On the relevance of multi-graph matching for sulcal graphs

R Yadav, François-Xavier Dupé, S Takerkart, G Auzias

ON THE RELEVANCE OF MULTI-GRAPH MATCHING FOR SULCAL GRAPHS

R.Yadav[⋆‡†], *F.X. Dupé*[†], *S. Takerkart*[⋆], *G. Auzias*[⋆]

[⋆]Institut de Neurosciences de la Timone UMR 7289, Aix-Marseille Université, CNRS
[‡]Aix Marseille Université, Institut Marseille Imaging, Marseille, France
[†]Laboratoire d'Informatique et Systèmes UMR 7020, Aix-Marseille Université, CNRS

Fine-scale characterization of the geometry of the folding patterns of the brain is a key processing step in neuroscience, with high impact applications such as for uncovering biomarkers indicative of a neurological pathology. *Sulcal graphs* constitute relevant representations of the complex and variable geometry of the cortex of individual brains. Comparing sulcal graphs is challenging due to variations across subjects in the number of nodes, graph topology and attributes (on both nodes and edges). Graph matching experiments on real data are limited by the absence of ground truth. In this paper we propose to generate synthetic graphs to benchmark graph matching methods and assess their robustness to noise on attributes and to the presence of un-matchable nodes. Three multi-graph matching methods are compared to one pairwise approach in various simulation settings, showing that good matching performances can be achieved even with highly perturbed sulcal graphs. An experiment on real data from a population of 134 subjects further unveil large performance differences across matching methods.

1. INTRODUCTION

Recent studies [1, 2] proposed to represent biologically relevant morphological descriptors as *sulcal graphs* by decomposing the cortical surface in *sulcal basins* and *sulcal pits*. As illustrated on Fig.1 and described in details in [1], sulcal basins are defined as concavities in the white matter surface which are bounded by convex ridges, and the deepest point in each bassin defines the corresponding sulcal pit. A sulcal graph is constructed by considering each sulcal basin (or corresponding pit) as a node, while the edges connect adjacent basins and thus represent their spatial organization, as detailed in [2]. Various geometrical information of a sulcal basin can then be attributed to graph nodes (such as the depth of the pit, its 3d location...), while the spatial organization of the basins is integrated into the edges and topology of the graph. As illustrated on Fig.1, variations across individuals manifest in the graphs as changes in the number of nodes, graph topology and in the attributes attached to nodes and edges. Formally, a sulcal graph $\mathcal{G}$ is a triplet of vertices, edges and attributes, namely $\mathcal{G} = (V, E, A)$ with the cardinality of the graph being the number of basins also noted $|\mathcal{G}| = n$. In this work, we consider as attributes the 3d coordinates of the sulcal pits on the sphere for the nodes and the geodesic distance for the edges.

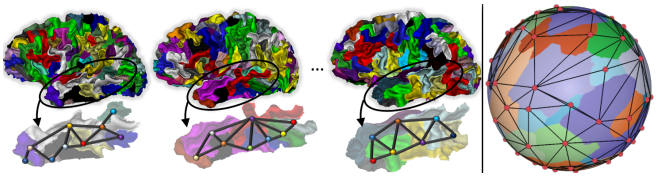

Fig. 1. Left: An illustration of the major variations across individuals in their respective sulcal graphs. Enlarged views of the temporal lobes show the variations in the number of nodes and graph adjacency. Right: The sulcal graph from each individual is mapped onto a common spherical domain which is convenient for comparing nodes coordinates.

