



HAL
open science

Population-wise Labeling of Sulcal Graphs using Multi-graph Matching

Rohit Yadav, François-Xavier Dupé, S. Takerkart, Guillaume Auzias

► **To cite this version:**

Rohit Yadav, François-Xavier Dupé, S. Takerkart, Guillaume Auzias. Population-wise Labeling of Sulcal Graphs using Multi-graph Matching. PLoS ONE, 2023, 10.1371/journal.pone.0293886 . hal-03963359v3

HAL Id: hal-03963359

<https://hal.science/hal-03963359v3>

Submitted on 7 Feb 2024

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.

RESEARCH ARTICLE

Population-wise labeling of sulcal graphs using multi-graph matching

Rohit Yadav^{1,2,3*}, François-Xavier Dupé³, Sylvain Takerkart¹, Guillaume Auzias¹

1 Institut de Neurosciences de la Timone UMR 7289, CNRS, Aix-Marseille Université, Marseille, France, **2** Institut Marseille Imaging, Aix Marseille Université, Marseille, France, **3** Laboratoire d'Informatique et Systèmes UMR 7020, CNRS, Aix-Marseille Université, Marseille, France

* rohit.yadav@univ-amu.fr



Abstract

Population-wise matching of the cortical folds is necessary to compute statistics, a required step for e.g. identifying biomarkers of neurological or psychiatric disorders. The difficulty arises from the massive inter-individual variations in the morphology and spatial organization of the folds. The task is challenging both methodologically and conceptually. In the widely used registration-based techniques, these variations are considered as noise and the matching of folds is only implicit. Alternative approaches are based on the extraction and explicit identification of the cortical folds. In particular, representing cortical folding patterns as graphs of sulcal basins—termed *sulcal graphs*—enables to formalize the task as a graph-matching problem. In this paper, we propose to address the problem of sulcal graph matching directly at the population level using multi-graph matching techniques. First, we motivate the relevance of the multi-graph matching framework in this context. We then present a procedure for generating populations of artificial sulcal graphs, which allows us to benchmark several state-of-the-art multi-graph matching methods. Our results on both artificial and real data demonstrate the effectiveness of multi-graph matching techniques in obtaining a population-wise consistent labeling of cortical folds at the sulcal basin level.

OPEN ACCESS

Citation: Yadav R, Dupé F-X, Takerkart S, Auzias G (2023) Population-wise labeling of sulcal graphs using multi-graph matching. PLoS ONE 18(11): e0293886. <https://doi.org/10.1371/journal.pone.0293886>

Editor: Xiao Luo, University of California Los Angeles, UNITED STATES

Received: August 21, 2023

Accepted: October 23, 2023

Published: November 9, 2023

Copyright: © 2023 Yadav et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All the source code is shared openly at https://www.github.com/gauzias/sulcal_graphs_matching. The MRI data used in this work were introduced in "Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults" and are available at <http://oasis-brains.org>.

Funding: This study was supported by Aix-Marseille Université (AMX-19-IET-002), awarded to RY, Agence Nationale de la Recherche (ANR-19-CE45-0014, ANR-21-NEU2-0005), awarded to GA.

1 Introduction

1.1 Quantitative comparison across brains is a crucial but open question

Comparing features extracted from brain MRI across individuals is necessary for estimating population statistics and ultimately discover markers of diseases. However, this task presents several challenges at both the methodological and conceptual levels. Indeed, the features extracted from two different individual brains are defined in two different spaces. Comparing such features thus requires to address the *methodological* problem of transferring them into a common space. The task of transferring information from different brains to a common space consists in defining spatial correspondences across these objects by compensating for the variations in their respective geometry. The challenge arises from the massive inter-individual variations in the geometry of the cortical surface, which make the identification of such spatial correspondences an ill-posed problem. Consequently, any solution to this problem inevitably requires the introduction of additional constraints based on assumptions about the biological

Competing interests: The authors have declared that no competing interests exist.

validity of the resulting spatial correspondences, which constitutes a challenge at the *conceptual* level. Indeed, the assumptions and constraints introduced in the definition of the spatial correspondences actually influence the derived statistics measured on the population of interest, and could therefore be considered as a source of bias in population-wise analyses [1].

One widely used type of approach to tackle this problem—termed here as the registration-based approach—consists in defining a mapping between each individual brain and an atlas serving as the common space by estimating a spatial transformation. See for instance [2–5] for examples of such approaches. As pointed above, the process of building the atlas and defining the associated projection operator which minimizes the error induced by the transformation remains an open research question. As a consequence, several registration techniques and atlases co-exist in the field [6, 7]. The variety of atlases, projection mechanisms and descriptors illustrate the ongoing exploration of putative biologically relevant features used to define these correspondences across individuals. One of the most widely used registration-based approaches [3] defines a mapping between cortical surfaces by imposing the alignment of a combination of curvature and convexity features estimated from a 2D mesh representing the geometry of the cortex. The cortical surface of a given subject is projected onto the atlas by matching its curvature and convexity, under the assumption that aligning these features induces biologically relevant anatomic-functional correspondences. In this process, as in any registration-based approach, variations across individuals are considered as noise or confounding perturbations to be minimized, including variations in the topology and number of folds (sulci). More generally, the registration-based approach could be seen as an oversimplification of the problem, as it does not take into account potentially relevant geometric information.

Alternative approaches consist in characterizing the geometry and organization of the cortical folds in each individual and then compare these features across the population.

1.2 Characterizing cortical folding patterns using graphs

Several approaches have been proposed to characterize cortical folding patterns, such as gyrification index, fractal dimension and curvature [8–10]. Although these measures capture relevant morphological features, they do not explicitly reflect the topology, i.e. the spatial relationships between sulci. [11] introduced an analysis framework based on the automatic extraction and labeling of the sulci allowing the characterization of their shape in terms of e.g. sulcus area, depth and length, but also their spatial pattern. This representation of the cortical geometry has been used for instance to characterize populations of healthy subjects [12], to quantify potential deviations from normal populations in various conditions such as schizophrenia [9] and autism spectrum disorder [13], or to estimate the heritability of the folding patterns [14]. Pursuing on this line of research, the *sulcal pits* were introduced as a concept allowing to decompose the sulci into smaller pieces and thus access finer scale geometrical information. As described in details in [15, 16], each fold is divided into *sulcal basins* that are defined as concavities in the white matter surface bounded by convex ridges, and the deepest point in each basin defines the associated sulcal pit. More recently, [17, 18] represented the geometrical relationships between sulcal basins as a *sulcal graph*. A sulcal graph is constructed by considering each sulcal basin (or associated pit) as a node, while the edges connect only adjacent basins and thus represent their spatial organization. Various geometrical features of a sulcal basin can then be attributed to graph nodes (such as the depth of the pit, its 3d coordinates. . .), while the spatial organization of the basins is encoded in the topology of the graph. Fig 1 illustrates this decomposition of the cortical folds into sulcal basins allowing to represent this complex geometry as a sulcal graph.

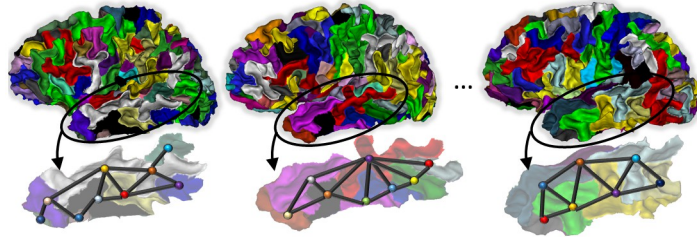


Fig 1. Example of sulcal graphs from three individual brains, superimposed with the underlying decomposition of the cortical surface in sulcal basins. Sulcal basins are shown in different colours, and their corresponding node in the graph are represented as spherical dots in the lower panel. The color of each node in the graph illustrates the value of a given attribute such as for instance the area or depth of corresponding sulcal basin.

<https://doi.org/10.1371/journal.pone.0293886.g001>

These sulcal graphs constitute particularly relevant representations because: 1) variations across individuals are preserved and are manifested as changes in both the topology of the graph and the value of the attributes attached to the nodes and edges; 2) the design of tools for the quantitative characterization of these variations can benefit from the extensive body of methods from the graph processing literature.

1.3 Problem statement and contributions

In the present work, we focus on the task of *matching together a set of sulcal graphs* in order to define *correspondences across a population of subjects*, under the specific constraint of explicitly taking into account the variations in *folding patterns*. Before moving to the formalization, we more precisely situate this problem with respect to the conceptual question of defining correspondences across sulcal graphs from different individuals, and with respect to the methodological problem of graph matching.

1.3.1 Unsupervised comparison and matching of sulcal graphs. The use of sulcal graphs to define correspondences across brains is highly relevant because all the geometrical information about the macroscopic cortical folding can be encoded in such graphs. However, several challenges need to be addressed in this context: 1) the large inter-individual variations in brain anatomy induce complex variations across sulcal graphs, including in their topology; 2) sulcal graphs can be contaminated by noise resulting from the imperfect segmentation of the individual cortical surface and corresponding sulcal basins; 3) there is no consensus on a nomenclature or atlas at the scale of sulcal basins covering the whole brain, that is a prerequisite to tackle the matching problem as a supervised learning task. Indeed, few studies investigated the matching of cortical folds across individuals as a supervised task [19–21]. All these works focused on the scale of sulci, i.e. considering large folds consisting of several of our sulcal basins. To our knowledge, only [22] attempted to tackle this problem at finer scale, probably because of the massive amount of efforts needed to gather sufficient amount of manually labeled data [23]. Indeed, ambiguities due to variations across individuals in the folding patterns become overwhelming at finer scale than sulci. This is illustrated by the tedious works advancing the definition of a fined-grained nomenclature of folds [24] and their relationship with underlying function [25]. The lack of widely accepted fined-grained nomenclature is also blatant in the related field of brain parcellation: more than 20 different fine-grained atlases co-exist [26], and even the most advanced multi-modal atlas [27] was validated only on a small portion of the cortex.

Matching sulcal graphs across individuals is thus a very challenging problem. Instead of relying on the few existing labeled data-sets that clearly deserve further validation, we decided to approach this question as an unsupervised learning task.

We now describe the few studies that have attempted to tackle the question of unsupervised labeling of sulcal graphs. The first approach was proposed by [15] and corresponds to the baseline in brain imaging field. Indeed graph matching is not the main approach to match sulcal patterns. As mentioned in sec.1.1, spherical registration techniques such as [3] implemented in Freesurfer are traditionally applied to warp cortical surface and indirectly match sulcal patterns. The approach from [15] consists in relying on such spherical registration to compute a map of the spatial density of sulcal pits across a population of subjects. The density map is computed by accumulating the pits from the different individuals in each vertex of an average surface. Thanks to the alignment of folds resulting from the registration, the density map shows spatial patterns following the major cortical folds that were consistently matched across individuals, with regions of higher density in deeper, more stable sulci. A watershed algorithm is then applied to this density map in order to separate the main *clusters of sulcal pits*, empirically defined as the regions of high density. An arbitrary label is then associated to each cluster, hereby defining an *ad-hoc* labeling of the pits across individuals, depending on the cluster to which they contributed in the density map. This procedure implicitly defines a matching of sulcal pits and corresponding basins across individuals. Exemplar applications of this method can be found in e.g. [15, 16, 28], with illustrations of density maps and induced labeling for various populations. We refer in the following to this category of methods as **Auzias et al.** since we used the open source implementation from that paper. The main limitation of this approach is that the labeling is driven only by the coordinates of the sulcal pit.

[29] introduced an alternative procedure for labeling the sulcal basins, hereby considering the geometry of the basin surrounding each sulcal pit in addition to its spatial location. We refer to this method as **Kaltenmark et al.** in the following. The authors of [29] also raised the question of the *consistency* of the labeling, a notion that we will develop further below. In this method, an explicit constraint is imposed to restrict the labeling to only one node per subject for each label. In addition, the nodes for which the labeling is ambiguous—i.e. for which several labels are equally plausible—remain unlabelled, which is often denoted as *partial matching* in the literature on graph processing. Importantly, the spatial relationship between adjacent sulcal basins and pits is not taken into account in any of these methods, since the different pits/basins from each subject are considered independently. In contrast, in the present work our aim is to exploit the spatial organization of the adjacent basins stored in the sulcal graph representation.

Few publications investigated the potential of graph matching in the context of sulcal graphs. In [17], the spectral graph matching technique [30] was applied to a set of 48 monozygotic twins, comparing a pair at a time. This study showed that the similarity of the sulcal graphs across pairs of twins are higher than for unrelated pairs, demonstrating the genetic influence on sulcal patterns, and the relevance of graph matching techniques in this context. This approach was used in follow-up papers from the same group, e.g for comparing brain lobes in [31] or for matching individuals onto an atlas in [32].

In [33], a population of 677 neonates was analyzed based on a sulcal graph comparison method similar to the one from [17]. The authors proposed to use different features of the sulcal basins such as the pit position, the pit depth, the basin area, the basin boundary and the local connectivity of the graph to construct different similarity matrices, one per feature. The similarity matrices were then merged using a matrix fusion technique [34]. A clustering algorithm was then applied to the fused similarity matrix to identify sub-populations of sulcal graphs, associated to specific folding patterns in the central, cingulate and superior temporal regions.

Critically, all these studies relied only on *pairwise* graph matching techniques. Comparing a population of graphs by pairs, in the presence of noise and large inter-individual variations, is clearly sub-optimal.

1.3.2 Multi-graph matching: A relevant framework for population studies. Given the large variations across subjects and imperfect sulcal basin extraction, examining jointly a group of sulcal graphs is key to reveal meaningful information not accessible by considering only pairs of subjects. This is the translation to sulcal graphs of the basic idea behind general population studies, that allowed researchers to uncover some of the mechanisms underlying the anatomo-functional organization of the brain. We follow this principle by investigating for the first time the potential of *multi-graph* matching techniques in the context of sulcal graphs. By considering several brains together, the geometrical information that is shared by the majority of individuals should help to regularize the matching problem and allow to identify putative noisy graph nodes in a more robust way than with pairwise matching. The multi-graph matching framework has the potential to uncover population-wise invariant patterns in sulcal graphs without imposing a priori, potentially biasing, assumptions.

