 (blue). Misclassifications due to outliers are penalized using slack variables ξ^i .

aims at determining the class of new, unclassified data points, after a training phase within the labeled dataset. Among the multitude of classification methods available in the literature, we consider here a type of algorithm known as Support Vector Machines (SVMs).

Since the early 2000s, the popularity of SVMs has continued to grow, due to their flexibility, simplicity and competitive performances compared to other classifiers. In essence, the idea behind SVMs is to formulate the classification task as an optimisation problem, seeking the separation hyperplane maximizing the width of the margin between classes, in the space spanned by the features representing the data points. In particular, the soft-margin C -SVM introduced by Cortes and Vapnik [12] is robust to outliers. Denoting by $i \in \{1, 2, \dots, N\}$ the index of each data point to be classified and defining the class indicator variable y for each data point as

$$y^i = \begin{cases} -1, & \forall \mathbf{x}^i \in C_1 \\ 1, & \forall \mathbf{x}^i \in C_2 \end{cases}$$

the C -SVM formulation is given by

$$\begin{aligned} \min_{\mathbf{w}, b} \quad & \|\mathbf{w}\|_2^2 + C \sum_{i=1}^N \xi^i \\ \text{subject to} \quad & y^i (\mathbf{w}^T \mathbf{x}^i + b) \geq 1 - \xi^i, \quad \forall i \\ & \xi^i \geq 0, \quad \forall i \end{aligned} \quad (1)$$

In Eq. (1), the objective function translates the maximization of the margin width as a minimization problem for convenience. The constant C is a parameter of the method and tunes the trade-off between the margin width and the misclassifications ξ^i related to outliers. The constraints force the labeled data points to be as correctly classified as possible. This training phase is visually summarized in Figure 2, for the two-dimensional case.

One interesting property of SVMs is their ability to generalize to non-linearly separable classes. Let us consider two

classes that are non-linearly separable in their feature space, and a non-linear mapping Φ that transforms the feature space into a new space of possibly higher dimension, in which classes become linearly separable. Such a mapping is in general hard to compute. Nevertheless, if we consider the dual formulation of Eq. (1) in the space induced by Φ :

$$\begin{aligned} \min_{\alpha^i} \quad & \frac{1}{2} \sum_{i,j=1}^N \alpha^i \alpha^j y^i y^j (\Phi^\top(\mathbf{x}^i) \Phi(\mathbf{x}^j)) - \sum_{i=1}^N \alpha^i \\ \text{subject to} \quad & 0 \leq \alpha^i \leq C, \forall i \in \{1, 2, \dots, N\} \\ & \sum_{i=1}^N \alpha^i y^i = 0 \end{aligned}$$

we observe that Φ only appears through an inner product $(\Phi^\top(\mathbf{x}^i) \Phi(\mathbf{x}^j)) = K(\mathbf{x}^i, \mathbf{x}^j)$. Hence, by knowing the result of the function K , called a kernel, for all pairs of points, we do not need to explicitly know the expression of Φ to solve the C -SVM optimization problem.

Let us notice that if Φ is the identity, the associated kernel $K(\mathbf{x}^i, \mathbf{x}^j) = (\mathbf{x}^i)^\top \mathbf{x}^j$ is called linear and corresponds to the linear separation in the original feature space. The advantage of this kernel is that it grants a weight to each feature ($\mathbf{w}$ in Eq. (1)), according to its discriminative power, and thus allows us to explain the classification result in terms of the features.

In the remaining parts of this section, we address methods specific to graph classifications. We first present two classification strategies based on the linear kernel. While the first one was already investigated in the literature, the second one is part of our contribution in this paper. Afterwards, we present the concept of a graph kernel. In particular, we propose two new graph kernels based on elements found in the literature. This constitutes the second contribution of this paper.

C. Classification strategies based on the linear kernel

Since for this part the kernel is fixed, the remaining task is to build a suitable feature vector describing a graph (see Figure 3A). The choice of features is crucial since they will constitute the fingerprint of a graph.