Defining spatial correspondences across the brains of a set of individuals is required to compute statistics at the scale of a population. Comparing brains using sulcal graph matching is highly relevant because all the geometrical information is encoded in these graphs. However, matching such graphs is challenging due to the complexity of brain geometry that is preserved in graph representations. The specific challenges to be addressed in this context are: 1) the presence of noise due to imperfect sulcal basins segmentation, resulting from inaccurate extraction of the cortical mesh; 2) the lack of ground truth data at the scale of sulcal basins. Defining ground truth data is already a tedious and ambiguous task at the scale of sulci [3]. Indeed, at sub-lobar scale, the anatomical ambiguity becomes a problem for any human expert, such that no nomenclature or validated atlas exists at the scale of sulcal basins. All this makes the problem of matching a pair of sulcal graphs an ill-posed problem. Indeed, the conjunction of major variations across individuals and non-neglectable level of noise constitute clearly unfavorable settings for the pairwise graph matching techniques to perform well.

In our previous work [4], we introduced a procedure to generate a set of synthetic sulcal graphs representative of a population, that served to benchmark state of the art *pairwise matching techniques*. We also included in our benchmark

one multi-graph approach which outperformed all the pairwise techniques. This served as a proof of concept of the superior efficiency of multi-graph matching approaches in this context. By considering several brains together, the geometrical information that is shared by the majority of individuals helps to regularize the matching problem and allows to identify putative noisy graph nodes in a much more robust way than in the case of a pairwise matching between two subjects.

In the present work, we benchmark a selection of recently published *multi-graph matching techniques* against the best pairwise technique from [4], and show variations in performances that might impact significantly subsequent analyses. We also propose several improvements to the simulation framework that yield more realistic synthetic graphs. In a second experiment, qualitative and quantitative evaluations on real data from 134 subjects allow us to refine our conclusions. Our results demonstrate the feasibility to compare a large population of brains based on multi-graph matching of sulcal graphs in fully acceptable computing time.

2. MULTI-GRAPH MATCHING

Graph matching (GM) refers to the problem of finding correspondences between the nodes of two or more graphs. GM methods are usually divided into those that provide exact or inexact matchings. Exact methods require that the correspondences between nodes strictly preserve the adjacency of the graphs. But such an isometry is inappropriate for sulcal graphs exhibiting topological variations. Hence this paper focuses on inexact methods which tolerate structural variations between graphs by allowing for some flexibility during the matching process.

Most GM methods are limited to the pairwise setting where two graphs are to be matched. When matching multiple graphs $\{\mathcal{G}^1, \cdots, \mathcal{G}^N\}$, the standard solution consists in using a pairwise algorithm to find the matchings between all pairs of graphs. However, this often yields inconsistencies: combining the matchings of $\mathcal{G}^1$ to $\mathcal{G}^2$ and of $\mathcal{G}^2$ to $\mathcal{G}^3$ might not be equivalent to the direct matching of $\mathcal{G}^1$ to $\mathcal{G}^3$, especially in the presence of noise. To overcome this issue, **mSync** [5] enforces a cycle consistency constraint between permutation matrices, which is defined as:

$$\forall i, j, k \in \{1, \ldots, N\} \quad \mathbf{X}_{ij} = \mathbf{X}_{ik}\mathbf{X}_{kj} \tag{1}$$

where $\mathbf{X}_{ij}$ is the permutation matrix from $\mathcal{G}^i$ to $\mathcal{G}^j$. The multi-matching solution is then obtained by solving Koopmans-Beckmann's quadratic assignment problem [6, 7, 8] using bulk matrices:

$$\max_{\mathbf{X} \in \mathcal{C}} \; tr(\mathbf{A}\mathbf{X}\mathbf{A}\mathbf{X}^\top) \tag{2}$$

where $\mathbf{X} = \{\mathbf{X}_{ij}\}_{1 \leq i,j \leq N}$ is the bulk matrix containing all permutations matrices and identity matrices as block diagonal, $\mathbf{A} = \{\mathbf{A}_{ij}\}_{1 \leq i,j \leq N}$ is the bulk matrix with the adjacency matrices and $\mathcal{C}$ is the set of constraints which at least include the consistency expressed in (1) and permutations requirement on $\{\mathbf{X}_{ij}\}_{1 \leq i,j \leq N}$.