1.3.3 Contributions. In our previous work [35], we introduced a framework to generate a set of synthetic sulcal graphs representative of a population, and used it to benchmark state of the art *pairwise* matching techniques in this context. In [36], we provided a proof of concept of the relevance of multi-graph matching techniques. In the present study, we extend these preliminary studies in several directions.

First, we introduce an improved simulation framework to generate populations of artificial sulcal graphs and demonstrate their biological plausibility through a quantitative comparison with real data. Second, we benchmark a selection of recently published multi-graph matching techniques against the best pairwise technique for this task (identified in from [35]), and report variations in performance that would clearly impact potential real-world applications, e.g in a clinical context. Third, we compare qualitatively and quantitatively the different graph matching techniques, as well as the previously published approaches Auzias et al. and Kaltenmark et al, on a real data-set of 137 subjects. Finally, we report an exemplar application of the multi-graph matching framework by assessing potential statistical differences in the depth of matched sulcal basins between a group of men and a group of women. Overall, our experiments demonstrate the feasibility of comparing a large population of sulcal graphs based on multi-graph matching techniques, in fully acceptable computing time.

2 Formal problem and state of the art

In this section, we define formally the problem of matching sulcal graphs, as well as the multi-graph framework. We then give an overview of the different methods proposed in the literature and provide a more detailed description of the multi-graph matching methods included in our experiments.

2.1 Undirected attributed sulcal graphs

We consider a population of N sulcal graphs, noted $\mathcal{G}_1 \dots \mathcal{G}_N$, representing the cortical folding pattern of an hemisphere from N different individuals. The sulcal graph from a given subject q is an undirected attributed graph formally defined as a quadruplet $\mathcal{G}_q = (V_q, E_q, A_q^V, A_q^E)$, where $V_q = \{v_1, v_2, \dots, v_{n_q}\}$ are the nodes in the graph and $|\mathcal{G}_q| = n_q$ is the number of nodes. $E_q \subseteq V_q \times V_q$ defines the set of e_q edges. $A_q^V = \{a_{v_1}^V, a_{v_2}^V, \dots, a_{v_{n_q}}^V\}$ is the set of attributes associated to each node in V_q , and $A_q^E = \{a_{e_1}^E, a_{e_2}^E, \dots, a_{e_{e_q}}^E\}$ is the set of attributes associated with each edge in E_q .

Note that the number of nodes n_q and edges e_q and corresponding attributes varies across graphs. As illustrated on Fig 2, the sulcal graph from each subject is then mapped onto the same common spherical domain using the surface inflation and registration tools from freesurfer v.5.1.0 (<https://surfer.nmr.mgh.harvard.edu/>, see [3] for details). The matching is computed in this common

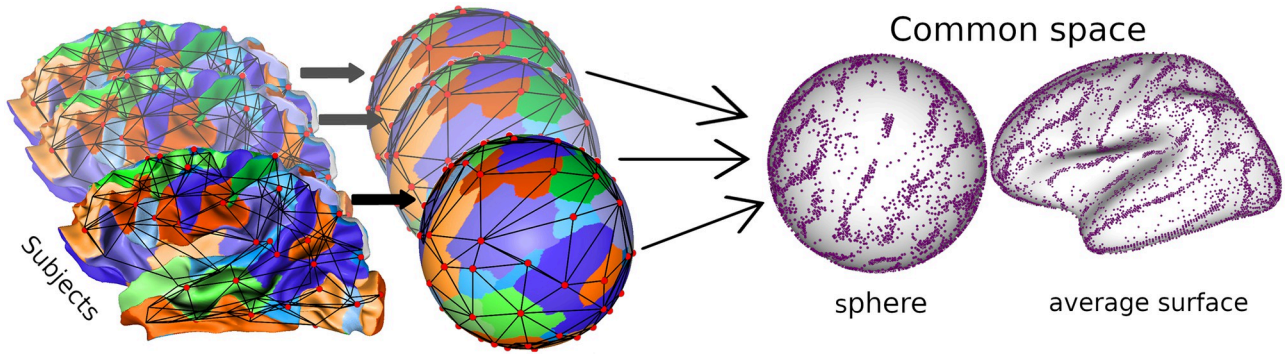


Fig 2. The sulcal graph from each subject is transferred onto a common sphere using the inflation and spherical registration tools from *freesurfer*. The sulcal graphs from every subjects can then be mapped onto either the common sphere or onto an average surface for visualization. Note that the spatial dispersion of the nodes of the graphs on the common spaces is heterogeneous, with dense clusters in cortical regions where the variations across individuals are lower.

<https://doi.org/10.1371/journal.pone.0293886.g002>

spherical domain. In this work, we consider as attributes of the nodes the 3D coordinates of the sulcal pits on the sphere. Regarding the attributes of the edges, we compute the length of the edge on the sphere as an approximation of the geodesic distance between neighboring pits.

2.2 Generalities and overview of pairwise graph matching methods

Pairwise graph matching refers to the problem of finding correspondences between the nodes of two graphs \mathcal{G}_1 and \mathcal{G}_2 . This problem can be formulated as a Quadratic Assignment Problem (QAP) [37]. Although different forms of QAP exist, the vast majority of the literature has focused on Lawler’s QAP [38]. Given two graphs \mathcal{G}_1 and \mathcal{G}_2 with number of nodes $|\mathcal{G}_1| = n_1$ and $|\mathcal{G}_2| = n_2$ respectively, the Lawler’s QAP consists in searching for the *assignment matrix* $\mathbf{X}_{12} \in \{0, 1\}^{n_1 \times n_2}$ such that $\mathbf{X}_{12}[i, j] = 1$ indicates that $v_i \in V_1$ corresponds to $v_j \in V_2$ and $\mathbf{X}_{12}[i, j] = 0$ otherwise, resulting from the following optimization problem:

$$\begin{aligned} \max J(\mathbf{X}_{12}) &= \text{vec}(\mathbf{X}_{12})^\top \Phi_{12} \text{vec}(\mathbf{X}_{12}) , \\ \text{subject to } \mathbf{X}_{12} \mathbf{1}_{n_2} &= \mathbf{1}_{n_1}, \mathbf{X}_{12}^\top \mathbf{1}_{n_1} \leq \mathbf{1}_{n_2}, \mathbf{X}_{12} \in \{0, 1\}^{n_1 \times n_2} , \end{aligned} \tag{1}$$

where $\text{vec}(\mathbf{X}_{12})$ denotes the column wise vectorization of \mathbf{X}_{12} ; $\mathbf{1}_{n_1}$ and $\mathbf{1}_{n_2}$ denote the column vectors of all ones of size n_1 and n_2 ; and $\Phi_{12} \in [0, 1]^{n_1 n_2 \times n_1 n_2}$ is the *affinity matrix* that is given as an input. The diagonal entries of Φ_{12} encode the similarity across nodes whereas non-diagonal entries encode the similarity across edges between the two graphs. The computation of the affinity matrix is context-dependent, and we detail the approach used in the present work in section 4.1.

The computation and storage in memory of the very large matrix Φ_{12} impedes the scalability of the matching problem based on this formulation. A solution to tackle this limitation is to reformulate the matching as a Koopmans-Beckmann’s problem [39] that is a special case of Lawler’s QAP:

$$\begin{aligned} \max J(\mathbf{X}_{12}) &= \text{tr}(\Psi_{12}^\top \mathbf{X}_{12}) + \text{tr}(\mathbf{A}_1 \mathbf{X}_{12} \mathbf{A}_2 \mathbf{X}_{12}^\top) , \\ \text{subject to } \mathbf{X}_{12} \mathbf{1}_{n_2} &= \mathbf{1}_{n_1}, \mathbf{X}_{12}^\top \mathbf{1}_{n_1} \leq \mathbf{1}_{n_2}, \mathbf{X}_{12} \in \{0, 1\}^{n_1 \times n_2} , \end{aligned} \tag{2}$$

where $\Psi_{12} \in [0, 1]^{n_1 \times n_2}$ denotes the affinity matrix *across nodes*, and $\mathbf{A}_1 \in \mathbb{R}^{n_1 \times n_1}$ and $\mathbf{A}_2 \in \mathbb{R}^{n_2 \times n_2}$ are the weighted adjacency matrices of two graphs respectively such that $\mathbf{A}[i, j] = w_{ij}$ if edge (v_i, v_j) exists with weight w_{ij} encoding the attributes on edges and $\mathbf{A}[i, j] = 0$ otherwise.

Koopmans-Beckmann’s formulation is a special case of Lawler’s where the edges can only be weighted by a scalar value (i.e. cannot support a vector of attributes on edges). Under this constraint, we can decompose the large matrix Φ_{12} into three smaller matrices Ψ_{12} , \mathbf{A}_1 and \mathbf{A}_2 , which provides better scalability than Lawler’s QAP.

These two formulations are combinatorial QAPs and are known to be NP-hard problems. Most methods therefore relax the hard constraints given in Eqs 1 and 2 and provide approximate solutions. Various approaches have been proposed to relax these problems, leading to a variety of graph matching methods. Exhaustive reviewing of these methods is beyond the scope of this work but we refer interested readers to the review [40].

Going back to our specific context, we reported in [35] a benchmark of the pairwise methods SMAC (Spectral Matching with Affine Constraints) [41], IPFP (Integer Projected Fixed Point algorithm) [42], RRWM (Reweighted Random Walks for graph Matching) [43], and KerGM (Kernelized Graph Matching) [44]. We observed that **KerGM** clearly outperforms the others in our context. Indeed, **KerGM** is well suited for sulcal graphs for several reasons. First, **KerGM** relies on Koopmans-Beckmann’s formulation which enables the use of attributes on both nodes and edges while limiting the memory usage. In addition, this method relies on Frank-Wolfe optimization that allows to follow an optimisation path that respects the constraint on each step, which induces a robustness to the presence of noise in graphs that is crucial in our context. In the present work, we included only **KerGM** as a representative of pairwise approaches in our benchmark because its performance was much higher than the other pairwise techniques. Note that **KerGM** also served to provide the initialization to all the multi-graph methods that are introduced in next section. The performance of **KerGM** in the multi-graph matching experiments thus represent the baseline to which the other techniques will be compared. Finally, note that since **KerGM** exploits both the attributes on nodes and edges, the information related to the topology of our sulcal graphs is implicitly taken into account in all the multi-graph matching techniques included in our study.

2.3 The multi-graph matching problem

We now focus on the problem of jointly matching a population of N graphs $\{\mathcal{G}_1, \dots, \mathcal{G}_N\}$, starting from pairwise assignment matrices \mathbf{X}_{ij} between graphs \mathcal{G}_i and \mathcal{G}_j (computed with **KerGM** in this work). The key concept behind multi-graph matching in our context is the *cycle consistency*. This concept states that a matching between two graphs \mathcal{G}_i and \mathcal{G}_j should be the same if we go through an intermediate graph \mathcal{G}_k to create a new mapping. Formally, a perfectly consistent, bijective mapping (every node is matched to one and only one other node) would satisfy:

$$\mathbf{X}_{ik} = \mathbf{X}_{ij}\mathbf{X}_{jk} \text{ ,} \tag{3}$$

for any i, j and k with $i \neq j \neq k$. A common way to estimate consistency at the population level is to compute the full bulk assignment matrix $\mathbb{X} \in \{0, 1\}^{m \times m}$ with $m = \sum_{q=1}^N |\mathcal{G}_q|$, that is obtained by assembling all individual pairwise matrices:

$$\mathbb{X} = \begin{bmatrix} \mathbf{X}_{11} & \mathbf{X}_{12} & \cdots & \mathbf{X}_{1N} \\ \mathbf{X}_{21} & \mathbf{X}_{22} & \cdots & \mathbf{X}_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{X}_{N1} & \mathbf{X}_{N2} & \cdots & \mathbf{X}_{NN} \end{bmatrix}$$

Intuitively, enforcing the consistency constraint will induce a reduction of the rank of this bulk matrix.

The first category of approaches explicitly aim at minimizing the rank of the bulk matrix using various approaches [45–48]. For instance, [49] solves a global optimization problem by using a projected power iterative method, and we detailed further [50].

The second category of techniques does not explicitly minimize the rank of the bulk matrix but rely on other types of formalization aiming at increasing the consistency across all graphs [51–54].

Finally, the third category corresponds to deep learning approaches that show promising performance in supervised tasks compared to previous methods, but are not suited for unsupervised tasks [55–62].

Some other interesting methods exploit the concept of consistency in order to solve the problem of jointly matching multiple images [63–66]. However, the extraction of the attributes on nodes is integrated and specific to images or videos (e.g. optical flow, SIFT. . .) in these methods. Application to sulcal graphs would require major modifications of the implementation that fall beyond the topic of current work.

2.4 Selection of the methods included in our benchmark

We used the following criteria to select the methods included in our benchmark: (i) *Availability of code*. We included only methods for which the authors have made their code openly available in order to avoid reimplementing issues and to ensure the full reproducibility of our results. (ii) *Scalability*. Since we are interested in performing population studies over large sets of individuals, we excluded methods that do not provide acceptable scalability. (iii) *Unsupervised methods*. Finally, as motivated in the introduction, we focus on unsupervised methods in the present study.

The methods that satisfy these selection criteria are **mALS** [50], **mSync** [45], **CAO** [51] and **MatchEig** [67]. We also identified the following methods for being relevant in our context, but they did not meet our inclusion criteria: HiPPI [49] was not included because the code was not provided; we never managed to run GAMGM [61] on our graphs despite our efforts (we suspect some well known stability issues with Sinkhorn's algorithm [68] but investigating such limitation was out of the scope of this article); we were not able to get interpretable results from LPMP [69], due to the sensitive tuning of many parameters; and IRGCL [70] did not scale with the memory requirement from our experiments, contrary to the other methods tested. We provide a detailed description of each of the methods included in our evaluation framework below.

