# Classifying connectivity graphs using graph and vertex attributes

Jonas Richiardi*†, Sophie Achard‡, Edward Bullmore§¶‖, and Dimitri Van De Ville*†
*Medical Image Processing Laboratory,Ecole Polytechnique Fédérale de Lausanne, Switzerland
†Medical Image Processing Laboratory,University of Geneva, Switzerland
‡GIPSA-Lab, Institut National Polytechnique de Grenoble, France
§Brain Mapping Unit, Department of Psychiatry, University of Cambridge, Cambridge, UK
¶Behavioural and Clinical Neurosciences Institute, University of Cambridge, Cambridge, UK
‖GSK Clinical Unit Cambridge, Addenbrooke's Hospital, Cambridge, UK

***Abstract*—Qualitative and quantitative description of functional connectivity graphs using graph attributes is of great interest to neuroscience, and has led to remarkable insights in the field. However, the statistical techniques used have generally been limited to whole-group, post-hoc studies. In this paper, we propose instead a novel approach to perform predictive inference on single subjects. It is based on a lossy embedding of connectivity graphs into a vector space using graph and vertex attributes, followed by the use of statistical machine learning to build a predictive model. The feature space proposed is easily interpretable for neuroscientists, and we illustrate the technique by revealing resting-state difference between young and elderly subjects.**

***Keywords*-graph attributes;connectivity decoding;graph embedding**

## I. Introduction

In pattern recognition literature, the graph classification task is generally approached through graph matching followed by classification (typically using $k$-NN) based on a matching cost (see e.g. [1]). When graphs are restricted to have a fixed-cardinality vertex sequence, however, the matching step is unnecessary and classification becomes suboptimal [2]. A recent class of approaches is to represent the graph in a vector space (graph embedding) [3], [4] , and to use statistical machine learning tools to perform classification. Thus, the design effort is spent on devising embeddings that ideally provide good discrimination between classes.

Here we propose a new embedding technique based on graph and vertex attributes. This is of interest because these attributes have been the focus of much attention in recent neuroscience research (see e.g. [5], [6]), and provide interpretable results, but have generally been applied in a post-hoc setting. The proposed embedding technique is lossy in the sense that the original graph cannot be reconstructed from the vector-space representation. However, our goal is to obtain a discriminative representation easy to interpret, rather than a good reconstruction. We illustrate the technique on a young-versus-old classification task.

## II. Representing connectivity graphs in vector space

### A. Connectivity graph computation

After motion correction and coregistration to an MNI template (SPM2 [1]), subjects displaying excessive head movement (in excess of 3 mm translation, or 0.38 rotation, in x, y, or z dimensions) are discarded. No spatial smoothing is used, and the data undergoes regional parcellation into 45 bilateral regions (total 90) using the anatomically labeled template image validated by Tzourio-Mazoyer et al. [7].

[1]available at http://www.fil.ion.ucl.ac.uk

The adjacency matrix for the functional connectivity graph in each subject is then constructed in the following manner [5]:

1) Regional mean time series are estimated by averaging the fMRI time series over all voxels in each region.
2) The confounding signals from the movement parameters are regressed out from the regional time series.
3) The pairwise inter-regional correlations between wavelet coefficients corresponding to 0.06-0.11 Hz are computed to produce a correlation matrix among regions. By removing the diagonal, this forms the adjacency matrix of an undirected, weighted graph.
4) The connectivity graphs are then explored by retaining the absolute correlations of statistical significance using multiple hypothesis test at 5%, and those greater than a given threshold chosen such that every graph has the same number of connections are retained.

This procedures yields a labeled simple graph $g = (V, E, \alpha, \beta)$, which is a 4-tuple consisting of a set of vertices $V$, a set of edges $E$, and labeling functions $\alpha$ and $\beta$ assigning respectively vertex and edge labels. In the present case the cardinality of the vertex set is fixed to the number of regions ($|V| = R$) for all graphs, the vertex labeling function $\alpha$ assigns unique and corresponding node labels for all graphs because of the atlasing procedure, and $\beta$ is a scalar function assigning correlation values as edge labels. For instance $\beta(1, 2)$ yields the correlation coefficient between regions 1 and 2.