Solving (2) is an NP-Hard problem since $\mathcal{C}$ is a non-convex set and consistency is difficult to enforce. Thus, **mSync** models consistency by introducing a virtual "universe" of nodes, which induces a specific permutation matrix for each graph. These matrices can be combined to produce the permutations between two graphs. This universe of node is computed via a synchronisation mechanism. In contrast, **mALS** [6] imposes a low-rank constraint on the bulk permutation matrix under the assumption that only few nodes are different between graphs.

The notion of universe of nodes is particularly interesting in the case of noisy graphs with different number of nodes. This way a consistent matching should satisfy $\mathbf{X}_{ij} = \mathbf{U}_i\mathbf{U}_j^T$, where $\mathbf{U}_i \in \{0,1\}^{n_i,k}$ represents the map from $\mathcal{G}^i$ to the universe, where k is the number of points in the universe, n_i is the size of $\mathcal{G}^i$ and $k \geq n_i$. The dimension k is directly linked to the rank of $\mathbf{X}$, meaning that **mALS** implicitly relies on such representation. On the contrary, **HiPPI** [7] parametrizes pairwise matchings in the universe of nodes instead of directly optimising on $\mathbf{X}$ with additional constraints. This method takes into account both cycle and graph attributes consistencies by solving (2) using a projected power method with an appropriate initialization.

While both **HiPPI** and **mALS** manipulate a low-rank version of $\mathbf{X}$, there is a big difference in their universe of nodes. $\mathbf{U}_i$ are permutation matrices, while a simple low-rank constraint is more general. **mALS** then allows more flexibility in the embedding of nodes with a threshold on the values to avoid bad matching, but it is implicit. As a consequence **HiPPI** formally offers a better enforcement of the cycle consistency than **mALS** and is therefore closer to **mSync**. We did also consider greedy methods such as CAO [9] but their higher complexity induced prohibitive computing time in our context.

3. EXPERIMENTS AND RESULTS

We now describe our experiments that aim at benchmarking multi-graph matching algorithms on population-wise sets of sulcal graphs. We work both on a real dataset, but also on sets of artificially generated sulcal graphs for which we know the ground-truth matching, which allows studying the robustness of these algorithms with respect to several types of variations present in the real graphs.

3.1. Generating realistic synthetic sulcal graphs

In order to generate realistic sulcal graphs, we benefit from the recent work [10] where an extensive description of a population of real sulcal graphs is provided: i) the average number of nodes across the sulcal graphs representative of a population of healthy adults is 85 ± 4; but ii) the labelling process

introduced in that work resulted in the identification of only 72 of these nodes, the less variable ones, leaving almost 20% of unlabelled nodes.

Based on these information, we first generate the nodes of a *reference sulcal graph* by randomly sampling 72 points on a sphere to match the number of the most reproducible nodes in the real data. Next, we simulate the variations across individuals in the location of the nodes and in the structure of the sulcal graphs using the four-steps process described below, which we repeat l times to generate the sulcal graphs of a population of l individuals: 1) We start by perturbing the coordinates of each reference node. Being on the sphere, we use the *Von-Mises Fisher distribution* $f_v(\mu, \kappa)$ to model the noise on the coordinates of the nodes. μ denotes the mean direction and κ is the concentration parameter (μ and $1/\kappa$ can be seen as homologous to μ and σ^2 for the Normal distribution); note that the amount of noise increases when κ decreases. 2) We draw O additional random points on the sphere, in order to simulate the unlabelled "outlier" nodes of the real sulcal graphs. 3) We compute the 3d convex hull of these 72+O points, which provides the edges of the graph and corresponding adjacency matrix in similar way as in the real graph since the edges between the nodes represent the neighbourhood relationship between sulcal basin (see [2] for precision). 4) Finally, we randomly drop 10% of these edges in order to best match the node degree distribution of real sulcal graphs.