In our experiments, these multigraph graph-matching techniques will be compared with the pairwise approach **KerGM**, and with the two methods from the literature specifically designed for labeling sulcal graphs already described in Sec. 1.3.1: **Auzias et al.** [16] and **Kaltenmark et al.** [29]. All the computations from this work were executed on a computing server with 32 physical cores of Intel XEON CPUs and 96 GB of RAM. All of the graph matching methods presented in this study were applied using the code provided by the authors, except for **MatchEIG** for which a very straightforward Matlab pseudo code was provided in their paper, allowing for re-implementation in just a few lines of code (available in our repository).

2.5 Description of the selected multi-graph matching methods

As described in section 2.3, the general objective of multi-graph matching methods is to match the nodes across several graphs together by enforcing consistency.

The authors of CAO [51] propose to maximize the affinity information and impose consistency at the same time instead of considering them separately. They assume that enforcing consistency acts as a regularizer in the affinity objective function, particularly when the matching is ambiguous due to noise. The approach is based on the search of an intermediate graph \mathcal{G}_q that allows to optimize the affinity score while progressively inducing consistency. They introduce the unitary consistency across a set of N pairwise matching solutions \mathbb{X} for a graph \mathcal{G}_q as:

$$C_u(\mathcal{G}_q, \mathbb{X}) = 1 - \frac{\sum_{i=1}^{N-1} \sum_{j=i+1}^N \|\mathbf{X}_{ij} - \mathbf{X}_{iq}\mathbf{X}_{qj}\|_F/2}{n_q N(N-1)/2}, \tag{4}$$

where $\|\cdot\|_F$ is the Frobenius norm. The authors propose several approaches to balance between consistency and affinity, leading to different variants of CAO. In particular, their best algorithm is able to elicit outlier nodes during the optimization, which is highly relevant in our context. However, the use of affinity information along with consistency and outlier elicitation increase the computational complexity of the method to $O(N^4)$. As a consequence, only the least resource-demanding algorithm CAO^{cst} did scale with the memory requirements imposed by the size of our graphs and number of subjects in our populations. We thus refer to that particular version in the rest of this article. This version of CAO enforces consistency through Eq 4, but ignores the affinity information.

The approach mSync [45] consists in estimating a mapping of each \mathbf{X}_{ij} to a common *universe* of assignment matrices, of size d :

$$\max_{\{U_i, U_j\} \in P} \sum_{i=1}^N \sum_{j=1}^N \langle U_i U_j, \mathbf{X}_{ij} \rangle, \tag{5}$$

$$\text{with } P = \{U \in \{0, 1\}^{n_q \times d} \mid U \mathbf{1}_d = \mathbf{1}_{n_q}\}. \tag{6}$$

Since solving Eq 5 is intractable in most applications, the authors relax the problem into a generalized Rayleigh problem. They further propose to use a *reference* graph in order to estimate the mapping to the *universe*. In the implementation provided by the authors, the first graph in the collection \mathcal{G}_1 is selected as the *reference* graph.

In contrast with mSync, MatchEig does not rely on a reference graph but uses the same building blocks. MatchEig uses a singular value decomposition to reduce the rank of the bulk matrix and applies the Hungarian method on the cross-correlation of corresponding singular vectors to compute the permutation matrix. As a result the consistency is not guaranteed, but in [67] the authors reported experimental results showing that MatchEig is robust to approximated estimation of the rank, and thus efficient in real conditions.

In mALS [50], the authors formalize the multi-graph matching as the following low rank matrix recovery problem:

$$\begin{aligned} f(\mathbb{X}) &= -\sum_{i=1}^N \sum_{j=1}^N \langle \Psi_{ij}, \mathbf{X}_{ij} \rangle + \alpha \langle \mathbf{1}, \mathbb{X} \rangle + \lambda \|\mathbb{X}\|_* , \\ &= -\langle \mathbb{K} - \alpha \mathbf{1}, \mathbb{X} \rangle + \lambda \|\mathbb{X}\|_* , \end{aligned} \tag{7}$$

where, $\langle \cdot, \cdot \rangle$ is the inner product, α controls the weight on sparsity, and $\mathbb{K} = \{\Psi_{ij}\}_{i,j=1}^N$ is the set of affinity matrices given as input. The cycle consistency is induced by the nuclear norm $\|\mathbb{X}\|_*$ that controls for the rank of \mathbb{X} while $\langle \mathbf{1}, \mathbb{X} \rangle$ favors bijective matchings across graphs.

Importantly, \mathbb{X} is treated as a real matrix such that $\mathbb{X} \in [0, 1]$. The matrix is binarized at the end of the optimization process using a threshold value t that is set by default as to $t = 0.5$. In addition, the authors leverage the work by [71, 72] for decomposing \mathbb{X} which allows to solve the problem in a lower dimension space using the ADMM method [73].

3 Generation of a population of synthetic sulcal graphs

A primary objective of our work is to investigate and evaluate different multi-graph matching techniques in the context of sulcal graphs. However, as mentioned in the introduction, there is no ground truth matching available for such graphs. We tackle this problem by designing a procedure allowing to generate a population of artificial sulcal graphs with correspondences defined by construction. Such populations of artificial graphs will constitute a ground truth against which the different matching methods can then be benchmarked. Generating artificial sulcal graphs for the purpose of a benchmark study induces the two following constraints: 1) The artificial graphs should be biologically plausible, i.e. they should respect as much as possible the intrinsic properties of a population of real sulcal graphs. 2) The generation of the artificial graphs should be as simple and straightforward as possible in order to facilitate the comparison of the performance obtained in the benchmark study and the interpretation of the differences, i.e. the generation procedure should rely only on a limited number of parameters, and potential biases should be avoided. As detailed below, these two contradictory constraints are balanced in the design of our generation procedure.

The procedure is summarized in Algo. 1 and consists in two main steps. First, we generate a set of points on the common spherical domain, that will serve as *reference nodes*. Then, we impose several types of perturbations to this set of reference nodes in order to generate a corresponding population of artificial sulcal graphs, while preserving the correspondences across graphs, i.e. the ground-truth matching. Such procedure provides the ground truth matching across the population, while controlling for the nature and amount of variations across artificial sulcal graphs (corresponding to different subjects in real data). We have implemented this complete pipeline in python and the code is openly accessible in the repository provided in section 1.3.3.

Algorithm 1 Procedure to generate a population of artificial sulcal graphs

```

Require:  $N, n_{ref}, K, \mu_{pert}, \sigma_{pert}, \rho$ 
Step1: create reference nodes ▷ See Sec.3.1
  for  $j = 1..10000$  do
    Sample  $n_{ref}$  points on the sphere
    Compute the minimum geodesic distance
  end for
  Choose the set of points with the largest min distance.
Step 2: generate a population of sulcal graphs ▷ See Sec.3.2
  for  $i = 1..N$  do
    Perturb location of the reference nodes ▷ See Sec.3.2.1
    Add outliers and suppress some nodes ▷ See Sec.3.2.2
    Compute the edges of the graph ▷ See Sec.3.2.3
  end for

```

3.1 Generation of a set of reference nodes

The first step consists in generating a set of reference nodes on the spherical domain while controlling for two specific distinct parameters: the *number of nodes* noted n_{ref} , that is typically set to match the average number of nodes across a real population, and the *minimum distance between the nodes*. Indeed, the nodes of the real sulcal graphs cannot be closer to each other than a minimum distance since they correspond to depth maxima that are not located in the

immediate proximity of the boundary of sulcal basins (see [16] for further description of the extraction of sulcal pits and basins). As a consequence the spatial distribution of the nodes on the sphere cannot be fully random. In order to generate this set of n_{ref} points on a sphere with pseudo-random spatial distribution, we adopted a simple brute force approach: we sample a set of n_{ref} points over the surface of the sphere 10000 times; and we select the set that has the largest minimum geodesic distance between neighbouring points. As we will show in sec.4.3.1, 10000 times is sufficient to get a set of reference nodes with a minimum distance between points that is realistic. Technically, the uniform sampling of points on the sphere is achieved by generating random rotations of the unit vector as described in [74, 75].

At this stage, we have defined on purpose a set of *reference nodes* that matches a real population in terms of size and of minimal distance between nodes. The next step consists in perturbing the reference nodes in order to generate the population of synthetic sulcal graphs.

3.2 Generation of an individual sulcal graphs

We now add perturbations of different natures to this set of reference nodes in order to obtain a population of artificial sulcal graphs, that corresponds to different subjects. These perturbations aim at mimicking the inter-individual variations that are observed in a healthy population, by affecting the features of the nodes and edges, but also the topology of the graphs. In order to generate a population of N artificial sulcal graphs, these operations are repeated N times independently.

3.2.1 Perturbation of the location of the reference nodes. The first step consists in adding random noise to the coordinates of the reference nodes on the sphere, in order to model the inter-individual variability that exists in the location of the sulcal pits. We used the von Mises-Fisher (vMF) distribution that is an approximation of Gaussian distribution on a sphere [76]. The two parameters of the vMF distribution μ and κ can be seen as the equivalent of the mean and of the inverse of the standard deviation ($\kappa \propto 1/\sigma$) for a Gaussian distribution. Therefore, we iterate across the reference nodes, and for each reference node, we produce a noisy one by sampling from the distribution $vMF(\mu, \kappa)$, where μ is the coordinates of this reference node. We control for the amount of noise on the coordinates of the perturbed nodes through the value of the parameter κ , that is common to all nodes from the reference set. Smaller values for κ will induce larger variations across the artificial sulcal graphs within the population. Importantly, note that since we perturb each node of the reference set independently, we keep the correspondence between each noisy node and its reference node, which will allow defining our ground truth matching at the population level.

3.2.2 Addition of outliers and suppression of nodes. Next, we simulate the inter-individual variations in the number of nodes across the sulcal graphs, which is of crucial importance for generating biologically plausible artificial populations. The aim is to model both false positive and false negative matchings, i.e. respectively nodes that are present in the reference set but not in a given graph, and nodes that are present in the graph but not in the reference set.

This is achieved by randomly adding a certain number n_o of nodes on top of the perturbed nodes—hereafter called *outlier nodes*, and by deleting n_s nodes amongst the perturbed nodes—hereafter called *suppressed nodes*. In order to randomly draw n_o and n_s , we use the β -binomial distribution $B(v, \alpha, \beta)$, which is a distribution of non-negative integers. The parameter v denotes the size of the support of the distribution, i.e the maximal value that can be sampled. The parameters α and β can be set so that $B(v, \alpha, \beta)$ approximates a Gaussian distribution. We describe the setting of these parameters and precise their link with μ and σ of a Gaussian in the S1 and S2 Figs. Since we want the average number of nodes across the population of perturbed graphs μ_{simu} to match the number of nodes in the reference set n_{ref} , we set $\mu_o = \mu_s = \mu_{pert}$ and

also $\sigma_o = \sigma_s = \sigma_{pert}$. This formulation allows us to control the standard deviation of the number of nodes across the population of artificial graphs with the two parameters μ_{pert} and σ_{pert} .

3.2.3 Construction of the edges. The last step consists in constructing each artificial sulcal graph with the sets of perturbed nodes as follows. We first compute the three-dimensional convex hull of each set of perturbed nodes located on the sphere. This yields a triangulation where only neighboring nodes on the sphere are connected, which is a simple way to simulate the region adjacency graph that is constructed from the sulcal basins in the real data. However, the average node degree in such triangulations is higher than for real sulcal graphs. Therefore, we finally delete a small percentage p of the edges in these triangulations, in order to obtain artificial graphs which match the average degree of real sulcal graphs.

Note that since the construction of the edges occurs after the previous perturbation steps (perturbations of the location, addition of outlier nodes and suppression of nodes), the resulting artificial sulcal graphs can show variations in their topology across individuals of a population, as we observe in real data, making them biologically-plausible in that respect.

4 Experiments and results

4.1 Computation of the affinity matrices

As described in Sec.2.2, we initialize all the multigraph matching methods using the pairwise results obtain from **KerGM**, which relies on the formalization of Eq 2. We thus need to compute the affinity matrices Ψ_{ij} , \mathbf{A}_i , \mathbf{A}_j that store the similarity between nodes and edges across every pairs of graphs in the population.

In the present work, we compute these affinity matrices using Gaussian kernels applied to the attributes. For two nodes $v \in \mathcal{G}$ and $v' \in \mathcal{G}'$ the affinity value is computed using the kernel defined as $\exp(-\gamma^V \|a_v^V - a_{v'}^V\|_2^2)$ and for two edges $e \in \mathcal{G}$ and $e' \in \mathcal{G}'$ the kernel is defined as $\exp(-\gamma^E \|a_e^E - a_{e'}^E\|_2^2)$. To estimate appropriate values for γ^V and γ^E we use a heuristic proposed in [18] that consists in using a cross-validation scheme to compute the inverse of the median of the distribution across all possible pairs of nodes/edges, independently for each attribute (3D coordinates on the sphere for the nodes and the geodesic distance for the edges).

4.2 Dummy nodes

Most graph matching methods assume a constant number of nodes across the graphs to be matched, which is not the case in our case (both synthetic and real graphs). We use the classical approach from the graph matching literature which consists in adding *dummy nodes* to smaller graphs so that all the graphs get the same number of nodes as the largest graph in the population. For each of these dummy nodes, we assign to 0 the corresponding values in the node and edge affinity matrices. This makes the optimization problem defined in Eq 2 independent from dummy nodes.

