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. 2023. hal-03963359v1

HAL Id: hal-03963359

https://hal.science/hal-03963359v1

Preprint submitted on 30 Jan 2023 (v1), last revised 7 Feb 2024 (v3)

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.

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

R.Yadav[⋆‡†], F.X. Dupé[†], S. Takerkart[⋆], G. Auzias[⋆]
[⋆]Institut de Neurosciences de la Timone UMR 7289, Aix-Marseille Université, CNRS
[‡]Aix Marseille Université, Institut Marseille Imaging, Marseille, France
[†]Laboratoire d'Informatique et Systèmes UMR 7020, Aix-Marseille Université, CNRS

January 30, 2023

Abstract

Population-wise matching of the cortical fold is necessary to identify biomarkers of neurological or psychiatric disorders. The difficulty comes from the massive inter-individual variations in the morphology and spatial organization of the folds. This task is challenging at both methodological and conceptual levels. 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 multi-graph matching framework in this context. We then introduce a procedure to generate populations of artificial sulcal graphs, which allows us benchmarking 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 to obtain a population-wise consistent labeling of cortical folds at the sulcal basins level.

Keywords: brain, sulcal graphs, multi-graph matching, sulcal pits, MRI.

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 mathematical spaces. Comparing such features thus requires to address the *methodological* problem of transferring them into a common space. The task of transferring information from one brain to another or to a common space consists in defining spatial correspondences

across these objects by compensating for their variations in their respective geometry. The challenge lies in the massive inter-individual variations of the morphology of the brain and in particular the geometry of the cortical surface, which make the identification of such spatial correspondences an ill-posed problem. As a consequence, any solution to this problem inevitably requires to introduce additional constraints based on assumptions on the biological 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 thus be considered as a source of bias in the analysis Van Essen, Glasser, Dierker, Harwell, and Coalson (2012).

One widely used 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. 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, and tools to enable comparison *across atlases* are then required (Devlin & Poldrack, 2007; Van Essen & Dierker, 2007). 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 approach (Fischl, Sereno, Tootell, & Dale, 1999) 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 anatomo-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 might be seen as an over simplification of the problem since potentially relevant geometrical information is not taken into account.

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 (Armstrong, Schleicher, Omran, Curtis, & Zilles, 1995; Cachia et al., 2008; Im et al., 2006). Although these measures capture relevant morphological features, they do not explicitly reflect the topology, i.e the spatial relationships between sulci. Mangin et al. (2004) introduced an analysis framework based on the automatic extraction and labeling of the sulci allowing to characterize their shape, size and pattern in terms of e.g. sulcus area, depth and length. This representation of the

cortical geometry has been used for instance to characterize populations of healthy subjects (Duchesnay et al., 2007), to quantify potential deviations from normal populations in various conditions such as schizophrenia (Cachia et al., 2008) and autism spectrum disorder (Auzias et al., 2014), or to estimate the heritability of the folding patterns (Pizzagalli et al., 2020). 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 Auzias, Brun, Deruelle, and Coulon (2015); Im et al. (2010), 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, Im et al. (2011); Takerkart, Auzias, Brun, and Coulon (2017) 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. Figure 1 illustrates this decomposition of the cortical folds into sulcal basins allowing to represent this complex geometry as a sulcal graph.

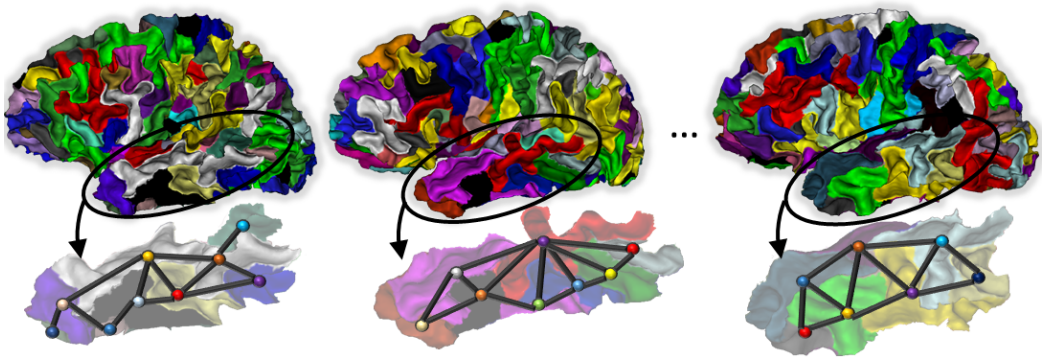

Figure 1: Example of sulcal graphs from three individual brains, superimposed with the underlying decomposition of the cortical surface in sulcal basins. Sulcal basins are shown in different colours, and their corresponding node in the graph are represented as spherical dots in the lower panel. The color 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.