### B. Graph and vertex attributes

As the graphs of interest have fixed-cardinality vertex sequences due to the atlasing procedure, each graph $g$ is uniquely defined by its adjacency matrix $\mathbf{A}$ and vice versa, because the ordering of vertices is not arbitrary. Hence, with the vertex problem correspondance solved a priori, this in turn entails that an isomorphism between graphs of this class only needs satisfiy equality of edges and edge labels (a problem subsumed by isomorphism of graphs with unique node labels [8]).

Thus, it is relatively easy to use or propose several vertex attributes that obey the definition of vertex invariants, namely, that return the same value for isomorphic graphs. In this paper, we used five basic vertex attributes already defined and used in [9], [5], [6].

- **Strength :** The pairwise correlations are averaged for each region in order to quantify the correlation weight for each region $i$, $1 \leq i \leq R$ :

$$S(i) = \frac{1}{R-1} \sum_{k=1, k \neq i}^{R} \beta(i, k)$$

978-0-7695-4399-4/11 $26.00 © 2011 IEEE
DOI 10.1109/PRNI.2011.18

- **Diversity :** The variance of the pairwise correlations for each regions
$$Div(i) = \frac{1}{R-1} \sum_{k=1,k\neq i}^{R} (\beta(i,k) - S_i)^2$$
Strength and diversity are computed on the adjacency matrix for the complete graph. The remaining three properties are computed on the thresholded adjacency matrix.
- **Degree :** The number of edges connected to each region $i$, $1 \leq i \leq R$ :
$$D(i) = \sum_{k=1,k\neq i}^{R} A_{ik}$$
- **Global efficiency :** The shortest path between pairs of regions $i$ and $k$, denoted $l_{ik}$. High global efficiency corresponds to regions that are highly connected to other distant regions.
$$E_g(i) = \frac{1}{R-1} \sum_{k=1,k\neq i}^{R} \frac{1}{l_{ik}}$$
- **Local efficiency :** Based on the connections between neighbours of a selected region. Let us denote $G_i$ the subgraph of vertex $i$ composed of the direct neighbours of $i$. High local efficiency corresponds to the ability to easily compensate for the loss of a region.
$$E_l(i) = \frac{1}{R_{G_i} - 1} \sum_{k,j\in G_i}^{R} \frac{1}{l_{jk}}$$

From these vertex attributes, graph attributes can be computed by averaging the vertex attributes over all vertices of the graph.

Since the thresholding procedure is somewhat arbitrary, we also sparsify the graph with several different thresholds, compute graph and vertex attributes, and average them across thresholds. The range of thresholds was chosen in order to cover the *small-world regime* [10]. Here, we define the range from 100 to 2000 edges with a step of 100 which corresponds to 2.5% and 50% of the total number of possible edges. We denote this as the "range" technique later in the paper.

### C. Lossy embedding with graph and vertex attributes

Formally, a graph attribute is the result of the computation of a function $f : g \mapsto \mathbb{R}$. A vertex attribute over a graph with $R$ vertices can be seen as a function $f_v : g \mapsto \mathbb{R}^R$. Thus, given a set of functions over graphs $F = \{f_1, \ldots, f_{n_1}\}$ and a set of functions over vertices $F_v = \{f_{v_1}, \ldots, f_{v_{n_2}}\}$, a graph can be represented in vector spaces $\mathbb{R}^{|F|}$ and $\mathbb{R}^{R\times|F_v|}$. This representation forms an embedding, which is lossy because the original adjacency matrix representing the graph cannot be reconstructed from the vector-space representation.

We note that this approach is different from [11], which applies to directed acyclic graphs, and relies on partial eigenspectrum sums in subgraphs to achieve vector-space embedding.