4.3 Benchmark on synthetic sulcal graphs

4.3.1 Description of synthetic data sets. We first tuned empirically the parameters to the values $\mu_{pert} = 12$, $\sigma_{pert} = 4$ and $p = 10\%$ to obtain variations in our synthetic graph populations that are in line with what is observed in real data. The distribution for number of nodes in the real data population is 88.27 ± 4.72 likewise in our simulated population for a randomly chosen κ value the distribution for number of nodes is 88.15 ± 4.45 for the selected value of μ_{pert} and σ_{pert} and is consistent across all κ values across all trials. We further provide in S3–S5 Figs additional materials showing the matching distributions between our simulated graphs and real data.

Furthermore, we varied the value of $\kappa \in [100, 200, 400, 1000]$, which controls the amount of variations across synthetic graphs within a population. Note that κ controls the spread of nodes coordinates around the reference nodes, which in turn induces variations in the topology and attributes of synthetic graphs.

For each value of κ , we generate 10 populations of $N = 137$ synthetic graphs (which corresponds to the number of subjects in our real population; see below) and report the average and standard deviation of the metrics described below. As illustrated on Fig 3, our populations of synthetic graphs show variations that are qualitatively very close to those observed across real graphs.

4.3.2 Evaluation metrics for synthetic data sets. In order to evaluate the different matching methods on simulated graphs, we use the classical *precision*, *recall* and F_1 -score:

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \in [0, 1] \quad (8)$$

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \in [0, 1] \quad (9)$$

$$F_1 = 2 \frac{(precision \times recall)}{precision + recall} \in [0, 1] \quad (10)$$

Thus, *Precision* is a ratio between the True positives (number of correct matches predicted by the algorithms) and all the positives (number of matches by the algorithms). Whereas, *Recall* is a ratio between True positives and True positives along with False negatives (number of correct matches not predicted by the algorithms). Finally, the F_1 score provides a balance between *Precision* and *Recall*. A F_1 -score of 1 reflects the ability of the algorithm to obtain a perfect matching

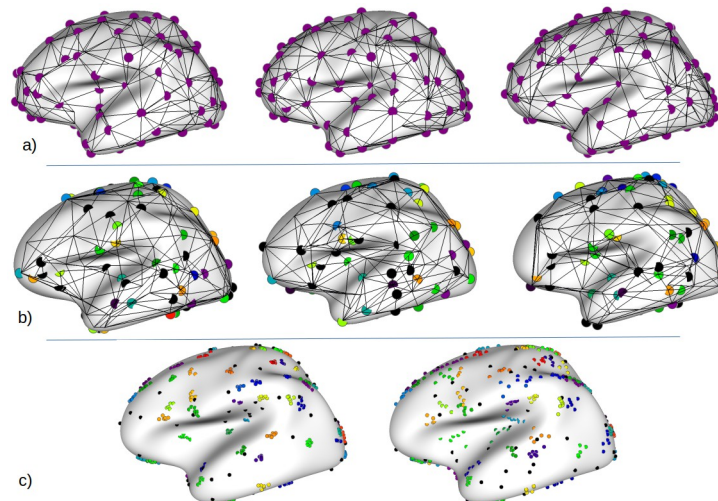


Fig 3. a) Real sulcal graphs from three randomly chosen individuals, and projected on the average surface. b) Simulated graphs randomly chosen for $\kappa = 1000$, showing the ground-truth correspondence across graphs in color. Nodes in black represent the outlier nodes that have no correspondence. c) Illustration of the impact of κ on the spatial dispersion of nodes: the nodes of six simulated graphs are shown on the average surface for $\kappa = 1000$ (left) and $\kappa = 200$ (right). The spread across the nodes for each cluster varies according to κ , while outlier nodes in black have random locations.

<https://doi.org/10.1371/journal.pone.0293886.g003>

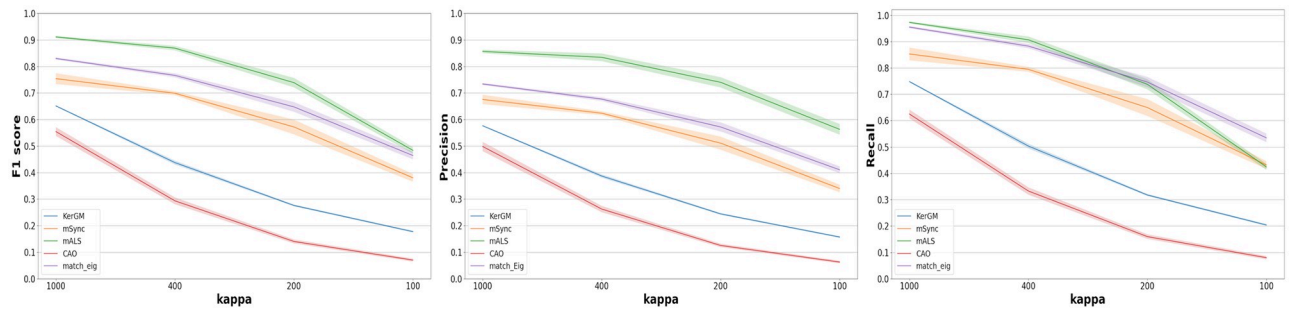


Fig 4. F_1 -score, precision and recall for $\kappa \in [1000, 400, 200, 100]$. For each method, we plot the average across the 10 simulated populations as a line and the standard deviation as the shaded region of the same color.

<https://doi.org/10.1371/journal.pone.0293886.g004>

of inlier nodes and accurate identification of outlier nodes. These metrics are relevant in our context to detect matching with outliers alongside the incorrect matches.

4.3.3 Results on synthetic data sets. We report in Fig 4 the mean and standard deviation of *Precision*, *Recall* and *F₁-score*, computed across the 10 synthetic populations for each value of κ .

First, we find that three multi-graph matching methods, **mALS**, **MatchEig** and **mSync**, vastly and consistently outperform **KerGM**, which has been identified as the best pairwise matching method for this task in [35]. This confirms our main hypothesis: considering the matching problem on the whole population using multi-graph matching allows an important gain in performance compared to only considering pairs of graphs.

Then, we observe a gradual decline in the performance of all methods as the noise increases (decrease of κ), as expected. The performance of the multigraph approaches **mALS**, **MatchEig** and **mSync** resist much more to this increase in variability than the pairwise approach. The performance of **mSync** is limited more specifically by the lower precision at any level of noise. This suggests that the difference in performance between **mSync**, **mALS** and **MatchEig** is mainly due to the hard constraint on the consistency in **mSync** that seems too restrictive. The higher performance of **MatchEig** compared to **mSync** while the two methods are based on the same building blocks supports this interpretation.

The performance of **mSync**, **MatchEig** and **mALS** are very similar when looking at the recall measure. However, the precision of **mALS** is higher than all the other methods, for all noise levels. Overall, **mALS** shows the best *F₁-score* for every κ values, thanks to a very high precision combined with very good recall. Indeed, the *F₁-score* for **mALS** is above 0.7 even for $\kappa = 200$ which corresponds to a configuration where the noise is quite strong.

Finally, the performance of **CAO** is very low, even lower than the pairwise technique **KerGM**. Such poor performance is likely a consequence of the optimization that considers only the consistency but ignores the affinity of nodes. As already mentioned in Sec.2.5, the other versions of **CAO** proposed in [51] could show much higher performance but did not scale with the size of our data.

4.4 Application to real data

4.4.1 Preprocessing of real data. For the evaluation on real data, we use the sulcal graphs from 137 young healthy adults (69 females and 68 males) selected from the publicly available database OASIS [77]. The preprocessing of these data (brain tissues segmentation, mesh extraction and sulcal graphs construction) has been detailed in [16, 18]. Across this population,

the number of nodes is 88 ± 4 , with a maximum size of 101 nodes/pits. Dummy nodes are thus added to all other graphs to get a constant size of 101, as explained above.

4.4.2 Evaluation metrics used with real data. In absence of ground truth matching for real data, we cannot compute the same scores as for the simulation experiments. We therefore combine a set of quantitative metrics with some qualitative assessments, which we describe below.

Consistency. According to [51], we compute the node consistency as follows: Given $\mathcal{G}_k \in \{\mathcal{G}_q\}_{q=1}^N$ and the bulk matrix \mathbb{X} , for node $v^k \in \mathcal{G}_k$, with index $i(v^k) \in \{1, \dots, |\mathcal{G}_k|\}$, its consistency is defined by:

$$C(v^k; \mathbb{X}) = 1 - \frac{\sum_{i=1}^{N-1} \sum_{j=i+1}^N \|\mathbf{Y}(v^k, :)\|_F / 2}{N(N-1)/2}, \in (0, 1], \quad (11)$$

where $\|\cdot\|_F$ is the Frobenius norm, $\mathbf{Y} = \mathbf{X}_{kj} - \mathbf{X}_{ki}\mathbf{X}_{ij}$ and $\mathbf{Y}(v^k, :)$ is the $i(v^k)$ -th row of matrix \mathbf{Y} . Note that it is different from Eq 4 which estimates the consistency at the graph level. This consistency measure is computed for each node of each graph, including dummy nodes. A value of 1 corresponds to the ideal case where each graph only contains nodes that have been matched in a consistent manner. This consistency measure cannot distinguish the matches of real nodes to dummy nodes from valid matches across real nodes. For methods imposing an explicit constraint on the consistency, a value of 1 is expected (and not informative), but for the other methods this measure is relevant and allows to assess the spatial pattern of the consistency across clusters.

Qualitative and quantitative assessment of the labeling induced by the matching. In terms of potential applications of the graph matching to sulcal graphs, a major outcome is the labeling of graph nodes that is induced. As already mentioned in the introduction, the assessment of the quality of the labeling and thus of the biological relevance of the matching across individuals is an ill-posed problem. The first problem is to retrieve a labeling from the assignment matrix resulting from the matching. In the case of a perfectly consistent matching where each node of each graph would be matched to one and only one node from every other graph in the population, the labeling would be trivial and would consist in simply associating a label to each row or column of the assignment matrix. This situation is however impossible since the number of nodes varies across individuals within our population of interest. Therefore, in the present work we take the largest graph as a reference, and we associate an arbitrary label to each of its nodes and then propagate these labels to every other graphs based on the assignment matrix resulting from each method.

Once the labeling of the nodes is retrieved, the nodes that share the same label across subjects are grouped together into what we will designate as *clusters*, that are different depending on the matching method. We then compute the coordinates of the centroid of each cluster, which enables to evaluate qualitatively the spatial distribution of the different clusters across the cortical surface.

This qualitative assessment is complemented with a quantitative measure of the compactness of the clusters. For this, we compute the silhouette coefficient of each node from each graph. As proposed in [78], the silhouette of a node corresponds to the ratio between the average Euclidean distance to the other nodes in the cluster and its distance to other nearby clusters. Since the distances are computed on the spherical domain, the use of Euclidean distance is sub-optimal but the errors induced are very low and independent from the matching method. The silhouette coefficient of a cluster is then obtained by averaging the silhouette values from corresponding nodes.

4.4.3 Results on real data. We first report in [Table 1](#) the quantitative measures that allow us to compare the different techniques at the whole brain level: the number of clusters (thus of labels) obtained with each method, the silhouette measure averaged across all nodes and graphs, the percentage of nodes remaining unlabeled, the consistency measure averaged across all nodes and graphs, and the computing time.

The number of clusters and percentage of unmatched nodes indicate that the two methods that allow partial matching **mALS** and **Kaltenmark et al.** result in a lower number of clusters, suggesting that the ambiguous nodes remain unlabeled instead of enforcing their matching into potentially unreliable clusters. The methods **MatchEig**, **mSync**, **CAO** and **KerGM** enforce the matching of every nodes, and result in a number of clusters equal to the size of the largest graph in the set, i.e. 101. The method **Auzias et al.** results in more clusters than the size of the largest graph, suggesting that some clusters correspond to highly variable nodes that cannot be matched consistently across individuals. This is confirmed by the consistency measure which is lower than for **mALS**. The consistency of **Auzias et al.** is still much higher than the value of 0.30 obtained with the pairwise technique **KerGM**. The performance of **MatchEig** lies between **mSync** and **Auzias et al.**. Note that the methods **mSync** and **CAO** explicitly enforce a perfect consistency, but this is possible only when considering the dummy nodes as pointed in section 4.4.2. Also note that the method **Kaltenmark et al.** also gets a perfect consistency. This is a consequence of the explicit constraint imposed in this technique by allowing one and only one node per subject to be matched for any given cluster.

The silhouette measures illustrate that a high consistency can be associated with a low compactness of the clusters as e.g. for **CAO** and **mSync** that get values close to the one of the pairwise technique **KerGM**. **MatchEig** get higher silhouette values than these techniques. The methods **Auzias et al.** and **Kaltenmark et al.** get much higher silhouette values which is expected since these techniques enforce the matching of nodes based essentially on their spatial proximity on the surface. The silhouette value of **mALS** is higher than these two techniques. Overall, **mALS** results in high silhouette and consistency values, at the cost of a high number of unmatched nodes (28.4%) compared to **Kaltenmark et al.** and **Auzias et al.**, indicating that this method was much more conservative in the matching, leaving more ambiguous nodes unmatched.

We then illustrate the matching across nodes from the different graphs (subjects), obtained for each method on [Fig 5](#). We do not show the results from **CAO** to save space, since the performance of this method on both simulations and real data were worse than the pairwise technique **kerGM**. The number and location of the different centroids (larger circles) is informative of the spatial distribution of the clusters of nodes across the cortical surface, for each method. On the first column (**mALS** and **Kaltenmark et al.**) some nodes remain unlabeled and are represented in black. The clusters seem more compact than for the methods

Table 1. Quantitative measures computed at the whole brain scale.