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 biologically relevant *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 individuals, and with respect to the methodological problem of graph matching.

1.3.1 Unsupervised comparison and matching of sulcal graphs

Comparing brains using sulcal graphs is highly relevant because all the geometrical information about the macroscopic cortical folding can be encoded into 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 (Behnke et al., 2003; Borne, Rivière, Mancip, & Mangin, 2020; Rivière et al., 2002). All these works focused at the scale of sulci, i.e. considering large folds consisting of several of our sulcal basins. To our knowledge, only Lyu et al. (2021) 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 Voorhies, Miller, Yao, Bunge, and Weiner (2021). 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 Sprung-Much and Petrides (2020) and their relationship with underlying function Willbrand et al. (2022). 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 (Eickhoff, Yeo, & Genon, 2018), and even the most advanced multi-modal atlas (Glasser et al., 2016) 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 Im et al. (2010) and consisted in computing a map of the spatial density of sulcal pits across a population of subjects. This density map was computed by accumulating the pits from the different individuals in each vertex of an average surface after aligning the folds using a registration technique. A watershed algorithm was 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 was 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.

Auzias et al. (2015); Im et al. (2010); Le Guen et al. (2017), 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 this paper. The main limitation of this approach is that the labeling is driven only by the coordinates of the sulcal pit.

Kaltenmark et al. (2020) 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 (Kaltenmark et al., 2020) 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 relationships between adjacent sulcal basins and pits are never 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 Im et al. (2011), the spectral graph matching technique (Leordeanu & Hebert, 2005) 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 Morton et al. (2019) or for matching individuals onto an atlas in Im et al. (2017).

In the work by Meng et al. (2018), a population of 677 neonates was analyzed based on a sulcal graph comparison method similar to the one of Im et al. (2011). The authors proposed to use different features of the sulcal pits such as the pit position, the pit depth, the basin area, the basin boundary and the pit local connectivity to construct different similarity matrices, one per feature, and merge them into a single one using a matrix fusion technique (B. Wang et al., 2014). 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 previous studies relied only on *pairwise* graph matching techniques. Comparing pairs of graphs independently, 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 basins 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 (Buskulic, Dupé, Takerkart, & Auzias, 2021), 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 the context of sulcal graphs. In Yadav, Dupé, Takerkart, and Auzias (2022), we provided a proof of concept of the relevance of multi-graph matching techniques in this context. 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. Secondly, we benchmark a selection of recently published multi-graph matching techniques against the best pairwise technique for this task (identified in from Buskulic et al. (2021)), and report variations in performances that would clearly impact potential real-world applications, e.g in a clinical context. Finally, 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 dataset of 137 subjects. In addition, our experiments demonstrate the feasibility of comparing a large population of sulcal graphs based on multi-graph matching techniques, in fully acceptable computing times. All the source code and data will be shared openly upon publication at `https://www.github.com/gauzias/sulcal_graphs_matching`.

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 $\mathscr{G}_1 \dots \mathscr{G}_N$, 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 $V_q = \{v_1, v_2, \ldots, v_{n_q}\}$ 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. $A_q^V = \{a_{v_1}^V, a_{v_2}^V, \ldots, a_{v_{n_q}}^V\}$ is the set of attributes associated to each node in V_q, and $A_q^E = \{a_{e_1}^E, a_{e_2}^E, \ldots, a_{e_{e_q}}^E\}$ 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 Fischl et al. (1999) for details). The matching is computed in this common 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.