Note that the computation of the convex hull for obtaining the edges comes after the spatial perturbation of reference nodes but also after the inclusion of outlier nodes. This induces important variations in the topology of the graphs in addition to the random drop of edges, as illustrated on Fig.2. Using this procedure, the amount of noise in our collection of l artificial sulcal graphs (i.e in our population of l subjects) is controlled by two parameters, κ and O.

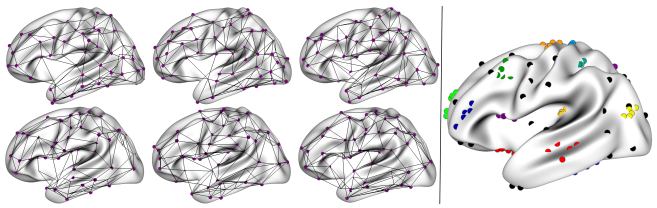

Fig. 2. Left: sulcal graphs mapped onto an average brain surface, showing variations in the number of nodes and the graph topology; top: real data from three individuals; bottom: three synthetic graphs. Right: visualization of the spatial dispersion of the nodes across a simulated population of $l = 5$ graphs with $\kappa = 200$ and $O = 8$. The ground truth matching is indicated by the color of the nodes, and outlier nodes are shown in black.

3.2. Results on synthetic sulcal graphs

We generate 10 populations of $l = 134$ synthetic graphs for each set of parameters. We explored the following sets of simulation parameters $\kappa \in [200, 800, 1400]$ and number of outlier nodes $O \in [0, 6, 12, 18]$. In the graph matching literature, most methods make use of an affinity matrix $\mathbf{K} \in \mathbb{R}^{n_1 n_2 \times n_1 n_2}$ that holds the node-to-node and edge-to-edge similarity values between any two graphs $\mathcal{G}^1$ and $\mathcal{G}^2$, where n_1 and n_2 are the number of nodes in each graph. The diagonal of $\mathbf{K}$ represents node-to-node similarities whereas the off-diagonal holds edge-to-edge similarities. In this paper, we compute the values in $\mathbf{K}$ using the Gaussian kernel on the squared norm of two vectors of node and edge attributes, as described in [2]. We then use $\mathbf{K}$ as weighted adjacency matrix for multi-graph methods described in section 2. We initialize each multi-graph method with the result of the pairwise matching obtained from **KerGM**. We run and evaluate the algorithms for each of the 10 populations independently, allowing to compute the average and standard deviation of our evaluation criterion, the standard F_1-score $\in$ [0,1], which is a function of the $precision$ and $recall$:

$$F_1 = 2 * (precision * recall)/precision + recall \quad (3)$$

where, $precision = \frac{\text{number of correct matches by the algorithm}}{\text{Total number of ground-truth matches}}$

$recall = \frac{\text{number of correct matches by the algorithm}}{\text{Total number of matches by the algorithm}}$

Indeed, for our sulcal graph matching task, the F_1-score is relevant to detect incorrect matching alongside matches with outliers. A F_1 score of 1 reflects the ability of the algorithm to obtain a perfect matching of inlier nodes and accurate identification of outlier nodes. Fig.3 shows the F_1-score obtained for different values of κ, fixing the number of outliers $O = 18$.

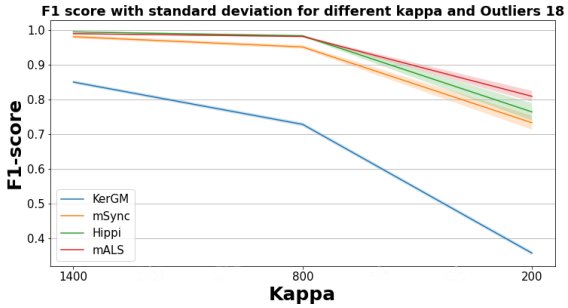

Fig. 3. Mean and standard deviation of F_1-score for each method, with $\kappa \in [1400, 800, 200]$ and the number of outliers O=18.