Method	Num. clusters	silhouette	Perc. unmatched	consistency	cpu time (min)
mALS	82	0.55 ± 0.22	28.4	0.91 ± 0.08	783
Kaltenmark et al	94	0.44 ± 0.23	17.0	1.0 ± 0.0	~ 180
Auzias et al	104	0.49 ± 0.18	0	0.82 ± 0.15	~ 30
MatchEig	101	0.18 ± 0.40	0	0.68 ± 0.02	13
mSync	101	0.08 ± 0.49	0	1.0 ± 0.0	31
CAO	101	-0.12 ± 0.45	0	1.0 ± 0.0	3255
KerGM	101	-0.04 ± 0.34	0	0.30 ± 0.17	1362

<https://doi.org/10.1371/journal.pone.0293886.t001>

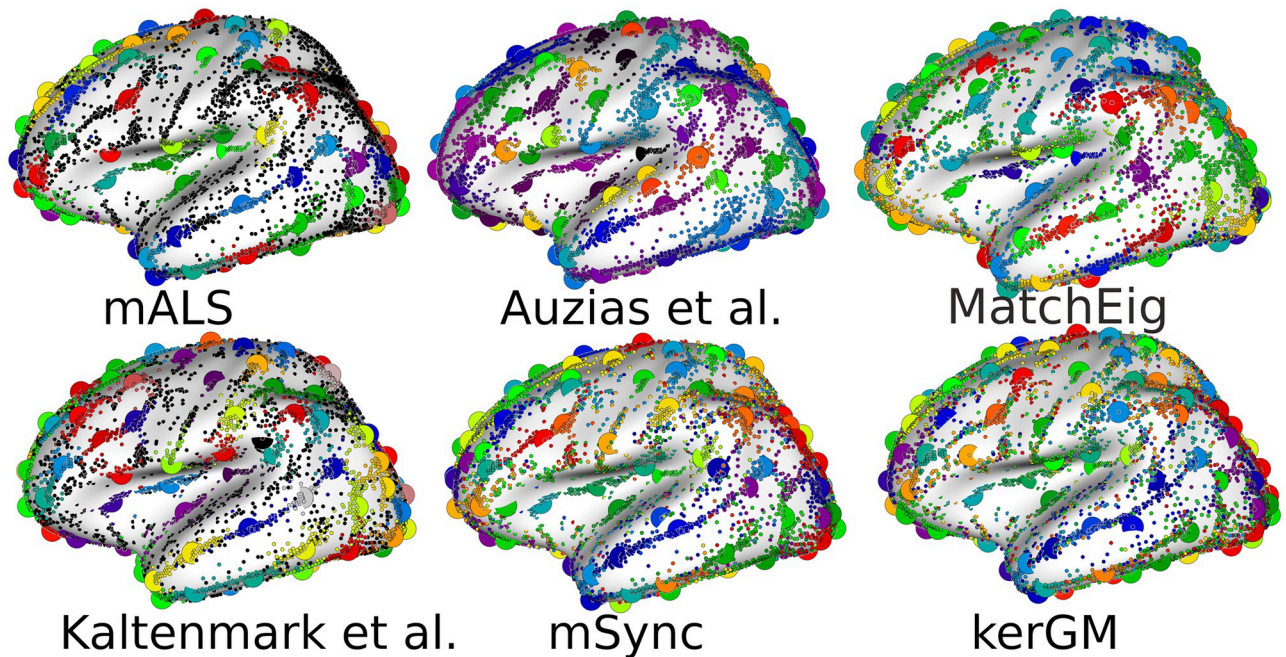


Fig 5. Labeling and corresponding cluster centroids (larger circles) for each method. Dots in black in the first column (for **mALS** and **Kaltenmark et al.**) correspond to unmatched nodes. See text for further description.

<https://doi.org/10.1371/journal.pone.0293886.g005>

Auzias et al., **MatchEig** and **mSync** that do not allow any node to remain unlabeled. With **kerGM**, the matching looks noisy, with clusters overlapping between each other in almost every cortical location, which illustrates the poor anatomical relevance of the matching.

For further evaluation of the performance of the different techniques, we show on [Fig 6](#) the silhouette values of every nodes across all graphs as well as the centroids of each cluster as a larger circle. The high silhouette values of the centroids for the methods **mALS**, **Auzias et al.** and **Kaltenmark et al.** are visible with mostly red and orange centroids. In contrast, we observe more centroids in green and blue for **KerGM**. Together with [Table 1](#), this figure illustrates the poor performance of pairwise matching approach with high spatial dispersion of nodes corresponding to each cluster for **KerGM**, associated to very low silhouette coefficients. The method **mSync** results in higher silhouette coefficients for some nodes, but lower value for others (nodes and centroids in blue on [Fig 6](#)), indicating that the matching was enforced also for ambiguous nodes located in highly variable regions. This is a consequence of the hard consistency constraint in **mSync** imposing a matching that is consistent across all graphs by construction, even in highly variable regions. The results from **MatchEig** are slightly better, with less centroids in blue than for **mSync**. For **Auzias et al.**, we observe that the clusters are organized around regions of high nodes density, but the nodes located relatively far from the centroids have a lower silhouette value (nodes in green on [Fig 6](#)). These observations are consistent with the algorithm that is based on a watershed applied to the sulcal pits density map as described in [Sec.1.3.1](#). For both **mSync** and **MatchEig**, we observe some clusters with low silhouette value located close to each other, suggesting that the number of clusters is too high.

The techniques **mALS** and **Kaltenmark et al.** result in much higher silhouette values, which is expected since they do not force the matching of highly variable nodes that are left unlabeled. The unlabeled nodes have a very low silhouette value (in violet on [Fig 6](#)), but since they do not belong to any cluster, this does not reduce the silhouette values of clusters. Note

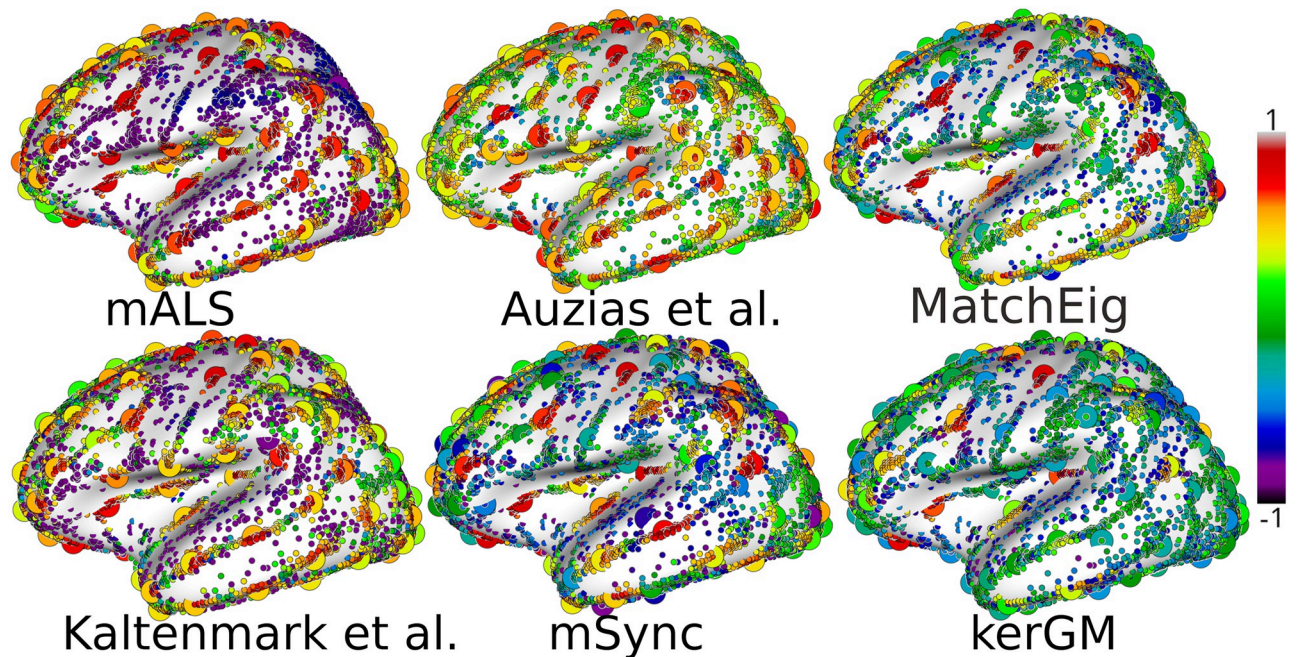


Fig 6. For each method, we show the silhouette coefficient of each node from every graphs, as well as corresponding centroids as larger circles. Each centroid (larger circles) is colored according to the average of the silhouette coefficient of corresponding nodes.

<https://doi.org/10.1371/journal.pone.0293886.g006>

that even for these methods, the clusters get closer with lower silhouette values in highly variable regions such as the anterior frontal and occipital lobes.

Across the different methods, we observe that the clusters showing a higher silhouette value relative to other clusters are located systematically in the same regions that are known to be less variable across individuals, such as the central sulcus, and the insula. For these clusters, the silhouette values are close across methods, confirming the lower ambiguity in the matching in these regions. In highly variable regions, the different methods produce different matchings. For instance in the occipital lobe, the clusters produced by **Kaltenmark et al.** show lower silhouette values compared to **mALS**, but we observe the opposite effect in the anterior frontal lobe.

On **Fig 7**, we show the consistency for every nodes and centroids, for the four methods that do not explicitly enforce a perfect consistency. Clearly, the pairwise technique **KerGM** results in inconsistent matching for every clusters, including the regions where the variations across individuals are known to be low (no centroid in green, even in the central sulcus and the

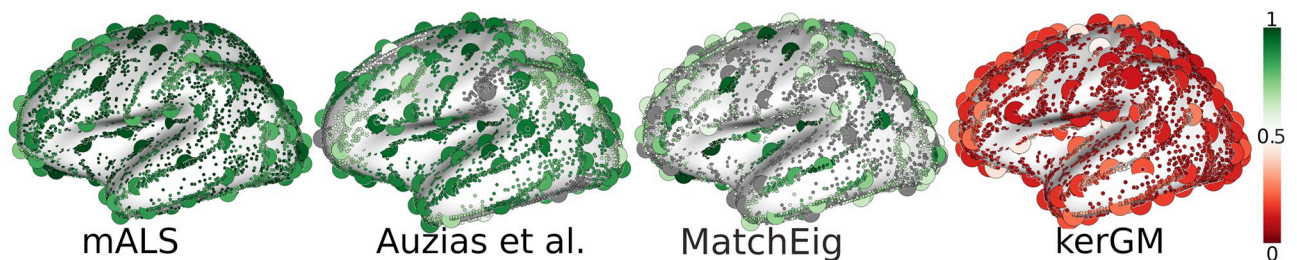


Fig 7. Node consistency computed for each node of each graph with respect to the remaining graphs, and then averaged across graphs. We adapted the colorbar to visualize the differences between the three methods, with the pairwise technique **KerGM** showing much lower values.

<https://doi.org/10.1371/journal.pone.0293886.g007>

insula). For **mALS** and **Auzias et al.**, we can observe the spatial variations of the consistency across cortical regions. Again, higher consistency is obtained in less variable regions (central sulcus, insula) for both techniques, and relatively lower values are visible in the frontal and occipital regions. With **MatchEig**, the clusters with lower consistency (in grey, close to .5) corresponds to the clusters with lower silhouette value on **Fig 6**. These clusters are located in highly variable regions such as the top of gyri. The consistency is higher for **mALS** than **Auzias et al.** for every clusters. Note that the spatial pattern of the consistency measure for **mALS** is anatomically relevant, with a consistent matching in the insula, the central and pre-central regions, and less consistent in the peri-sylvian regions. At a more local scale, we observe a cluster in the superior temporal sulcus that is more consistent than those located anteriorly or posteriorly, which is in line with previous studies describing variations and stabilities across individuals in this region [79].

4.4.4 Exemplar application: Group statistics. We report an exemplar application of the multi-graph matching framework in the context of a statistical comparison between two groups of subjects, which is a classical task in the literature. In this experiment, we divide our population of 137 subjects into two groups depending on their sex: a group of 69 females and a group of 68 males. We then use the matching across graphs resulting from our previous experiment to define correspondences across the sulcal basins from all the 137 individuals from the two groups. Thanks to this matching, comparing the two population is trivial and one can simply compute a t-test between the two groups in order to assess statistically potential differences related to their sex in any of the features stored in the graphs as attributes of nodes. On **Fig 8**, we report the t-value from the t-test computed to assess potential difference in the depth of sulcal basins between males and females.