1) *Topological and spectral features*: A natural choice of features investigated among others by Li et al. [13] is to consider coefficients related to the topological properties of the graph, and to the spectral properties of its weighted adjacency matrix. Among the multitude of such coefficients available in the literature, we consider the following:

- | | |
|-----------------------------------|-------------------------|
| a) Mean degree | h) Radius |
| b) Mean strength | i) Diameter |
| c) Density of edges | j) Modularity |
| d) Average clustering coefficient | k) Assortativity |
| e) Transitivity coefficient | l) Diffusion efficiency |
| f) Characteristic path length | m) Spectral radius |
| g) Global efficiency | n) Frobenius norm |
| | o) Energy |

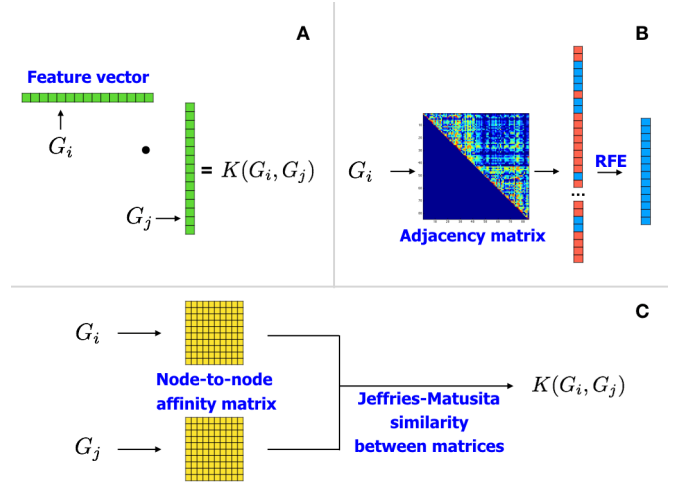


Fig. 3. Methodological considerations about kernel methods. **A** The linear kernel between two connectomes G_i and G_j is defined as the scalar product between their respective feature vectors. **B** For graphs having the perfect node correspondence property, we can build a feature vector by vectorizing the adjacency matrix and then applying Recursive Features Elimination. **C** Using node-to-node affinity matrices to represent two graphs G_i and G_j having the perfect node correspondence property, we can define a kernel being the similarity between these matrices.

Topological coefficients (a-l) are described in [14] and were computed thanks to the Brain Connectivity Toolbox¹. The spectral radius is defined as the largest eigenvalue of the weighted adjacency matrix, in absolute value [15]. The Frobenius norm of a real weighted adjacency matrix is the square root of the sum of its squared elements [15]. The energy of a graph, as defined by Gutman in [16] (although other definitions exist), is the $L1$ -norm of the spectrum of its weighted adjacency matrix.

Following this strategy, the kernel between two graphs G_i and G_j is thus defined as the scalar product between their respective feature vector, as illustrated in Figure 3A.

2) *Bag-of-Edges*: It is known that we can completely describe a graph using its weighted adjacency matrix. Hence, we could vectorize this matrix to build a feature vector that would be a perfect fingerprint of the graph. Unfortunately, this would result in a large feature vector whose size would impact the performance of the SVM. To overcome this issue, we apply a procedure known as Recursive Features Elimination (RFE).

Recursively, the RFE procedure trains an SVM with linear kernel, sorts features according to the weights granted by the SVM, and reduces the size of the feature vector by only keeping a given percentage of the most discriminative features. The RFE is stopped when a given number of features is reached.

The Bag-of-Edges strategy thus builds a feature vector whose components are the most discriminative edges of the graph, according to the classes under consideration. This is illustrated in Figure 3B. This approach has received little

¹<https://sites.google.com/site/bctnet/>

attention in the literature on graph classification since most of existing methods were designed to classify variable size graphs. Furthermore, in the context of connectomes classification, this approach allows us to locate the origin of the difference in terms of anatomical links.

D. Classification strategies based on graph kernels

Intuitively, the value of a kernel $K(G_i, G_j)$ can be understood as a measure of similarity between G_i and G_j . From this, several graph kernels have been developed such as the Random Walk kernel [17] that performs random walks on both graphs and counts the number of matching walks, or the Weisfeiler-Lehman kernel [18] that defines a measure of similarity based on the local neighbourhood of the nodes, up to a given depth. These kernels are suitable for the classification of graphs having different sizes. However, given the perfect node correspondence of the connectomes under consideration, we seek to take advantage of this property to design new graph kernels.