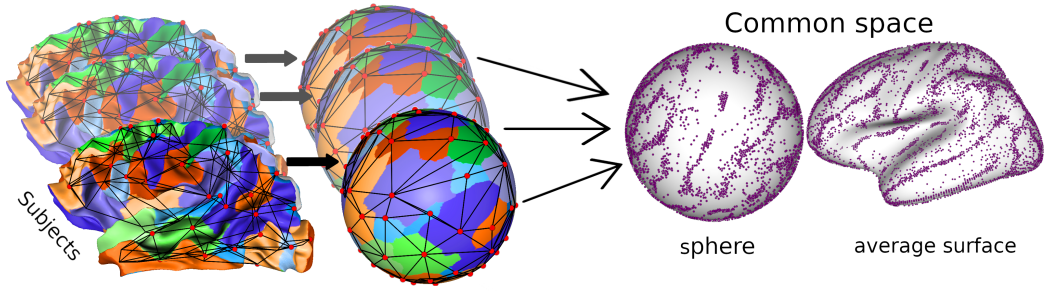

Figure 2: The sulcal graph from each subject is transferred onto a common sphere using the inflation and spherical registration tools from freesurfer. The sulcal graphs from every subjects can then be mapped onto either the common sphere or onto an average surface for visualization. Note that the spatial 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.

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 is usually divided into two categories: exact and partial matching. Exact matching methods consider graph matching to be a special case of the graph isomorphism problem. It aims at finding the bijection between two graphs, which implies that both the nodes and edges of the different graphs are strictly matched. This requirement is too strict for most real-world tasks and in particular in our context where the number of nodes and edges varies across graphs. Therefore, we focus on the partial matching problem. This problem can be formulated as a Quadratic Assignment Problem (QAP) (Loiola, Silva, & Galati, 2007). Although different forms of QAP exist, the vast majority of the literature has focused on Lawler's QAP (Lawler, 1963). 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:

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

where $vec(\mathbf{X}_{12})$ 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 $\mathbf{\Phi}_{12} \in [0,1]^{n_1 n_2 \times n_1 n_2}$ is the *affinity matrix* that is given as an input. The diagonal entries of $\mathbf{\Phi}_{12}$ 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 $\mathbf{\Phi}_{12}$ 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 (F. Zhou & De la Torre, 2015) that is a special case of Lawler's QAP:

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

where $\mathbf{\Psi}_{12} \in [0,1]^{n_1 \times n_2}$ 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} 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 $\mathbf{\Phi}_{12}$ into three smaller matrices $\mathbf{\Psi}_{12}$, $\mathbf{A}_1$ 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 Eq.(1) and (2) and provide approximate solutions. Various approaches have been proposed to relax these problems, leading to a variety of graph matching methods. Discussing these methods is beyond the scope of this work but we refer interested readers to the review Yan, Yin, et al. (2016).

Going back to our specific context, we reported in Buskulic et al. (2021) a benchmark of the pairwise methods SMAC (Spectral Matching with Affine Constraints) (Cour, Srinivasan, & Shi, 2007), IPFP (Integer Projected Fixed Point algorithm) (Leordeanu, Hebert, & Sukthankar, 2009), RRWM (Reweighted Random Walks for graph Matching) (Hutchison et al., 2010), and KerGM (Kernelized Graph Matching) (Zhang, Xiang, Wu, Xue, & Nehorai, 2019). We observed that **KerGM** clearly outperforms the others in our context. Conceptually, **KerGM** well suits sulcal graphs as it relies on Frank-Wolfe optimization that allows to follow an optimisation path that respects the constraint on each step. This induces a robustness to the presence of noise in graphs that is crucial in our context. In the present work, **KerGM** is included in our benchmark as a representative of pairwise approaches. It

is also used to define the initialization of all the multi-graph methods that are introduced in next section.

2.3 The multi-graph matching problem

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

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

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 $m = \sum_{q=1}^{N} |\mathscr{G}_q|$, 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. Multi-graph matching techniques can be divided into three categories as follows.

The first category of approaches explicitly aim at minimizing the rank of the bulk matrix using various approaches (Chen, Guibas, & Huang, 2014; Hu, Huang, Thibert, & Guibas, 2018; Pachauri, Kondor, & Singh, 2013; Q. Wang, Zhou, & Daniilidis, 2018). For instance, (Bernard, Thunberg, Swoboda, & Theobalt, 2019) solves a global optimization problem by using a projected power iterative method, and we detailed further (X. Zhou, Zhu, & Daniilidis, 2015).