## III. Experiments

### A. Dataset

*1) Subjects:* Thirty healthy human volunteers were recruited in two age groups: 17 younger participants aged 18-33 years, mean age = 24.3 years, nine male; and 13 older participants aged 62-76 years, mean age = 67.3 years, six male. Two young and two old subjects were excluded due to excessive motion, resulting in a sample of 15 young and 11 old subjects. Exclusion criteria included a history of neurological or psychiatric disorder, current treatment with vasoactive or psychotropic medication, or any contraindications to MRI or study drug. Prior to functional MRI scanning, each participant also had an electrocardiogram and a structural MRI scan reviewed as normal by a physician. All participants gave wrtten informed consent. The study was approved by the Addenbrooke's National Health Service Trust Local Research Ethics Committee, Cambridge, UK. Two different analyses of data acquired on this sample have been previously reported [10], [12].

*2) fMRI acquisition:* Each participant was scanned lying quietly at rest with eyes closed for 9 min, 37.5s. Gradient-echo echoplanar imaging (EPI) data depicting BOLD contrast were acquired using a Medspec S300 3 T scanner (Bruker Medical) in the Wolfson Brain Imaging Centre (Cambridge, UK). We acquired 525 volumes with the following parameters: number of slices, 21 (interleaved); slice thickness, 4 mm; interslice gap, 1 mm; matrix size, 64 $\times$ 64; flip angle, 90°; repetition time (TR), 1100 ms; echo time, 27.5 ms; in-plane resolution, 3.125 mm. The first seven volumes were discarded to allow for T1 saturation effects, leaving 518 volumes available for analysis of resting state connectivity in each subject.

### B. Classification tasks

We perform young-versus-old classification experiments in a leave-one-subject-out crossvalidation setting. The goal is to see whether a particular feature, or combination of features, is discriminative in this respect. To lower the risk of falsely concluding to an absence of discriminative power, we test several classifiers, both generative and discriminative. The classifiers used are Naïve Bayes with kernel densities ($NB_k$), SVMs with linear and 2nd order normalised polynomial kernels ($SVM_l$ and $SVM_p$, cost 10), a radial basis function network ($RBF$, 3 clusters), a multi-layer perceptron ($MLP$), and trees(C4.5 ($C4.5$), functional tree ($FT$) with minimum 3 items per branch, random forest ($RF$) with 401 trees and 2 features per tree). All classifiers are implemented in Weka [13] and use default settings except where mentioned. The RF classifier was provided by Abhishek Jaiantilal[2].

*1) Graph attributes:* The first set of experiments is based on using whole-brain attributes, global and local efficiency. For each parcellation scale, these are extracted and made into two $E_g, E_l \in \mathbb{R}^{1\times1}$ feature vectors, and one joint feature vector $(E_g, E_l)^T \in \mathbb{R}^{2\times1}$ for each subject. Other graph properties could be combined in the same way. The results are reported in Table I. We also examine the between-group difference in means (or medians) using a single-factor ANOVA (or Kruskall-Wallis test), selecting the hypothesis test according to Gaussianity of the data as established by a Jarque-Bera test at $p < 0.05$.

For global efficiency, 2 classifiers fail at achieving above-chance accuracy in both classes, both for attributes computed for 400 edges and attributes computed over a range of edges. For local efficiency, only the polynomial SVM fails. This indicates that both attributes are good predictors of age group, because a variety of optimisation techniques and decision boundary forms are able to predict reasonably well. Using the range technique leads to marked improvements in the discriminative power afforded by local efficiency.