1) *DeltaCon kernel*: Recently, Koutra and colleagues proposed a new similarity measure between graphs having the perfect node correspondence property, called DeltaCon [19]. The principle underlying this metric is summarized in Figure 3C. First, we compute for each graph a node-to-node affinity matrix. This matrix, also known as a kernel *on* a graph in the field of semi-supervised learning, encodes a particular measure of similarity between the nodes of the graph under consideration. In [19], the authors suggest using the Fast Belief Propagation (FaBP) affinity matrix derived in [20]. Then, the DeltaCon kernel (or similarity) K_{DC} between two graphs G_i and G_j is defined as

$$K_{DC}(G_i, G_j) = (1 + d_{JM}(\mathbf{S}_i, \mathbf{S}_j))^{-1}$$

where the matrices $\mathbf{S}_i$ and $\mathbf{S}_j$ are the node-to-node affinity matrices of G_i and G_j respectively, and where d_{JM} stands for the Jeffries-Matusita distance between matrices.

It turns out that the DeltaCon measure has interesting properties, as demonstrated in [19]. However, it has not yet been applied as a kernel in a supervised classification framework, which is why we want to investigate its performance in this paper.

2) *Kernel based on the regularized Laplacian*: Motivated by the previous approach, we want to investigate the influence of the node-to-node affinity measure on the quality of the DeltaCon kernel. Let us introduce another such measure, known as the normalized Laplacian $\mathcal{L}$ of a graph:

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

In Eq. (2), $\mathbf{I}$ is the identity matrix, $\mathbf{D}$ is the diagonal matrix encoding the degree of the nodes and $\mathbf{A}$ is the weighted adjacency matrix of the graph. Then, we can define the regularized Laplacian $\mathbf{S}_{RL}$:

$$\mathbf{S}_{RL} = (\mathbf{I} + \alpha \mathcal{L})^{-1} \quad (3)$$

In Eq. (3), α is a parameter of the method. Eventually, to test the influence of the node-to-node affinity measure, we build a new kernel K_{RL} following the same approach as the DeltaCon kernel (see Figure 3C), but using the regularized Laplacian instead of the Fast Belief Propagation.

III. RESULTS

In this section, we begin by presenting the experimental setup we used in order to test our methods, followed by the associated results. Then, we explain our findings by interpreting the origin of the classification outcome.

A. Experimental setup

A test consists of a random shuffling of the entire dataset, which is then split in training and test sets. After that, model selection is performed on the training set using a 5-fold cross-validation scheme. Finally, the performances of our methods are assessed on the remaining unseen test set by computing the classification accuracy, i.e. the proportion of correct classifications. We performed 30 such tests and averaged the results.

Since our dataset is slightly unbalanced (113 over 204 connectomes were from female subjects), we also consider Cohen's kappa κ as a second performance metric. Introduced by Jacob Cohen in [21], κ belongs to $[-1, 1]$ and can detect issues induced by unbalanced classes. By considering both classification accuracy and κ , we can assess the performance of the methods as well as the reliability of the results.

For the aspects related to SVMs, RFE and the performance evaluation procedure, we used the Scikit-learn PYTHON toolbox². The MATLAB implementation of both Random Walk and Weisfeiler-Lehman kernels are publicly available on the internet³. The DeltaCon kernel was computed using the code publicly available on the website⁴ of the first author of [19].

B. Classification with respect to gender

The classification results are reported in Table I. We can observe that all methods succeed in finding a difference between male and female brain networks, and that the Bag-of-Edges strategy is the best for both performance metrics. We also see that both Random Walk and Weisfeiler-Lehman kernels are less efficient than the other methods. Eventually, let us remark that the regularized Laplacian seems to induce better results than the Fast Belief Propagation affinity measure proposed in [19].

C. The inter-hemispheric ratio

When looking at Table I, we observe that the Bag-of-Edges (BoE) strategy succeeds in finding a linear separation between male and female connectomes in the space spanned by the weights of the edges. Intrigued by this observation, we want to investigate the origin of this separation. In Figure 4, we depict the 400 most discriminative edges obtained by Recursive Features Elimination. These edges are traced in the

²<http://scikit-learn.org/>

³<http://www.mit.edu/~pinary/kdd/>

⁴<http://web.eecs.umich.edu/~dkoutra/>

TABLE I

CLASSIFICATION RESULTS. TSF: TOPOLOGICAL AND SPECTRAL FEATURES, BoE: BAG-OF-EDGES, DC: DELTA CON KERNEL, RL: KERNEL BASED ON THE REGULARIZED LAPLACIAN, RW: RANDOM WALK KERNEL, WL: WEISFEILER-LEHMAN KERNEL. MEAN $\pm$ STANDARD DEVIATION, AVERAGED OVER 30 INDEPENDENT TESTS. THE HIGHER IS THE BETTER FOR BOTH PERFORMANCE METRICS. $\dagger$: METHODS INTRODUCED IN THIS PAPER.