As we can observe on this figure, the group statistics are strongly influenced by variations in the matching resulting from the different techniques. More specifically, few regions show a

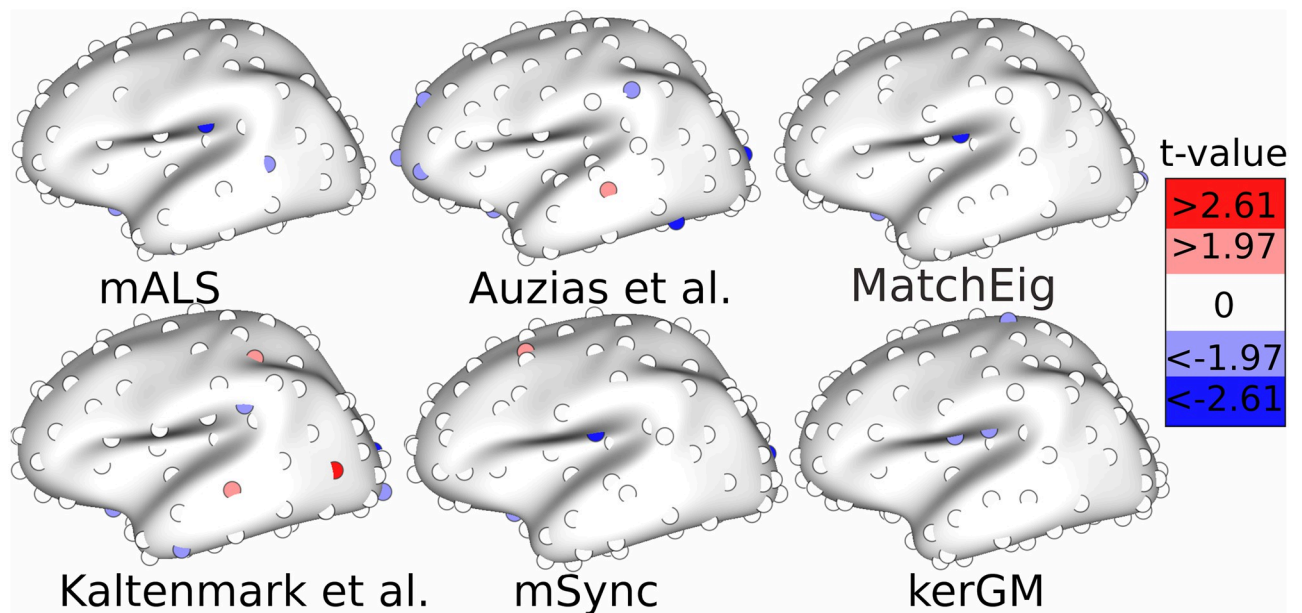


Fig 8. For each method, the clusters are colored following t-values obtained from the t-test comparing the depth of sulcal basins between males and females, with $t > 0$ $M > F$. To facilitate the comparison across the different methods, we applied two thresholds to the t-values: t-values superior to 1.98 or inferior to -1.98 correspond to $p < .05$ (two-sided test with a dof 135), and t-values superior to 2.61 or inferior to -2.61 correspond to $p < .01$, and t-values closer to 0 corresponding to a non-significant difference are colored in white.

<https://doi.org/10.1371/journal.pone.0293886.g008>

weak statistical difference ($p < .05$) between the two groups with the matching from **kerGM**. The sulcal graph matching techniques from previous literature **Auzias et al.** and **Kaltenmark et al.** result in very different statistics. With **Auzias et al.**, we observe one cluster with a lower depth in males ($p < .01$) in the basal temporal lobe while the most significant cluster with **Kaltenmark et al.** is located in the posterior temporal cortex and corresponds to a higher depth in males ($p < .01$). We also observe a region with higher depth ($p < .05$) in males with the two methods in the anterior temporal lobe. Across the three graph matching techniques **mALS**, **mathEIG** and **mSync**, the most significant cluster ($p < .01$) is consistently located in the posterior insula and corresponds to a lower depth in males. Therefore, this experiment confirms that the matching is key to enable statistical comparisons, and the choice of the method has a strong influence on the resulting statistics. Further interpretation of these group statistics fall beyond the scope of the current methodological work, but we refer readers interested in such statistical analysis at the scale of sulcal basins to e.g. [16] where the authors reported an experiment on the asymmetry across left and right hemispheres, [80] for a statistical analysis of the relationship between basins frequency and IQ, or [81–83] for applications to psychiatric disorders. Note that a strong influence of the sulcal basins matching technique on the resulting statistics as we observed in Fig 8 is expected for all these publications. Finally, we emphasize that extending the analysis to other features stored as attributes of nodes is straightforward. Many other features of interest can be easily extracted and stored as attributes of nodes such as for instance cortical thickness, curvature or an estimation of the cortical myelin [27].

5 Discussion

In this work, we explored the potential of graph matching methods applied to a population of sulcal graphs to uncover correspondences across individuals driven by the local patterns of folds. Indeed, these graph matching methods can use the characteristics of individual sulcal basins as well as their topological organization to construct the correspondences. Our results on both simulations and real data support the biological relevance of the correspondences across individual resulting from multi-graph matching techniques.

5.1 Relevance of simulated graphs relative to real data for evaluating matching techniques

To overcome the lack of ground truth for real data, we proposed a procedure allowing to generate artificial graphs that approximate the features of real sulcal graphs while controlling the variations across graphs. This simulation procedure enabled to benchmark various pairwise and multi-graph matching techniques. The evaluation of the performance of the different methods and their robustness to controlled variations in the simulated graphs was informative for probing their effectiveness in this context. The performance of the pairwise approach **KerGM** was limited even when the level of perturbations was minimal. Note that we reported in [35] that alternative pairwise techniques perform even worse on this task. Amongst the different multi-graph matching techniques tested, **mALS** showed better performance than the others in all conditions, and a good robustness to increasing noise levels. These observations were confirmed by our application on real data. Overall, our set of experiments confirmed the intuition that multi-graph matching techniques are highly relevant in our context, while pairwise techniques show limited performance and might thus be restricted to initialization purpose.

Of note, our aim was not to push the biological plausibility of our simulated graphs. Keeping the simulations simple enables straightforward interpretation of the variations in the performance across the different approaches. This trade-off is visible in the procedure in

particular when we sample the reference nodes uniformly on the sphere. Indeed, our simulation procedure cannot produce realistic non-uniform spatial distribution of nodes across the population. While this could be achieved by adapting the sampling of the reference points, this would induce variations in the performance of the matching techniques depending on the location on the sphere, which in turn would make the comparison across methods much more difficult.

Beyond the present work, our procedure for simulating sulcal graphs could be instrumental to assess future improvements in graph matching techniques.

5.2 Potential methodological improvements and considerations relative to deep-learning approaches

As already mentioned in section 2.3, many other graph matching techniques can be found in the literature but were not included in the present work. More specifically, deep learning approaches outperform traditional approaches in supervised learning task [84]. Recent works such as e.g. [85, 86] showed that the structural information can be learnt by a Graph Neural Network (GNN), providing that manually labelled ground-truth data is available.

In addition, the rise of semi-supervised learning approaches represents an opportunity in the context of graphs with partial matching ground-truth. Such approaches are worth considering in our context, since we observed marked variations across cortical regions in the ambiguity of the matching. The work by [87] considers a semi-supervised framework for handling the matching problem where the ground-truth correspondences are only given for a small subset of nodes. In addition, their approach imposes an explicit inductive bias to find correspondences across graphs, based on neighbourhood consensus that does not allow adjacent nodes from being mapped to different regions in other graphs. This is appealing in the case of sulcal graph matching where we would like to enforce the matching of nodes located in some specific regions more than in others. Such a framework could benefit from the recent work [22] on context-aware data augmentation, which could be instrumental to overcome the bottleneck of the lack of ground-truth labeling data.

Another avenue for potential gains in performance consists in improving the definition and integration of the attributes on nodes and edges. Many other geometrical features could be considered to enrich the attributes on nodes, such as e.g. shape index and curvedness [88], or the local gyrification index [89]. Note however that the complementarity of the different attributes on nodes is crucial to improve the matching performance.

On the other hand, the attributes on edges are most often reduced to a scalar value (i.e. a simple weight), due to technical limitations of the graph matching methods. Indeed, sulcal graphs can be enriched with various types of attributes on edges, which would greatly help the matching algorithm. For instance, in [80] the authors proposed to use as an attribute on edges the depth of the shallower point separating two neighboring sulcal basins, denoted as ‘ridges’. The depth of the ridge between two basins is a good descriptor of the local cortical geometry since the two basins can appear almost as separated as two different sulci if the ridge is very superficial, while in other cases, sulcal basins can constitute a long continuous sulci with no interruption when the ridge is deep. Other highly relevant attributes on edges would be those related to the connectivity across cortical areas, such as the structural connectivity extracted from diffusion weighted MRI [90] or the functional connectivity extracted from resting state functional MRI [91]. Note however that here also, an appropriate assessment of the complementarity of the information carried by the different attributes is missing. Such analysis would in addition face the problem of the lack of methods in the literature attacking the problem of learning edge representations [92]. In particular, the methods included in the present work

cannot handle *vectors* of attributes on edges. Some recent deep learning methods such as [58] can exploit such vectors of attributes, but their scalability is limited by the size of the affinity matrices. We proposed in [93] to overcome this limitation by leveraging the recent matrix factorization method from [44]. We also identified very recent works that are particularly relevant in this context. [94] introduced innovative strategies to address noisy matching at both node and edge levels. [95] leverages representation learning techniques to acquire universal points for partial matching. These approaches offer various possibilities for potential improvement in current context, which might be assessed in future studies.

5.3 Data-driven nomenclature of sulcal basins

The present work extends previously reported experimental results illustrating the major impact of the matching strategy on the induced correspondences and data-driven nomenclature. In [29], the number of clusters obtained at the group level varied from 90 to 114 for the right hemisphere using either the approach proposed in [16, 29] respectively, on the same population of subjects. We provide a much more detailed comparison. We show on Fig 9 the superimposition of the centroids from different methods on the same average surface. This visualization shows that the location of some of the centroids are very consistent across methods (indicated by arrows), corresponding to cortical regions where variations across individuals are known to be low, such as the central sulcus, the insula, the inferior precentral or superior temporal sulcus. Other clusters differ across methods. The clusters indicated by squares are those resulting from **Auzias et al.** and **mSync** (resp.) that do not match clusters from **Kaltenmark et al.** (see also Table 1). These are located either in highly variable regions such as the frontal lobe, or on the top of gyri such as the superior temporal gyrus and the inferior frontal gyrus. While **mALS** and **Kaltenmark et al.** result in centroids that are highly similar (crosses and rings are often superimposed on the panel on the left), the two methods do not result in the same matching in highly variable regions such as the inferior frontal and parietal regions (indicated by diamonds). In conjunction with our results on synthetic (Sec.4.3.3) and real data (Sec.4.4.3), these observations confirm that the conceptual differences between the approaches yield different matching. Our experiment on group statistics in Sec.4.4.4 confirms that different correspondences across individuals induce strong variations in the subsequent statistical analysis. Indeed, graph-matching techniques are able to take into account the

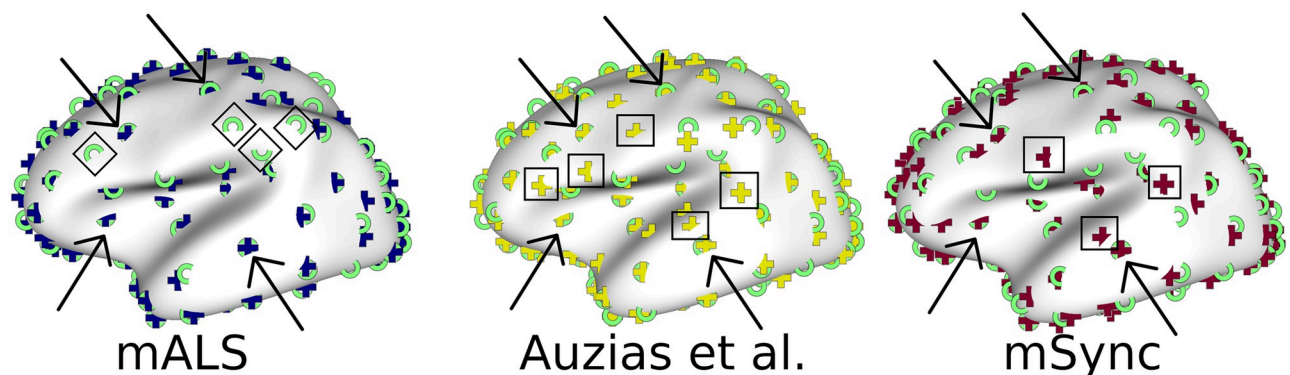


Fig 9. Superimposition of the centroids from mALS, Auzias et al., and mSync shown as crosses with those of Kaltenmark et al. shown as green rings. mSync is representative of the methods kerGM, MatchEig and CAO that also result in 101 clusters. The arrows point to centroids that are robust across methods. The squares indicate centroids corresponding to small clusters located on gyri. Diamonds indicate centroids that differ between mALS and Kaltenmark et al.

<https://doi.org/10.1371/journal.pone.0293886.g009>

topological information encoded in the graphs, i.e the spatial organization of neighbouring folds, while **Kaltenmark et al.** relies only on the geometry of sulcal basins, considering different folds separately.

The next step will be to assess the biological relevance of the induced correspondences across subjects by visualizing the matching on the cortical surface of the individuals. Given the variations across methods in the location of clusters observed on [Fig 9](#), we expect to observe important differences between the techniques at the individual level, especially in highly variable regions such as the parietal lobe. More specifically, our expectation is that graph matching techniques should allow solving potential anatomical ambiguities in a much more relevant way than **Auzias et al.** and **Kaltenmark et al.**, by exploiting the topological information of the neighbouring folding pattern. Furthermore, we also aim to validate our findings in larger, multi-modal databases like HCP [96], thereby confirming the reproducibility of the acquired nomenclature and to gain a understanding on its relevance with respect to the functional organisation of the brain.

6 Conclusion

In this study, we explored the potential of several graph matching methods selected from the literature to define population-wise correspondences across individual cortical geometries. In the absence of a ground-truth labeling for real data, we first proposed a procedure to generate simulated sulcal graphs that follow the intrinsic structure and properties of real sulcal graphs. We then compared the approaches on our simulated sulcal graphs with ground truth correspondences defined by construction.

We also evaluated the methods on 137 real graphs, and compared the results with two other methods from the literature. We computed the silhouette value of each node of the graph that measures the degree of compactness of each cluster, giving us insights on the matching across graphs produced by the different methods. The consistency measure gave us an insight into the variability across the population for each cluster. Finally, we demonstrated the influence of the matching on the statistical analysis that depend on the induced correspondences across individuals. Overall, our experiments on both artificial and real data showed the high relevance of multi-graph methods for sulcal graph matching. We observed that **mALS**, **MatchEig** and **mSync** outperform **CAO** and the pairwise approach **KerGM**. While **mALS** proved to be very robust to noise compared to other methods, the much lower complexity of **mSync** and **match-eIG** makes them also relevant candidates for further studies and extensions to larger populations.

Supporting information

S1 Fig. Beta-binomial effect. Effect on β -binomial mass function for different values of ν fixing $\alpha = 7.15$ and $\beta = 28.62$.