From this figure and the other parameter settings not shown, we observe that all the multi-graph matching methods (**mALS, mSync** and **Hippi**) outperform the pairwise matching method **KerGM** which was the best among all pairwise methods in [4]. This confirms that pairwise matching is not

competitive in our application where matching information is spread across multiple graphs. The performances of all the algorithms deteriorate as kappa decreases (higher perturbation). Although the three multi-graph methods seems to perform well when perturbations are strong, **mALS** is slightly better than **mSync** and **Hippi**. However the computation time of **mALS** is much higher than for the other methods (see Table.1). When varying the number of outlier nodes (not shown), we observe also that **Hippi** and **mSync** are slightly more prone to incorrect matching compared to **mALS**, that is less strict in enforcing the consistency constraint.

3.3. Application to real data

For the evaluation on real data, we work with the sulcal graphs from 134 young healthy adults taken from the open OASIS database [11]. The cortical meshes and sulcal graphs were obtained following the procedure described in [1, 2]. In absence of ground truth matching, we cannot compute the F_1-score. Instead, we compute the matching consistency of each node of each graph as proposed in e.g. [9, 4], and report the average across all graphs and nodes. The matching consistency measures the uniqueness of the pairwise matchings by different composition orders and varies between 0 and 1. A value of 1 corresponds to the ideal case where each graph only contain nodes that are matchable.

Then, we examine the characteristics of node clusters, where a cluster is defined as the set of nodes that share the same label across all graphs, taking the largest graph as a reference. For this, we compute the silhouette coefficient of each cluster, which is a good indicator of both the concentration of nodes within a cluster and the separation between clusters. It is the average within a cluster of the silhouette coefficients of all its nodes, which is given by the ratio between the average Euclidean distance to the other nodes in the cluster and its distance to other nearby clusters [12].

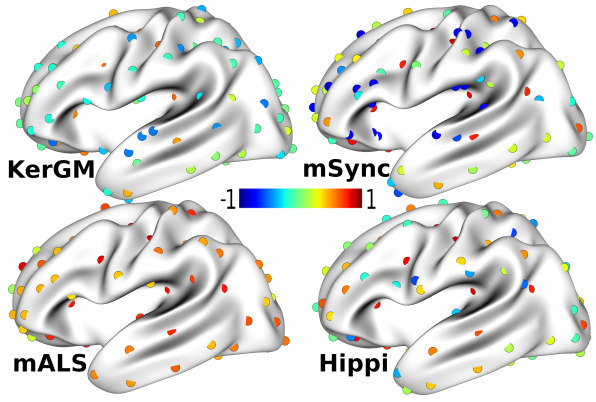

Fig. 4. For each method, the silhouette coefficient of each cluster is indicated by the color of the circle. The location of each circle corresponds to the coordinates of the centroid of each cluster.

On Fig.4, the silhouette coefficient of each cluster is indicated by the color of the circle, and the location of each circle corresponds to the coordinates of the centroid of each cluster, allowing to compare spatial patterns across methods. Together with Table 1, this figure illustrates the high spatial dispersion of nodes corresponding to each cluster with **KerGM**, associated to very low silhouette coefficients. The method **mSync** results in higher silhouette coefficients for some nodes, but lower value for others, indicating that the matching was enforced also for ambiguous nodes located in highly variable regions. This is expected since the solution of **mSync** is a matching that is consistent across all graphs by construction, every nodes being matched across all graphs. For **mALS**, the silhouette coefficients are very high but the number of unmatched nodes also, indicating that this method was much more restrictive in the matching, leaving more than 27% of the nodes unmatched. The results of **Hippi** are balanced, resulting in high silhouette coefficients in less variable regions where **mSync** and **mALS** did also perform well, and lower coefficients in regions known to be variable, still enforcing the matching for almost 98% of the nodes. Across all methods, the clusters with a high silhouette coefficient are consistently located in regions such as the central and pre-central sulci and the superior temporal region, which are known to be less variable across individuals than e.g. inferior temporal regions. This pattern supports the biological relevance of the correspondences across individual graphs resulting from multi-graph matching techniques.