	TSF	BoE †	DC †	RL †	RW	WL
Accuracy [%]	66.99 $\pm$ 7.4	83.09 $\pm$ 5.5	75.37 $\pm$ 5.7	77.32 $\pm$ 6.9	55.80 $\pm$ 7.7	62.20 $\pm$ 5.7
100 $\times$ κ [-]	35.45 $\pm$ 12.6	66.92 $\pm$ 11.2	47.95 $\pm$ 10.4	52.80 $\pm$ 12.7	19.41 $\pm$ 13.5	19.89 $\pm$ 13.9

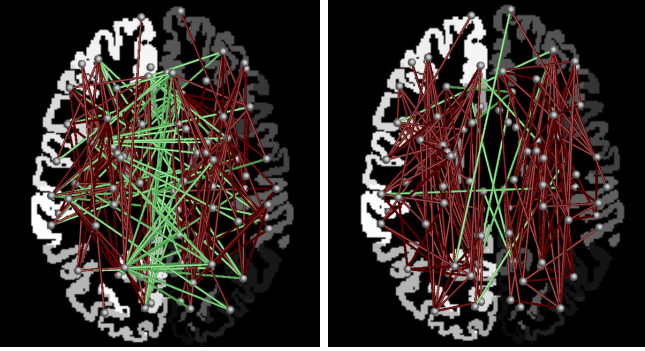


Fig. 4. Left: connectome corresponding to the average of the adjacency matrices related to female subjects. Right: connectome corresponding to the average of the adjacency matrices related to male subjects. The 400 most discriminative edges obtained by RFE are depicted in the connectome where their normalized weight is the highest. Inter-hemispheric edges are traced in green, intra-hemispheric edges are traced in red.

connectome where their normalized weight is the highest. An interesting observation in this Figure is that inter-hemispheric connections are more present in female brain.

Let us consider a connectome $G(\mathcal{V}, \mathcal{E})$ with vertex set $\mathcal{V}$ and edge set $\mathcal{E}$. We denote by $\mathcal{E}_{inter}$ the set of inter-hemispheric edges (i.e., linking two nodes lying in opposed hemispheres). The function $w : \mathcal{E} \rightarrow \mathbb{R} : e \rightarrow w(e)$ gives the normalized weight of an edge e . We define the inter-hemispheric ratio as

$$\zeta = \frac{\sum_{e' \in \mathcal{E}_{inter}} w(e')}{\sum_{e \in \mathcal{E}} w(e)}$$

In the left part of Figure 5, we depict the histograms of the inter-hemispheric ratio for both male and female connectomes. If we consider the red dashed line as a threshold ($\zeta^* = 0.11$) for the classification between male and female connectomes, 75% of the connectomes constituting our dataset are correctly classified. This is not as much as the accuracy reached by the BoE strategy in Table I, but we emphasize here the fact that one single feature is able to discriminate male and female connectomes more efficiently than the Random Walk and Weisfeiler-Lehman graph kernels. The right part of Figure 5 presents the Receiver Operating Characteristic (ROC) curve associated with these histograms. The area under the curve is 0.8058. Similar results about the inter-hemispheric ratio are discussed by Ingahlalikar et al. in [22].