[2]https://code.google.com/p/randomforest-matlab/

Table I

LEAVE-ONE-OUT CLASSIFICATION ACCURACIES (IN PERCENTAGE) ON GRAPH ATTRIBUTES $E_g$ AND $E_l$. $P_{y,o}$ INDICATES ACCURACY FOR YOUNG AND OLD RESPECTIVELY, $P_b$ IS THE BALANCED ACCURACY. YELLOW INDICATES THE BEST RESULTS.

| $\|E\|$ | Class. | $E_g$ | | | $E_l$ | | | $(E_g, E_l)$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $P_y$ | $P_o$ | $P_b$ | $P_y$ | $P_o$ | $P_b$ | $P_y$ | $P_o$ | $P_b$ |
| 400 | $NB_k$ | 87 | 64 | 75 | 80 | 64 | 72 | 80 | 64 | 72 |
| | $SVM_l$ | 87 | 64 | 75 | 80 | 73 | 76 | 80 | 64 | 72 |
| | $SVM_p$ | 87 | 0 | 43 | 100 | 0 | 50 | 93 | 9 | 51 |
| | $RBF$ | 80 | 64 | 72 | 67 | 55 | 61 | 80 | 64 | 72 |
| | $MLP$ | 87 | 55 | 71 | 67 | 64 | 65 | 73 | 64 | 68 |
| | $C4.5$ | 87 | 45 | 66 | 60 | 82 | 71 | 60 | 82 | 71 |
| | $FT$ | 87 | 64 | 75 | 80 | 64 | 72 | 73 | 64 | 68 |
| | $RF$ | 73 | 64 | 68 | 60 | 55 | 57 | 60 | 55 | 57 |
| | $p$-value | 0.001 ($F = 14.01$) | | | <0.003 ($\chi^2 = 9.22$) | | | n/a | | |
| range | $NB_k$ | 87 | 55 | 71 | 87 | 82 | 84 | 80 | 82 | 81 |
| | $SVM_l$ | 80 | 73 | 76 | 87 | 91 | 89 | 80 | 91 | 85 |
| | $SVM_p$ | 100 | 0 | 50 | 67 | 0 | 33 | 100 | 18 | 59 |
| | $RBF$ | 80 | 45 | 63 | 93 | 64 | 78 | 80 | 64 | 72 |
| | $MLP$ | 93 | 55 | 74 | 87 | 82 | 84 | 87 | 82 | 84 |
| | $C4.5$ | 93 | 64 | 78 | 87 | 82 | 84 | 80 | 55 | 67 |
| | $FT$ | 80 | 73 | 76 | 87 | 91 | 89 | 80 | 82 | 81 |
| | $RF$ | 80 | 55 | 67 | 80 | 73 | 76 | 80 | 55 | 67 |
| | $p$-value | <0.001 ($F = 14.26$) | | | <0.001($\chi^2 = 11.21$) | | | n/a | | |

The best results are obtained using local efficiency (range technique) as a feature together with a linear SVM or functional tree classifier. It is in fact quite remarkable that a single scalar is able to distinguish so well between old and young subjects. This is also hinted at by the low p-values.

The combination of global and local efficiency generally fails to improve predictive ability over the best 1D feature, or even worsens accuracy, presumably because the two are strongly correlated ($\rho = 0.83$). It would be of interest to include other more orthogonal graph attributes.

*2) Vertex attributes:* The second set of experiments attempts to classify graphs based on vectors of vertex attributes, for each subject $\mathbb{R}^{R\times 1}$ vectors. As an example, a composite vector $(\mathbf{E}_g, \mathbf{E}_l)^T$ is also extracted by concatenating local attribute vectors into a $\mathbb{R}^{2R\times 1}$ vector per subject. Several vertex attribute vectors could be combined in this way. Table II shows the results.