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 (Yan, Cho, Zha, Yang, & Chu, 2016; Yan, Cho, et al., 2016; Yan et al., 2014, 2013; Yan, Wang, Zha, Yang, & Chu, 2015).

Finally, the third category corresponds to deep learning approaches that show promising performances in supervised tasks compared to previous methods, but are not suited for unsupervised tasks (Rolínek et al., 2020; R. Wang, Yan, & Yang, 2019, 2020a, 2020b, 2021; Yu, Wang, Yan, & Li, 2019, 2021; Zanfir & Sminchisescu, 2018).

Some other interesting methods exploit the concept of consistency in order to solve the problem of jointly matching multiple images (Faktor & Irani, 2013; Rubinstein, Joulin,

Kopf, & Liu, 2013; Tron, Zhou, Esteves, & Daniilidis, 2017; T. Zhou, Jae Lee, Yu, & Efros, 2015). However, these do not take into account the connectivity of the graphs.

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 reimplementation issues and to ensure the full reproducibility of our results. *(ii) Methods exploiting graph topology.* We selected the methods that take into account the topology of the graph, which is crucial to exploit the spatial adjacency information encoded in the sulcal graphs. *(iii) Scalability.* Since we are interested in performing population studies over large sets of individuals, we excluded methods that do not provide acceptable scalability. *(iv) Unsupervised methods.* Finally, as motivated in the introduction, we focus on unsupervised methods in the present study.

The method that satisfy these selection criteria are **mALS** (X. Zhou et al., 2015), **mSync** (Pachauri et al., 2013) and **CAO** (Yan, Cho, et al., 2016). We provide a detailed description of each of these methods below. In our experiments, these multigraph graph-matching techniques will be compared with the pariwise 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.** (Auzias et al., 2015) and **Kaltenmark et al.** (Kaltenmark et al., 2020).

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** (Yan, Cho, et al., 2016) 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 $\mathscr{G}_q$ 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 $\mathscr{G}_q$ as:

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

where $\|.\|_F$ 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** (Pachauri et al., 2013) consists in estimating a mapping of each $\mathbf{X}_{ij}$ to a common *universe* of assignment matrices, of size d:

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

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

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 **mALS** X. Zhou et al. (2015), the authors formalize the multi-graph matching as the following low rank matrix recovery problem:

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

where, $\langle ., . \rangle$ is the inner product, α controls the weight on sparsity, and $\mathbb{K} = \{\mathbf{\Psi}_{ij}\}_{i,j=1}^{N}$ 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 Hastie, Mazumder, Lee, and Zadeh (2015) and Cabral, De la Torre, Costeira, and Bernardino (2013) for decomposing $\mathbb{X}$ which allows to solve the problem in a lower dimension space using the ADMM method (Eckstein & Bertsekas, 1992).

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 performances 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).

Algorithm 1 Procedure to generate a population of artificial sulcal grahs

Require: $N, n_{ref}, \kappa, \mu_{pert}, \sigma_{pert}, p$
  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 Auzias et al. (2015) 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 dis-

tance 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 Blaser, Fryzlewicz, Blaser, and Fryzlewicz (2016); Lefèvre et al. (2018).

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 (Von Mises, 1964). 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 $B(\nu, \alpha, \beta)$, which is a distribution of non-negative integers. The parameter ν 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 $B(\nu, \alpha, \beta)$ approximates a Gaussian distribution. We describe the setting of these parameters and precise their link with μ and σ of a Gaussian in Appendix A. 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 $\mathbf{\Psi}_{ij}, \mathbf{A}_i, \mathbf{A}_j$ 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 $v \in \mathscr{G}$ and $v' \in \mathscr{G}'$ the affinity value is computed using the kernel defined as $\exp\left(-\gamma^V \left\| a_v^V - a_{v'}^V \right\|_2^2\right)$ and for two edges $e \in \mathscr{G}$ and $e' \in \mathscr{G}'$ the kernel is defined as $\exp\left(-\gamma^E \left\| a_e^E - a_{e'}^E \right\|_2^2\right)$. To estimate appropriate values for γ^V and γ^E we use a heuristic proposed in Takerkart et al. (2017) 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 Appendix B 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*:

$$\textit{Precision} = \frac{\textit{True Positives}}{\textit{True Positives + False Positives}} \in [0, 1] \tag{8}$$