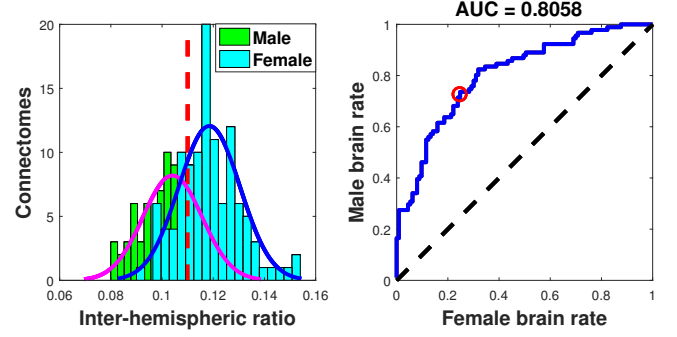


Fig. 5. Left: histograms of the inter-hemispheric ratio for both male and female connectomes. Right: Corresponding Receiver Operating Characteristic curve. The red circle corresponds to the red dashed threshold line on the histograms. AUC: Area Under the Curve. AUC=1 corresponds to a perfect separation of the two classes. AUC=0.5 corresponds to totally overlapping classes.

IV. DISCUSSION

A. Choice of kernel

The choice of a kernel for graph classification is still an open question. In the present work, we investigated the performance of two methods based on the linear kernel, with the motivation to explain *a posteriori* the classification result. In particular, the Bag-of-Edges approach allowed us to partly detect the origin of the difference between male and female connectomes, by defining the inter-hemispheric ratio. Among the remaining methods, we proposed two graph kernels taking advantage of the perfect node correspondence property and observed that taking this property into account can help the classification process.

Nevertheless, our intuition is that each kernel captures a particular facet of what a graph is. From this statement, we strongly believe that more attention should be paid to Multiple Kernel Learning (MKL) approaches in future work because they aim at combining several kernels representing different aspects of the same object, in order to improve the classification result.

Eventually, no matter what kernel is chosen, the positive semidefiniteness of the resulting $N \times N$ symmetric kernel matrix $\mathbf{K}$, defined as

$$[\mathbf{K}_{ij}]_{i,j=1}^N = [K(\mathbf{x}^i, \mathbf{x}^j)]_{i,j=1}^N$$

must be assessed in order to ensure that the C -SVM optimization problem is well-posed, as required by Mercer's theorem. The linear kernel guarantees this condition, as well as the Random Walk and the Weisfeiler-Lehman kernels. The DeltaCon

kernel and the kernel based on the regularized Laplacian that we introduced in this paper gave rise to positive semidefinite kernel matrices during our experiments. Nonetheless, there is no proof (to our knowledge) that it will always be the case. Consequently, future work using these kernels should always pay attention to this aspect.

B. Implications of the difference between male and female

The present work highlighted structural brain differences related to gender. In [22], the authors argue that these differences directly impact the behaviour of the individuals, and that male and female brains are not designed for the same tasks. However, it is not yet clear if these differences are truly at the origin of the biological distinction between men and women, or if they are rather consequences of different educational, social and environmental conditions. Moreover, the link between brain structure and function is still fuzzy, and inferring the impact of changes in brain circuitry on behaviour is not straightforward.

Notwithstanding, being aware of differences in brain structure is of crucial importance when conducting group studies in neuroscience. We emphasized in this work structural differences related to gender, but further studies should examine if other factors are related to differences in brain structure, since these differences may bias experiments.

V. CONCLUSION

Modeling the human brain as a complex network offers several advantages. In particular, we can benefit from the formalism of graph theory to study the intricate patterns shaping brain structure. The objectives of the present work were twofold. Firstly, we showed how advances in the analysis of diffusion and structural MRI data allow us to build a structural brain network. Secondly, after having built our own dataset composed of 204 such connectomes, we addressed the question of detecting individual differences in brain structure that are related to a particular label. Our contribution for this part is the introduction of three new supervised graph classification methods for the specific case of graphs exhibiting the perfect node correspondence property. The efficiency of the proposed methods was assessed comparatively to existing approaches, in the framework of male-female brain classification. Eventually, we demonstrated the ability of one of our methods to provide clues about the origin of the detected difference, showing that on average women have stronger anatomical connections between hemispheres than men.

Although the present work restricted its application to the particular case of gender differences, the scope of the proposed methods can be extended in future work to other labels, related for example to pathological conditions or behavioral habits.