Table II

LEAVE-ONE-OUT CLASSIFICATION ACCURACIES (IN PERCENTAGE) ON VERTEX ATTRIBUTE VECTORS. $P_b$ IS THE BALANCED ACCURACY. $\oplus$, RESP. $\ominus$, INDICATES AN ACCURACY GAIN, RESP. ACCURACY LOSS, OVER THE BEST OF $\mathbf{E}_g$ OR $\mathbf{E}_l$ BY COMBINING THEM. YELLOW INDICATES THE BEST RESULTS FOR EACH VERTEX ATTRIBUTE.

| $\|E\|$ | Class. | $\mathbf{E}_g$ | $\mathbf{E}_l$ | $\mathbf{D}$ | $\mathbf{Str}$ | $\mathbf{Div}$ | $(\mathbf{E}_g, \mathbf{E}_l)$ |
|---|---|---|---|---|---|---|---|
| | | | | $P_b$ | | | $P_b$ |
| 400 | $NB_k$ | 66 | 54 | 78 | 54 | 81 | 58 $\ominus$ |
| | $SVM_l$ | 73 | 66 | 76 | 66 | 72 | 75 $\oplus$ |
| | $SVM_p$ | 59 | 72 | 65 | 71 | 73 | 74 $\oplus$ |
| | $RBF$ | 43 | 44 | 57 | 50 | 67 | 55 $\oplus$ |
| | $MLP$ | 73 | 62 | 63 | 66 | 72 | 75 $\oplus$ |
| | $C4.5$ | 76 | 36 | 89 | 80 | 68 | 76 |
| | $FT$ | 68 | 52 | 72 | 67 | 71 | 72 $\oplus$ |
| | $RF$ | 74 | 51 | 66 | 52 | 72 | 68 $\ominus$ |
| range | $NB_k$ | 62 | 76 | 56 | 54 | 81 | 72 $\ominus$ |
| | $SVM_l$ | 73 | 74 | 68 | 66 | 72 | 76 $\oplus$ |
| | $SVM_p$ | 64 | 80 | 64 | 71 | 73 | 65 $\ominus$ |
| | $RBF$ | 46 | 53 | 44 | 50 | 67 | 44 $\ominus$ |
| | $MLP$ | 73 | 71 | 70 | 66 | 72 | 76 $\oplus$ |
| | $C4.5$ | 61 | 54 | 68 | 80 | 68 | 61 |
| | $FT$ | 72 | 71 | 66 | 67 | 71 | 81 $\oplus$ |
| | $RF$ | 69 | 59 | 65 | 57 | 67 | 77 $\oplus$ |

The first notable results is that all 5 vertex properties yield balanced accuracies above chance in general, and that they can reach around 80% accuracy given the proper classifier. Overall, the peaks of accuracy are reached by the NB or C4.5 classifiers.

*3) Direct embedding:* For comparison, we perform the same classification task using direct connection label sequence embedding [14], whereby the upper triangular part of the unthresholded adjacency matrix is vectorised and used directly as a feature vector. Summarising the results, the best performance is obtained by a linear SVM, resulting in $P_y = 87\%, P_o = 64\%, P_b = 76\%$. Embedding the thresholded 400-edges weighted graphs in the same way, the best result is obtained by a C4.5 tree, at $P_y = 93\%, P_o = 73\%, P_b = 83\%$ This is below the best results that can be obtained by using vertex or graph attributes on this task, hinting that it can indeed be beneficial to extract graph attributes to better represent structural relationships in graphs. These may be difficult or impossible to learn using classical machine learning on a vector-space representation, where the relational information is lost or captured to a low degree only. The relative small dimensionality of the feature space generated by graph and vertex attributes further underlines the potential of using a representation based on these attributes.

### C. Regional importances

Based on the classifiers trained on vertex attributes and their classification performance, we can compute a mapping in brain space to show which graph regions are (jointly) more discriminative between the two classes. As usual, in this respect two main types of classifiers exist: those that use a sparse discriminant function (in our case, trees), where not all features are selected and/or assigned a weight, and those that use all features in the discriminant function (SVM (trained with L2 loss), MLP, RBF), yielding a "dense" map. For "sparse" classifiers that do not assign a weight to features, the number of times a feature is picked can serve as a proxy of its importance. For "dense" maps derived from linear classifiers, the (normalised) weight attached to a feature can be used. We note that the C4.5 classifier indeed does pick very few regions to achieve good discrimination (as little as 3 for the $\mathbf{D}$ attribute, concentrating on orbito-frontal regions and the amygdala, which is consistent with previous work on the dataset), and yields shallow trees.