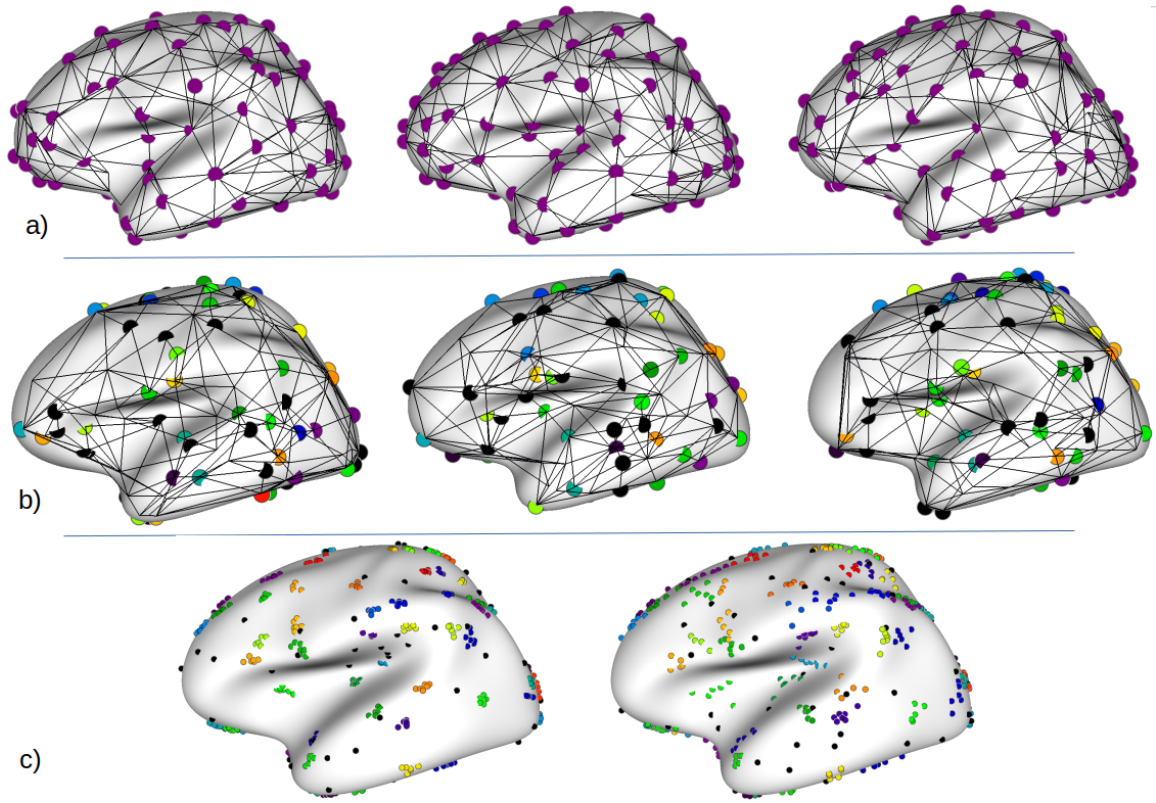

Figure 3: a) Real sulcal graphs from three randomly chosen individuals, and projected on the average surface. b) Simulated graphs randomly chosen for $\kappa = 1000$, showing the ground-truth correspondence across graphs in color. Nodes in black represent the outlier nodes that have no correspondence. 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.

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

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

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 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 of Fig 4 the mean and standard deviation of *Precision*, *Recall* and F_1-*score*, computed across the 10 synthetic populations for each value of κ.

First, we find that two multi-graph matching methods, **mALS** and **mSync**, vastly and consistently outperform **KerGM**, which has been identified as the best pairwise matching method for this task in Buskulic et al. (2021). 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 performances of all methods as the noise increases (decrease of κ), as expected. The performances of the multigraph approaches **mALS** and **mSync** resist much more to this increase in variability than the pairwise approach. The performances of **mSync** are limited more specifically by the lower precision at any level of noise. This suggests that the difference in performances between the two methods are mainly due to the hard constraint on the consistency in **mSync** that seems too restrictive. On the other hand, the recall indicates that **mSync** is more robust to increasing noise than **mALS**, with very close value when $\kappa = 100$. However, **mALS** performs better for lower noise values. Overall, **mALS** shows the best F_1-*score* for every κ values, thanks to a very high precision combined with very good recall. Indeed, the F_1-*score* for **mALS** is above 0.7 even for $\kappa = 200$ which corresponds to a configuration where the noise is quite strong.