ACKNOWLEDGMENT

Benjamin Chi  m is a FRIA (F.R.S.-FNRS) Fellow. This paper presents research results of the Concerted Research Action (ARC) programme supported by the Federation Wallonia-Brussels (contract ARC 14/19-060). Data were provided by

the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

REFERENCES

- [1] E. Bullmore and O. Sporns, "Complex brain networks: graph theoretical analysis of structural and functional systems," *Nature Reviews Neuroscience*, 10(3), 186-198, 2009.
- [2] U. Braun, S. F. Muldoon and D. S. Bassett, "On human brain networks in health and disease," *eLS*, 2015.
- [3] R. C. Craddock, P. E. Holtzheimer, X. P. Hu and H. S. Mayberg, "Disease state prediction from resting state functional connectivity," *Magnetic resonance in Medicine*, 62(6), 1619-1628, 2009.
- [4] W. Jamal, S. Das, I. A. Oprea, K. Maharatna, F. Apicella and F. Sicca, "Classification of autism spectrum disorder using supervised learning of brain connectivity measures extracted from synchronostates," *Journal of neural engineering*, 11(4), 046019, 2014.
- [5] K. Supekar, M. Musen, V. Menon, "Development of large-scale functional brain networks in children," *PLoS biology*, 7(7), e1000157, 2009.
- [6] D. C. Van Essen, S. M. Smith, D. M. Barch, T. E. Behrens, E. Yacoub, K. Ugurbil and Wu-Minn HCP Consortium, "The WU-Minn human connectome project: an overview," *Neuroimage*, 80, 62-79, 2013.
- [7] J. Tournier, F. Calamante and A. Connelly, "MRtrix: diffusion tractography in crossing fiber regions," *International Journal of Imaging Systems and Technology*, 22(1), 53-66, 2012.
- [8] R. S. Desikan, F. S  gonne, et al., "An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest," *Neuroimage*, 31(3), 968-980, 2006.
- [9] B. Patenaude, S.M. Smith, D.N. Kennedy, and M. Jenkinson, "A Bayesian model of shape and appearance for subcortical brain segmentation," *Neuroimage*, 56(3), 907-922, 2011.
- [10] J.D. Tournier, F. Calamante, and A. Connelly, "Improved probabilistic streamlines tractography by 2nd order integration over fibre orientation distributions," In *Proc. 18th Annual Meeting of the Intl. Soc. Mag. Reson. Med. (ISMRM)* (p. 1670), 2010.
- [11] R.E. Smith, J.D. Tournier, F. Calamante and A. Connelly, "SIFT: spherical-deconvolution informed filtering of tractograms," *Neuroimage*, 67, 298-312, 2013.
- [12] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, 20(3), 273-297, 1995.
- [13] G. Li, M. Semerci, B. Yener and M.J. Zaki, "Effective graph classification based on topological and label attributes," *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 5(4), 265-283, 2012.
- [14] M. Rubinov, O. Sporns, "Complex network measures of brain connectivity: uses and interpretations," *Neuroimage*, 52(3), 1059-1069, 2010.
- [15] G.H. Golub and C.F. Van Loan, "Matrix computations," (Vol. 3), JHU Press, 2012.
- [16] I. Gutman, "The energy of a graph: old and new results," *Algebraic combinatorics and applications*, 196-211, 2001.
- [17] S.V.N. Vishwanathan, N.N. Schraudolph, R. Kondor and K.M. Borgwardt, "Graph kernels," *Journal of Machine Learning Research*, 11(Apr), 1201-1242, 2010.
- [18] N. Shervashidze, P. Schweitzer, E.J.V. Leeuwen, K. Mehlhorn and K.M. Borgwardt, "Weisfeiler-lehman graph kernels," *Journal of Machine Learning Research*, 12(Sep), 2539-2561, 2011.
- [19] D. Koutra, J.T. Vogelstein, and C. Faloutsos, "Deltacon: A principled massive-graph similarity function," in *Proceedings of the 2013 SIAM International Conference on Data Mining* (pp. 162-170), Society for Industrial and Applied Mathematics, 2013.
- [20] D. Koutra, T.Y. Ke, U. Kang, D.H.P. Chau, H.K.K. Pao and C. Faloutsos, "Unifying guilt-by-association approaches: Theorems and fast algorithms," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (pp. 245-260), Springer, Berlin, Heidelberg, 2011.
- [21] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and psychological measurement*, 20(1), 37-46, 1960.
- [22] M. Ingahlhalikar, A. Smith, D. Parker, T.D. Satterthwaite et al., "Sex differences in the structural connectome of the human brain," *Proceedings of the National Academy of Sciences*, 111(2), 823-828, 2014.