Figures 1 and 2 present dense results for the linear SVM classifier on global efficiency and joint local and global efficiency. For global efficiency, we note that the most discriminative regions are located ventrally. The lingual gyrus, not found in previous analyses, is thought to show decreased activation with age in memory tasks [15], a plausible finding during resting state. This would entail that the decrease in global efficiency is due to the decrease in activation.

The analysis in terms of combined global and local efficiency, where the importance of a region is given as the mean importance of $E_g$ and $E_l$ in that region, shows again that ventral areas (with the exception of the inferior frontal triangularis cortex) are more discriminative between young and old subjects, with limbic and frontal regions having more marked differences.

## IV. DISCUSSION

Comparing our results (Section III-C) with those obtained previously on the same dataset [10] we find numerous areas of agreement - the importance of the orbitofrontal regions, the importance of the amygdalae and the parahippocampal formation are also underlined in the present study under a predictive setting. These area are also well-known targets of age-related changes, thus it should come as no surprise that some of their connectivity properties show distinct differences.

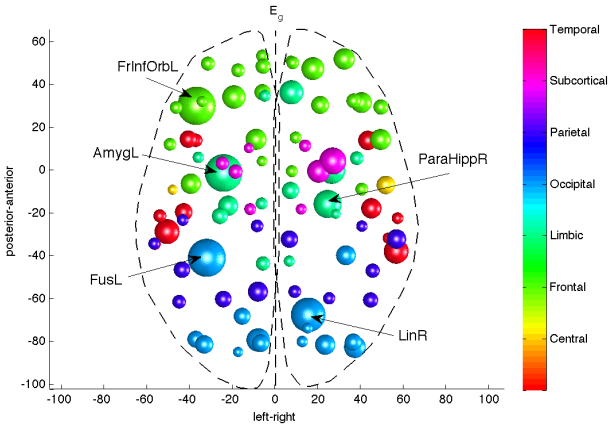


Figure 1. Importance of brain regions in discrimination between old and young in terms of global efficiency (linear SVM classifier). Larger spheres represent more importance in the discriminative function.

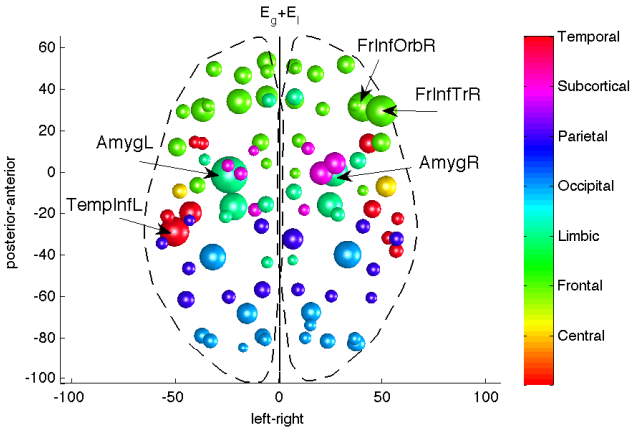


Figure 2. Importance of brain regions in discrimination between old and young in terms of combined global and local efficiency (linear SVM classifier).

However, care must be used when interpreting these "dense" maps: the regions are jointly discriminative, but not necessarily when taken in isolation, as is the case for univariate tests. Nevertheless, the good classification performance of classifiers using a very sparse subset of regions suggests that around 10-15% of brain regions carry the essential discriminative information.

An intriguing possibility for future work is to combine graph and vertex attributes, and to investigate more fully feature generation to avoid redundant features.