Finally, the performances of **CAO** are very low, even lower than the pairwise technique **KerGM**. Such poor performances are 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 Yan, Cho, et al. (2016) could show much higher performances but did not scale with the size of our data.

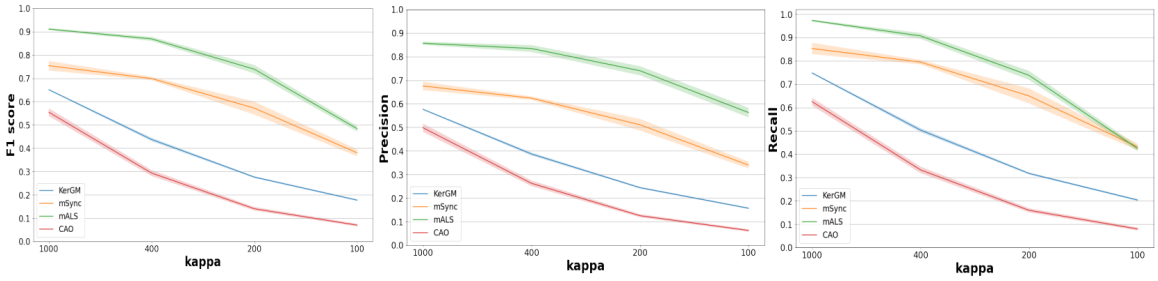

Figure 4: F_1-score, Precision and recall for $\kappa \in [1000, 400, 200, 100]$. For each method, we plot the average across the 10 simulated populations as a line and the standard deviation as the shaded region of the same color.

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 taken from the publicly available database OASIS (Marcus et al., 2007). The preprocessing

of these data (brain tissues segmentation, mesh extraction and sulcal graphs construction) has been detailed in Auzias et al. (2015); Takerkart et al. (2017). 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 Yan, Cho, et al. (2016), we compute the node consistency as follows: *Given* $\mathscr{G}_k \in \{\mathscr{G}_q\}_{q=1}^N$ *and the bulk matrix* $\mathbb{X}$*, for node* $v^k \in \mathscr{G}_k$*, with index* $i(v^k) \in \{1, \ldots, |\mathscr{G}_k|\}$*, its consistency is defined by:*

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

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 Rousseeuw (1987), 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.

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
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

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 three methods **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**. 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**. 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. 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 unlabelled and are represented in black. The clusters seem more compact than for the methods of the second column (**Auzias et al.** and **mSync**) that do not allow any node to remain unlabeled. On the third column (**CAO** and **kerGM**) the matching looks noisy, with clusters overlapping between eachother in almost every cortical location, which illustrates the poor anatomical relevance of the matching.