Table 1. Quantitative comparison on real data.

Method	Num. labels	silhouette	Perc. unmatched	consistency	cpu time (min)
mALS	82	0.54 ± 0.23	27.7	0.91 ± 0.08	783.81
Hippi	92	0.20 ± 0.42	2.1	0.95 ± 0.04	54.23
mSync	101	0.05 ± 0.52	0	1.0 ± 0.0	25.54
KerGM	101	-0.05 ± 0.31	0	0.30 ± 0.17	1362.72

4. CONCLUSION

In this paper, we apply mutli-graph matching to define population-wise correspondences across individual cortical geometries. Our experiments on both synthetic and real data demonstrate the benefits of multi-graph matching techniques compared to pairwise approaches. Furthermore, clear differences emerge between the different state-of-the-art multi-graph matching algorithms that we benchmarked. Only **mALS** and **HiPPI** provide matchings that are biologically plausible, whereas in this context **mSync** is hampered by algorithmic constraints. Finally, the superior scalability offered by **HiPPI** makes it the best candidate method for future large-scale studies of brain anatomy based on sulcal graphs.

5. REFERENCES

[1] G. Auzias, L. Brun, C. Deruelle, and O. Coulon, "Deep sulcal landmarks: Algorithmic and conceptual improvements in the definition and extraction of sulcal pits," *NeuroImage*, vol. 111, pp. 12–25, May 2015.

[2] Sylvain Takerkart, Guillaume Auzias, Lucile Brun, and Olivier Coulon, "Structural graph-based morphometry: A multiscale searchlight framework based on sulcal pits," *Medical Image Analysis*, vol. 35, pp. 32–45, Jan. 2017.

[3] Léonie Borne, Denis Rivière, Martial Mancip, and Jean-François Mangin, "Automatic labeling of cortical sulci using patch-or cnn-based segmentation techniques combined with bottom-up geometric constraints," *Medical Image Analysis*, vol. 62, pp. 101651, 2020.

[4] N. Buskulic, F.X. Dupé, S. Takerkart, and G. Auzias, "Labelling sulcal graphs across indiviuals using multi-graph matching," in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, 2021, pp. 1486–1490.

[5] Deepti Pachauri, Risi Kondor, and Vikas Singh, "Solving the multi-way matching problem by permutation synchronization," in *Advances in neural information processing systems*, 2013, pp. 1860–1868.

[6] Xiaowei Zhou, Menglong Zhu, and Kostas Daniilidis, "Multi-image matching via fast alternating minimization," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 4032–4040.

[7] Florian Bernard, Johan Thunberg, Paul Swoboda, and Christian Theobalt, "Hippi: Higher-order projected power iterations for scalable multi-matching," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 10284–10293.

[8] Roberto Tron, Xiaowei Zhou, Carlos Esteves, and Kostas Daniilidis, "Fast Multi-image Matching via Density-Based Clustering," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Oct. 2017, pp. 4077–4086, IEEE.

[9] Junchi Yan, Minsu Cho, Hongyuan Zha, Xiaokang Yang, and Stephen M Chu, "Multi-graph matching via affinity optimization with graduated consistency regularization," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 6, pp. 1228–1242, 2015.

[10] Irene Kaltenmark, Christine Deruelle, Lucile Brun, Julien Lefèvre, Olivier Coulon, and Guillaume Auzias, "Cortical inter-subject correspondences with optimal group-wise parcellation and sulcal pits labeling," *Medical Image Analysis*, 2020.

[11] Daniel S Marcus, Tracy H Wang, Jamie Parker, John G Csernansky, John C Morris, and Randy L Buckner, "Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented older adults," *Journal of cognitive neuroscience*, vol. 19, no. 9, pp. 1498–1507, 2007.

[12] Peter J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, Nov. 1987.