## Acknowledgments

This work was supported in part by the Swiss National Science Foundation (grant PP00P2-123438), and in part by the Center for Biomedical Imaging (CIBM) of the Geneva and Lausanne Universities, EPFL, and the Leenaards and Louis-Jeantet foundations. The computations of graph attributes were done using the R software *brainwaver* (http://cran.r-project.org/web/packages/brainwaver/). SA was partly founded by Réseau National des Systèmes Complexes (RNCS). SA and EB would like to thank the British council and the French Foreign Affairs Ministry for their support through the Alliance program.

## References

[1] K. Riesen and H. Bunke, "Approximate graph edit distance computation by means of bipartite graph matching," *Image and Vision Computing*, vol. 27, pp. 950–959, 2009.

[2] J. Richiardi, D. Van De Ville, K. Riesen, and H. Bunke, "Vector space embedding of undirected graphs with fixed-cardinality vertex sequences for classification," in *Proc. 20th Int. Conf. on Pattern Recognition (ICPR)*, 2010.

[3] R. Wilson, E. Hancock, and B. Luo, "Pattern vectors from algebraic graph theory," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 7, pp. 1112–1124, 2005.

[4] K. Riesen and H. Bunke, "Graph classification by means of lipschitz embedding," *IEEE Trans. on Man, Systems, and Cybernetics, part B*, vol. 39, no. 6, pp. 1472–1483, 2009.

[5] S. Achard, R. Salvador, B. Whitcher, J. Suckling, and E. Bullmore, "A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs," *Journal of Neuroscience*, vol. 26, no. 1, pp. 63–72, Jan. 2006.

[6] M. Lynall, D. S. Bassett, R. Kerwin, P. J. McKenna, M. Kitzbichler, U. Muller, and E. Bullmore, "Functional connectivity and brain networks in schizophrenia." *J Neurosci*, vol. 30, no. 28, pp. 9477–9487, Jul 2010.

[7] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot, "Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain," *NeuroImage*, vol. 15, pp. 273–289, 2002.

[8] P. Dickinson, H. Bunke, A. Dadej, and M. Kraetzl, "Matching graphs with unique node labels," *Pattern Analysis & Applications*, vol. 7, no. 3, pp. 243–254, Sep. 2004. [Online]. Available: http://dx.doi.org/10.1007/BF02683991

[9] V. Latora and M. Marchiori, "Economic small-world behavior in weighted networks," *European Physical Journal B*, vol. 32, pp. 249–263, 2003.

[10] S. Achard and E. Bullmore, "Efficiency and cost of economical human brain functional networks," *PLoS Computational Biology*, vol. 3, p. e17, 2007.

[11] A. Shokoufandeh and S. Dickinson, "A unified framework for indexing and matching hierarchical shape structures," in *Proc. 4th Int. Workshop on Visual Form (IWVF)*, C. Arcelli, L. Cordella, and G. di Baja, Eds., vol. 2059, Capri, Italy, May 2001, pp. 67–84.

[12] D. Meunier, S. Achard, A. Morcom, and E. Bullmore, "Age-related changes in modular organization of human brain functional networks." *Neuroimage*, vol. 44, no. 3, pp. 715–23, 2009.

[13] I. H. Witten and E. Frank, *Data Mining: Practical machine learning tools and techniques*, 2nd ed. Morgan Kaufman, 2005.

[14] J. Richiardi, H. Eryilmaz, S. Schwartz, P. Vuilleumier, and D. Van De Ville, "Decoding brain states from fMRI connectivity graphs," *NeuroImage (Special Issue on Multivariate Decoding and Brain Reading)*, 2010, (in press).

[15] W. E. Mencl, K. R. Pugh, S. E. Shaywitz, B. A. Shaywitz, R. K. Fulbright, R. T. Constable, P. Skudlarski, L. Katz, K. E. Marchione, C. Lacadie, and J. C. Gore, "Network analysis of brain activations in working memory: Behavior and age relationships," *Microscopy Research and Techique*, vol. 51, no. 1, pp. 64–74, 2000.