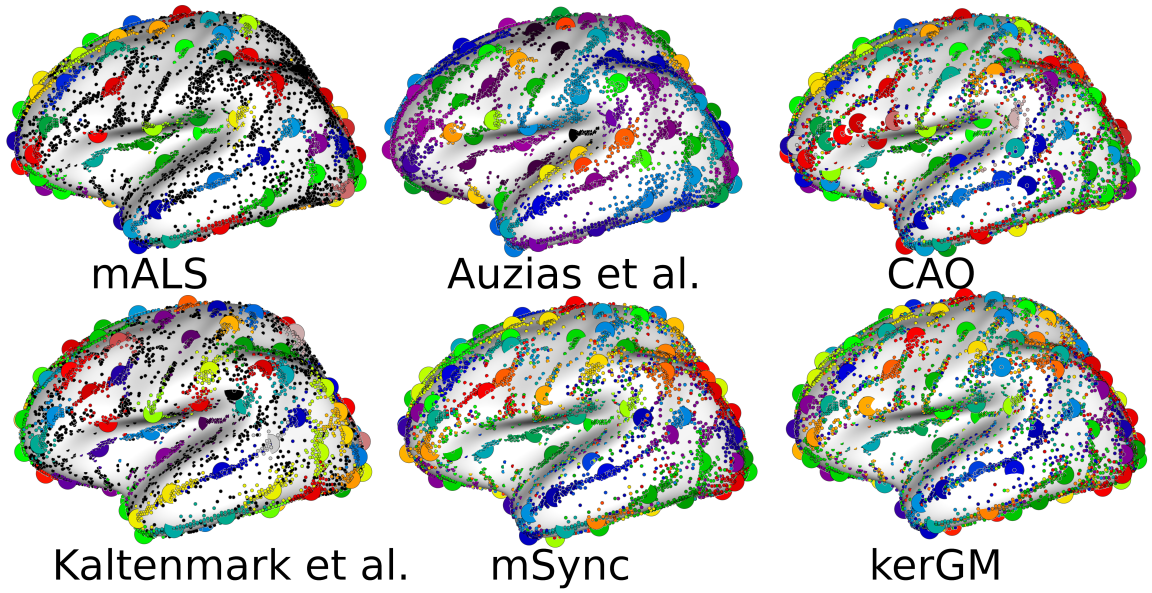

Figure 5: Labeling and corresponding cluster centroids (larger circles) for each method. Dots in black in the first column (for **mALS** and **Kaltenmark et al.**) correspond to unmatched nodes. See text for further description.

For further evaluation of the performances 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 **CAO** and **KerGM**. Together with Table 1, this figure illustrates the poor performances 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. 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 **Auzias et al.**, 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 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 three 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 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. 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 Leroy et al. (2015).

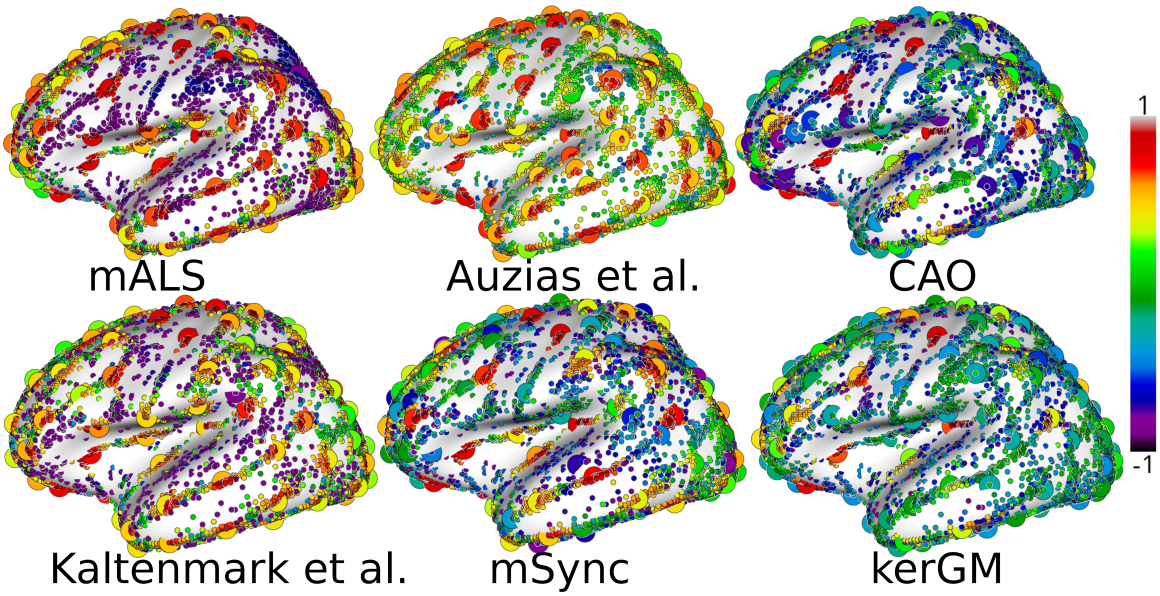

Figure 6: For each method, we show the silhouette coefficient of each node from every graphs, as well as corresponding centroids as larger circles. Each centroid (larger circles) is colored according to the average of the silhouette coefficient of corresponding nodes.

Figure 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 pariwise technique **KerGM** showing much lower values.