(PNG)

S2 Fig. Beta-binomial distribution. β -binomial distributions for identical mean: $\mu, \mu_1 = 12$ but different standard deviations: $\sigma = 3, \sigma_1 = 5$. The dotted lines signifies the mean of the distribution where as the shaded area is the standard deviation across 5 trials.

(PNG)

S3 Fig. Distribution for number of nodes. Distribution for number of node in the simulated population corresponding to different κ value with the distribution for number of nodes in the real sulcal graphs.

(PNG)

S4 Fig. Distribution for geodesic distance. Distribution for geodesic distance in the simulated and real population of 137 graphs. The shaded region corresponds to standard deviations across graphs in the population.

(PNG)

S5 Fig. Distribution for node degrees. Degree distribution in the simulated and real population of 137 graphs. The shaded region corresponds to standard deviations across graphs in the population.

(PNG)

Author Contributions

Conceptualization: Rohit Yadav, François-Xavier Dupé, Sylvain Takerkart, Guillaume Auzias.

Data curation: Rohit Yadav, Sylvain Takerkart, Guillaume Auzias.

Formal analysis: Rohit Yadav, François-Xavier Dupé, Sylvain Takerkart, Guillaume Auzias.

Funding acquisition: François-Xavier Dupé, Guillaume Auzias.

Investigation: Rohit Yadav, François-Xavier Dupé, Sylvain Takerkart, Guillaume Auzias.

Methodology: Rohit Yadav, François-Xavier Dupé, Sylvain Takerkart, Guillaume Auzias.

Project administration: Guillaume Auzias.

Resources: Guillaume Auzias.

Software: Rohit Yadav, Guillaume Auzias.

Supervision: François-Xavier Dupé, Sylvain Takerkart, Guillaume Auzias.

Validation: Guillaume Auzias.

Visualization: Rohit Yadav, Guillaume Auzias.

Writing – original draft: Rohit Yadav, François-Xavier Dupé, Sylvain Takerkart, Guillaume Auzias.

Writing – review & editing: Rohit Yadav, François-Xavier Dupé, Sylvain Takerkart, Guillaume Auzias.

References

1. Van Essen DC, Glasser MF, Dierker DL, Harwell JW, Coalson TS. Parcellations and hemispheric asymmetries of human cerebral cortex analyzed on surface-based atlases. *Cerebral cortex* (New York, NY: 1991). 2012; 22(10):2241–62. <https://doi.org/10.1093/cercor/bhr291> PMID: 22047963
2. Auzias G, Colliot O, Glaunes JA, Perrot M, Mangin JF, Trouve A, et al. Diffeomorphic Brain Registration Under Exhaustive Sulcal Constraints. *IEEE Transactions on Medical Imaging*. 2011; 30(6):1214–1227. <https://doi.org/10.1109/TMI.2011.2108665> PMID: 21278014
3. Fischl B, Sereno MI, Tootell RB, Dale AM. High-resolution intersubject averaging and a coordinate system for the cortical surface. *Human brain mapping*. 1999; 8(4):272–284. [https://doi.org/10.1002/\(SICI\)1097-0193\(1999\)8:4<272::AID-HBM10>3.0.CO;2-4](https://doi.org/10.1002/(SICI)1097-0193(1999)8:4<272::AID-HBM10>3.0.CO;2-4) PMID: 10619420
4. Lyu I, Kang H, Woodward ND, Styner MA, Landman BA. Hierarchical spherical deformation for cortical surface registration. *Medical Image Analysis*. 2019; 57:72–88. <https://doi.org/10.1016/j.media.2019.06.013> PMID: 31280090
5. Robinson EC, Garcia K, Glasser MF, Chen Z, Coalson TS, Makropoulos A, et al. Multimodal surface matching with higher-order smoothness constraints. *NeuroImage*. 2018; 167:453–465. <https://doi.org/10.1016/j.neuroimage.2017.10.037> PMID: 29100940

6. Devlin JT, Poldrack RA. In praise of tedious anatomy. *NeuroImage*. 2007; 37(4):1033–41; discussion 1050–8. <https://doi.org/10.1016/j.neuroimage.2006.09.055> PMID: 17870621
7. Van Essen DC, Dierker DL. On navigating the human cerebral cortex: response to 'in praise of tedious anatomy'. *NeuroImage*. 2007; 37(4):1050–4; discussion 1066–8. <https://doi.org/10.1016/j.neuroimage.2007.02.021> PMID: 17766148
8. Armstrong E, Schleicher A, Omran H, Curtis M, Zilles K. The ontogeny of human gyrification. *Cerebral cortex*. 1995; 5(1):56–63. <https://doi.org/10.1093/cercor/5.1.56> PMID: 7719130
9. Cachia A, Paillère-Martinot ML, Galinowski A, Januel D, de Beaurepaire R, Bellivier F, et al. Cortical folding abnormalities in schizophrenia patients with resistant auditory hallucinations. *Neuroimage*. 2008; 39(3):927–935. <https://doi.org/10.1016/j.neuroimage.2007.08.049> PMID: 17988891
10. Im K, Lee JM, Yoon U, Shin YW, Hong SB, Kim IY, et al. Fractal dimension in human cortical surface: multiple regression analysis with cortical thickness, sulcal depth, and folding area. *Human brain mapping*. 2006; 27(12):994–1003. <https://doi.org/10.1002/hbm.20238> PMID: 16671080
11. Mangin JF, Riviere D, Cachia A, Duchesnay E, Cointepas Y, Papadopoulos-Orfanos D, et al. Object-based morphometry of the cerebral cortex. *IEEE transactions on medical imaging*. 2004; 23(8):968–982. <https://doi.org/10.1109/TMI.2004.831204> PMID: 15338731
12. Duchesnay E, Cachia A, Roche A, Rivière D, Cointepas Y, Papadopoulos-Orfanos D, et al. Classification based on cortical folding patterns. *IEEE transactions on medical imaging*. 2007; 26(4):553–65. <https://doi.org/10.1109/TMI.2007.892501> PMID: 17427742
13. Auzias G, Viellard M, Takerkart S, Villeneuve N, Poinso F, Da Fonséca D, et al. Atypical sulcal anatomy in young children with autism spectrum disorder. *NeuroImage: Clinical*. 2014; 4:593–603. <https://doi.org/10.1016/j.nicl.2014.03.008> PMID: 24936410
14. Pizzagalli F, Auzias G, Yang Q, Mathias SR, Faskowitz J, Boyd JD, et al. The reliability and heritability of cortical folds and their genetic correlations across hemispheres. *Communications Biology*. 2020; 3(1):1–12. <https://doi.org/10.1038/s42003-020-01163-1> PMID: 32934300
15. Im K, Jo HJ, Mangin JF, Evans AC, Kim SI, Lee JM. Spatial distribution of deep sulcal landmarks and hemispherical asymmetry on the cortical surface. *Cerebral cortex* (New York, NY: 1991). 2010; 20(3):602–11. <https://doi.org/10.1093/cercor/bhp127> PMID: 19561060
16. Auzias G, Brun L, Deruelle C, Coulon O. Deep sulcal landmarks: Algorithmic and conceptual improvements in the definition and extraction of sulcal pits. *NeuroImage*. 2015; 111:12–25. <https://doi.org/10.1016/j.neuroimage.2015.02.008> PMID: 25676916
17. Im K, Pienaar R, Lee JM, Seong JK, Choi YY, Lee KH, et al. Quantitative comparison and analysis of sulcal patterns using sulcal graph matching: a twin study. *Neuroimage*. 2011; 57(3):1077–1086. <https://doi.org/10.1016/j.neuroimage.2011.04.062> PMID: 21596139
18. Takerkart S, Auzias G, Brun L, Coulon O. Structural graph-based morphometry: A multiscale search-light framework based on sulcal pits. *Medical Image Analysis*. 2017; 35:32–45. <https://doi.org/10.1016/j.media.2016.04.011> PMID: 27310172
19. Rivière D, Mangin JF, Papadopoulos-Orfanos D, Martinez JM, Frouin V, Régis J. Automatic recognition of cortical sulci of the human brain using a congregation of neural networks. *Medical image analysis*. 2002; 6(2):77–92. [https://doi.org/10.1016/S1361-8415\(02\)00052-X](https://doi.org/10.1016/S1361-8415(02)00052-X) PMID: 12044997
20. Borne L, Rivière D, Mancip M, Mangin JF. Automatic labeling of cortical sulci using patch-or CNN-based segmentation techniques combined with bottom-up geometric constraints. *Medical Image Analysis*. 2020; 62:101651. <https://doi.org/10.1016/j.media.2020.101651> PMID: 32163879
21. Behnke KJ, Rettmann ME, Pham DL, Shen D, Resnick SM, Davatzikos C, et al. Automatic classification of sulcal regions of the human brain cortex using pattern recognition. In: *Proceedings of SPIE*. vol. 5032; 2003. p. 1499–1510.
22. Lyu I, Bao S, Hao L, Yao J, Miller JA, Voorhies W, et al. Labeling lateral prefrontal sulci using spherical data augmentation and context-aware training. *NeuroImage*. 2021; 229:117758. <https://doi.org/10.1016/j.neuroimage.2021.117758> PMID: 33497773
23. Voorhies WI, Miller JA, Yao JK, Bunge SA, Weiner KS. Cognitive insights from tertiary sulci in prefrontal cortex. *Nature Communications*. 2021; 12(1):5122. <https://doi.org/10.1038/s41467-021-25162-w> PMID: 34433806
24. Sprung-Much T, Petrides M. Morphology and Spatial Probability Maps of the Horizontal Ascending Ramus of the Lateral Fissure. *Cerebral Cortex*. 2020; 30(3):1586–1602. <https://doi.org/10.1093/cercor/bhz189> PMID: 31667522
25. Willbrand EH, Parker BJ, Voorhies WI, Miller JA, Lyu I, Hallock T, et al. Uncovering a tripartite landmark in posterior cingulate cortex. *Science Advances*. 2022; 8(36):eabn9516. <https://doi.org/10.1126/sciadv.abn9516> PMID: 36070384

26. Eickhoff SB, Yeo BTT, Genon S. Imaging-based parcellations of the human brain. *Nature Reviews Neuroscience*. 2018; 19(11):672–686. <https://doi.org/10.1038/s41583-018-0071-7> PMID: 30305712
27. Glasser MF, Coalson TS, Robinson ESJ, Hacker CD, Harwell J, Yacoub E, et al. A multi-modal parcellation of human cerebral cortex. *Nature*. 2016; 536(7615):171–178. <https://doi.org/10.1038/nature18933> PMID: 27437579
28. Le Guen Y, Auzias G, Leroy F, Noulhiane M, Dehaene-Lambertz G, Duchesnay E, et al. Genetic Influence on the Sulcal Pits: On the Origin of the First Cortical Folds. *Cerebral Cortex*. 2017;(2015):1–12. <https://doi.org/10.1093/cercor/bhx098>
29. Kaltenmark I, Deruelle C, Brun L, Lefèvre J, Coulon O, Auzias G. Cortical inter-subject correspondences with optimal group-wise parcellation and sulcal pits labeling. *Medical Image Analysis*. 2020;.
30. Leordeanu M, Hebert M. A spectral technique for correspondence problems using pairwise constraints. 2005;.
31. Morton SU, Maleyeff L, Wypij D, Yun HJ, Newburger JW, Bellinger DC, et al. Abnormal Left-Hemispheric Sulcal Patterns Correlate with Neurodevelopmental Outcomes in Subjects with Single Ventricular Congenital Heart Disease. *Cerebral Cortex*. 2019; p. 1–12. <https://doi.org/10.1093/cercor/bhz101>
32. Im K, Guimaraes A, Kim Y, Cottrill E, Gagoski B, Rollins C, et al. Quantitative Folding Pattern Analysis of Early Primary Sulci in Human Fetuses with Brain Abnormalities. *American Journal of Neuroradiology*. 2017; 38(7):1449–1455. <https://doi.org/10.3174/ajnr.A5217> PMID: 28522661
33. Meng Y, Li G, Wang L, Lin W, Gilmore JH, Shen D. Discovering cortical sulcal folding patterns in neonates using large-scale dataset. *Human brain mapping*. 2018; 39(9):3625–3635. <https://doi.org/10.1002/hbm.24199> PMID: 29700891
34. Wang B, Mezlini AM, Demir F, Fiume M, Tu Z, Brudno M, et al. Similarity network fusion for aggregating data types on a genomic scale. *Nature methods*. 2014; 11(3):333–337. <https://doi.org/10.1038/nmeth.2810> PMID: 24464287
35. Buskalic N, Dupé FX, Takerkart S, Auzias G. Labelling Sulcal Graphs Across Individuals Using Multi-graph Matching. In: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI); 2021. p. 1486–1490. <https://doi.org/10.1109/ISBI48211.2021.9434035>
36. Yadav R, Dupé FX, Takerkart S, Auzias G. On The Relevance of Multi-Graph Matching for Sulcal Graphs. In: 2022 IEEE International Conference on Image Processing (ICIP); 2022. p. 2536–2540.
37. Loiola CF, Silva DAd, Galati EAB. Phlebotomine fauna (Diptera: Psychodidae) and species abundance in an endemic area of American cutaneous leishmaniasis in southeastern Minas Gerais, Brazil. *Memórias do Instituto Oswaldo Cruz*. 2007; 102:581–585. <https://doi.org/10.1590/S0074-02762007005000050> PMID: 17710302
38. Lawler EL. The quadratic assignment problem. *Management science*. 1963; 9(4):586–599. <https://doi.org/10.1287/mnsc.9.4.586>
39. Koopmans TC, Beckmann M. Assignment problems and the location of economic activities. *Econometrica: journal of the Econometric Society*. 1957; p. 53–76. <https://doi.org/10.2307/1907742>
40. Yan J, Yin XC, Lin W, Deng C, Zha H, Yang X. A short survey of recent advances in graph matching. In: Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval; 2016. p. 167–174.
41. Cour T, Srinivasan P, Shi J. Balanced Graph Matching. In: Advances in Neural Information Processing Systems 19. MIT Press; 2007. p. 313–320. Available from: <http://papers.nips.cc/paper/2960-balanced-graph-matching.pdf>.
42. Leordeanu M, Hebert M, Sukthankar R. An Integer Projected Fixed Point Method for Graph Matching and MAP Inference. In: Advances in Neural Information Processing Systems 22. Curran Associates, Inc.; 2009. p. 1114–1122. Available from: <http://papers.nips.cc/paper/3756-an-integer-projected-fixed-point-method-for-graph-matching-and-map-inference.pdf>.
43. Hutchison D, Kanade T, Kittler J, Kleinberg JM, Mattern F, Mitchell JC, et al. Reweighted Random Walks for Graph Matching. In: Computer Vision—ECCV 2010. vol. 6315. Berlin, Heidelberg: Springer Berlin Heidelberg; 2010. p. 492–505. Available from: http://link.springer.com/10.1007/978-3-642-15555-0_36.
44. Zhang Z, Xiang Y, Wu L, Xue B, Nehorai A. KerGM: Kernelized Graph Matching. In: Advances in Neural Information Processing Systems 32; 2019. p. 3335–3346. Available from: <http://papers.nips.cc/paper/8595-kergm-kernelized-graph-matching.pdf>.
45. Pachauri D, Kondor R, Singh V. Solving the multi-way matching problem by permutation synchronization. In: Burges CJC, Bottou L, Welling M, Ghahramani Z, Weinberger KQ, editors. Advances in Neural Information Processing Systems 26. Curran Associates, Inc.; 2013. p. 1860–1868. Available from: <http://papers.nips.cc/paper/4987-solving-the-multi-way-matching-problem-by-permutation-synchronization.pdf>.
46. Chen Y, Guibas LJ, Huang QX. Near-optimal joint object matching via convex relaxation. arXiv preprint arXiv:14021473. 2014;.

47. Wang Q, Zhou X, Daniilidis K. Multi-image semantic matching by mining consistent features. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2018. p. 685–694.
48. Hu N, Huang Q, Thibert B, Guibas LJ. Distributable consistent multi-object matching. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2018. p. 2463–2471.
49. Bernard F, Thunberg J, Swoboda P, Theobalt C. Hippo: Higher-order projected power iterations for scalable multi-matching. In: Proceedings of the IEEE/CVF International Conference on Computer Vision; 2019. p. 10284–10293.
50. Zhou X, Zhu M, Daniilidis K. Multi-image matching via fast alternating minimization. In: Proceedings of the IEEE International Conference on Computer Vision; 2015. p. 4032–4040.
51. Yan J, Cho M, Zha H, Yang X, Chu SM. Multi-Graph Matching via Affinity Optimization with Graduated Consistency Regularization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2016; 38(6):1228–1242. <https://doi.org/10.1109/TPAMI.2015.2477832> PMID: 26372208
52. Yan J, Tian Y, Zha H, Yang X, Zhang Y, Chu SM. Joint optimization for consistent multiple graph matching. In: Proceedings of the IEEE international conference on computer vision; 2013. p. 1649–1656.
53. Yan J, Li Y, Liu W, Zha H, Yang X, Chu SM. Graduated consistency-regularized optimization for multi-graph matching. In: European Conference on Computer Vision. Springer; 2014. p. 407–422.
54. Yan J, Wang J, Zha H, Yang X, Chu S. Consistency-Driven Alternating Optimization for Multigraph Matching: A Unified Approach. *IEEE Transactions on Image Processing*. 2015; 24(3):994–1009. <https://doi.org/10.1109/TIP.2014.2387386> PMID: 25576568
55. Zanfir A, Sminchisescu C. Deep learning of graph matching. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2018. p. 2684–2693.
56. Wang R, Yan J, Yang X. Combinatorial learning of robust deep graph matching: an embedding based approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2020;.
57. Wang R, Yan J, Yang X. Learning combinatorial embedding networks for deep graph matching. In: Proceedings of the IEEE/CVF international conference on computer vision; 2019. p. 3056–3065.
58. Wang R, Yan J, Yang X. Neural graph matching network: Learning lawler’s quadratic assignment problem with extension to hypergraph and multiple-graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2021;.
59. Yu T, Wang R, Yan J, Li B. Deep Latent Graph Matching. In: Meila M, Zhang T, editors. Proceedings of the 38th International Conference on Machine Learning. vol. 139 of Proceedings of Machine Learning Research. PMLR; 2021. p. 12187–12197. Available from: <https://proceedings.mlr.press/v139/yu21d.html>.
60. Yu T, Wang R, Yan J, Li B. Learning deep graph matching with channel-independent embedding and hungarian attention. In: International conference on learning representations; 2019.
61. Wang R, Yan J, Yang X. Graduated assignment for joint multi-graph matching and clustering with application to unsupervised graph matching network learning. *Advances in Neural Information Processing Systems*. 2020; 33:19908–19919.
62. Rolínek M, Swoboda P, Zietlow D, Paulus A, Musil V, Martius G. Deep graph matching via blackbox differentiation of combinatorial solvers. In: European Conference on Computer Vision. Springer; 2020. p. 407–424.
63. Rubinstein M, Joulín A, Kopf J, Liu C. Unsupervised joint object discovery and segmentation in internet images. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2013. p. 1939–1946.
64. Faktor A, Irani M. “Clustering by Composition”—Unsupervised Discovery of Image Categories. *IEEE transactions on pattern analysis and machine intelligence*. 2013; 36(6):1092–1106. <https://doi.org/10.1109/TPAMI.2013.251>
65. Tron R, Zhou X, Esteves C, Daniilidis K. Fast Multi-image Matching via Density-Based Clustering. In: 2017 IEEE International Conference on Computer Vision (ICCV). Venice: IEEE; 2017. p. 4077–4086. Available from: <http://ieeexplore.ieee.org/document/8237699/>.
66. Zhou T, Jae Lee Y, Yu SX, Efros AA. Flowweb: Joint image set alignment by weaving consistent, pixel-wise correspondences. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2015. p. 1191–1200.
67. Maset E, Arrigoni F, Fusiello A. Practical and Efficient Multi-view Matching. In: 2017 IEEE International Conference on Computer Vision (ICCV). Venice: IEEE; 2017. p. 4578–4586. Available from: <http://ieeexplore.ieee.org/document/8237751/>.
68. Peyré G, Chizat L, Vialard FX, Solomon J. Quantum entropic regularization of matrix-valued optimal transport. *European Journal of Applied Mathematics*. 2019; 30(6):1079–1102. <https://doi.org/10.1017/S0956792517000274>

69. Swoboda P, Mokarian A, Theobalt C, Bernard F, et al. A convex relaxation for multi-graph matching. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2019. p. 11156–11165.
70. Shi Y, Li S, Lerman G. Robust multi-object matching via iterative reweighting of the graph connection Laplacian. *Advances in Neural Information Processing Systems*. 2020; 33:15243–15253.
71. Hastie T, Mazumder R, Lee JD, Zadeh R. Matrix completion and low-rank SVD via fast alternating least squares. *The Journal of Machine Learning Research*. 2015; 16(1):3367–3402. PMID: [31130828](https://pubmed.ncbi.nlm.nih.gov/31130828/)
72. Cabral R, De la Torre F, Costeira JP, Bernardino A. Unifying nuclear norm and bilinear factorization approaches for low-rank matrix decomposition. In: Proceedings of the IEEE international conference on computer vision; 2013. p. 2488–2495.
73. Eckstein J, Bertsekas DP. On the Douglas—Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*. 1992; 55(1):293–318. <https://doi.org/10.1007/BF01581204>
74. Blaser R, Fryzlewicz P, Blaser R, Fryzlewicz P. *Random Rotation Ensembles*. 2016;.
75. Lefèvre J, Pepe A, Muscato J, De Guio F, Girard N, Auzias G, et al. SPANOL (SPectral ANalysis of Lobes): A Spectral Clustering Framework for Individual and Group Parcellation of Cortical Surfaces in Lobes. *Frontiers in Neuroscience*. 2018; 12:1–14. <https://doi.org/10.3389/fnins.2018.00354> PMID: [29904338](https://pubmed.ncbi.nlm.nih.gov/29904338/)
76. Von Mises R. *Mathematical theory of probability and statistics*. Academic press; 1964.
77. Marcus DS, Wang TH, Parker J, Csernansky JG, Morris JC, Buckner RL. Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *Journal of cognitive neuroscience*. 2007; 19(9):1498–1507. <https://doi.org/10.1162/jocn.2007.19.9.1498> PMID: [17714011](https://pubmed.ncbi.nlm.nih.gov/17714011/)
78. Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*. 1987; 20:53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
79. Leroy F, Cai Q, Bogart SL, Dubois J, Coulon O, Monzalvo K, et al. New human-specific brain landmark: The depth asymmetry of superior temporal sulcus. *Proceedings of the National Academy of Sciences*. 2015; 112(4):1208–1213. <https://doi.org/10.1073/pnas.1412389112> PMID: [25583500](https://pubmed.ncbi.nlm.nih.gov/25583500/)
80. Im K, Choi YY, Yang JJ, Lee KH, Kim SI, Grant PE, et al. The relationship between the presence of sulcal pits and intelligence in human brains. *NeuroImage*. 2011; 55(4):1490–1496. <https://doi.org/10.1016/j.neuroimage.2010.12.080> PMID: [21224005](https://pubmed.ncbi.nlm.nih.gov/21224005/)
81. Brun L, Auzias G, Viellard M, Villeneuve N, Girard N, Poinso F, et al. Localized Misfolding Within Broca's Area as a Distinctive Feature of Autistic Disorder. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*. 2016; 1(2):160–168. <https://doi.org/10.1016/j.bpsc.2015.11.003> PMID: [29560874](https://pubmed.ncbi.nlm.nih.gov/29560874/)
82. Lefrere A, Auzias G, Favre P, Kaltenmark I, Houenou J, Piguat C, et al. Global and local cortical folding alterations are associated with neurodevelopmental subtype in bipolar disorders: a sulcal pits analysis. *Journal of Affective Disorders*. 2023; 325:224–230. <https://doi.org/10.1016/j.jad.2022.12.156> PMID: [36608853](https://pubmed.ncbi.nlm.nih.gov/36608853/)
83. Li XW, Jiang YH, Wang W, Liu XX, Li ZY. Brain morphometric abnormalities in boys with attention-deficit/hyperactivity disorder revealed by sulcal pits-based analyses. *CNS Neuroscience & Therapeutics*. 2021; 27(3):299–307. <https://doi.org/10.1111/cns.13445> PMID: [32762149](https://pubmed.ncbi.nlm.nih.gov/32762149/)
84. LeCun Y, Bengio Y, Hinton G. Deep learning. *nature*. 2015; 521(7553):436–444. <https://doi.org/10.1038/nature14539> PMID: [26017442](https://pubmed.ncbi.nlm.nih.gov/26017442/)
85. Scarselli F, Gori M, Tsoi AC, Hagenbuchner M, Monfardini G. The Graph Neural Network Model. *IEEE Transactions on Neural Networks*. 2009; 20(1):61–80. <https://doi.org/10.1109/TNN.2008.2005605> PMID: [19068426](https://pubmed.ncbi.nlm.nih.gov/19068426/)
86. Xu K, Hu W, Leskovec J, Jegelka S. How Powerful are Graph Neural Networks? In: *International Conference on Learning Representations*; 2019.
87. Fey M, Lenssen JE, Morris C, Masci J, Kriege NM. Deep Graph Matching Consensus. In: *International Conference on Learning Representations*; 2020.
88. Awate SP, Yushkevich Pa, Song Z, Licht DJ, Gee JC. Cerebral cortical folding analysis with multivariate modeling and testing: Studies on gender differences and neonatal development. *NeuroImage*. 2010; 53(2):450–459. <https://doi.org/10.1016/j.neuroimage.2010.06.072> PMID: [20630489](https://pubmed.ncbi.nlm.nih.gov/20630489/)
89. Rabiei H, Richard F, Coulon O, Lefèvre J. Local Spectral Analysis of the Cerebral Cortex: New Gyrfication Indices. *IEEE Transactions on Medical Imaging*. 2017; 36(3):838–848. <https://doi.org/10.1109/TMI.2016.2633393> PMID: [27913336](https://pubmed.ncbi.nlm.nih.gov/27913336/)

90. Honey CJ, Kötter R, Breakspear M, Sporns O. Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *Proceedings of the National Academy of Sciences*. 2007; 104(24):10240–10245. <https://doi.org/10.1073/pnas.0701519104> PMID: 17548818
91. Friston KJ. Functional and effective connectivity in neuroimaging: a synthesis. *Human brain mapping*. 1994; 2(1-2):56–78. <https://doi.org/10.1002/hbm.460020107>
92. Hsu HHH, Shen Y, Cremers D. A Graph Is More Than Its Nodes: Towards Structured Uncertainty-Aware Learning on Graphs. *arXiv preprint arXiv:221015575*. 2022;.
93. Dupé FX, Yadav R, Auzias G, Takerkart S. Kernelized multi-graph matching. In: *ACML 2022*; 2022.
94. Lin Y, Yang M, Yu J, Hu P, Zhang C, Peng X. Graph Matching with Bi-level Noisy Correspondence. In: *ICCV*; 2023.
95. Nurlanov Z, Schmidt FR, Bernard F. Universe Points Representation Learning for Partial Multi-Graph Matching. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2023;37(2):1984–1992.
96. Glasser MF, Smith SM, Marcus DS, Andersson JL, Auerbach EJ, Behrens TE, et al. The human connectome project's neuroimaging approach. *Nature neuroscience*. 2016; 19(9):1175–1187. <https://doi.org/10.1038/nn.4361> PMID: 27571196