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 take into account 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 performances of the different methods and their robustness to controlled variations in the simulated graphs was informative for probing their effectiveness in this context. The performances of the pairwise approach **KerGM** were limited even when the level of perturbations was minimal. Note that we reported in Buskulic et al. (2021) that alternative pairwise techniques perform even worse on this task. Amongst the different multi-graph matching techniques that were tested, **mALS** showed better performances 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 performances 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 performances 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 performances 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 Considerations relative to deep-learning approaches and potential methodological improvements

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 (LeCun, Bengio, & Hinton, 2015). Recent works such as e.g. (Scarselli, Gori, Tsoi, Hagenbuchner, & Monfardini, 2009; Xu, Hu, Leskovec, & Jegelka, 2019) 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 re-

gions in the ambiguity of the matching. The work by Fey, Lenssen, Morris, Masci, and Kriege (2020) considers a semi-supervised framework for handling the matching problem where the ground-truth correspondence 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 Lyu et al. (2021) 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 Awate, Yushkevich, Song, Licht, and Gee (2010), or the local gyrification index Rabiei, Richard, Coulon, and Lefèvre (2017). On the other hand, the literature on learning edge representations is very scarce Hsu, Shen, and Cremers (2022), and the attributes on edges are most often reduced to a scalar value (i.e. a simple weight). In particular, the methods included in the present work cannot handle *vectors* of attributes on edges. Some recent deep learning methods such as (R. Wang et al., 2021) can exploit such vectors of attributes, but their scalability is limited by the size of the affinity matrices. We proposed in Dupé, Yadav, Auzias, and Takerkart (2022) to overcome this limitation by leveraging the recent matrix factorization method from Zhang et al. (2019). We will further investigate the potential of these methods in our future studies.

5.3 Data-driven nomenclature of sulcal basins

The present work extends previously reported experimental results illustrating the major impact of the labeling strategy on the induced correspondences and data-driven nomenclature. In Kaltenmark et al. (2020), the number of clusters obtained at the group level varied from 90 to 114 for the right hemisphere using either the approach proposed in Kaltenmark et al. (2020) or Auzias et al. (2015) respectively, on the same population of subjects. We provide a much more detailed comparison. We show on Fig. 8 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 matchings and thus different correspondences across individuals. Indeed, graph-matching techniques such as **mALS** are able to take into account the 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.

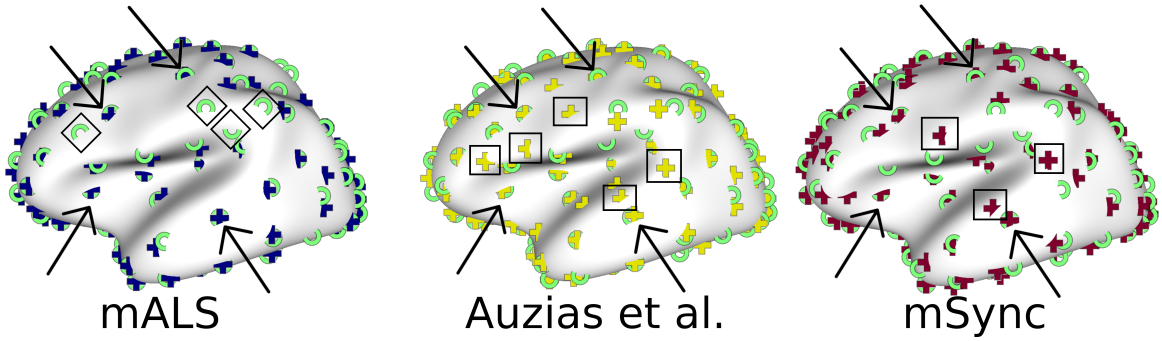

Figure 8: Superimposition of the centroids from **mALS**, **Auzias et al.**, and **mSync** shown as crosses with those of **Kaltenmark et al.** shown as green rings. **mSync** is representative of the methods **kerGM** and **CAO** that also result in 101 clusters. The arrows point to centroids that are 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.**.

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.8, 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.

6 Conclusion

In this study, we explored the potential of several graph matching methods chosen 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 graph. We then compared the approaches on our simulated sulcal graphs with ground-truth correspondences defined by construction.

We also evaluated the methods on real data. 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. We also computed a consistency measure that gave us an insight on the variability across the population for each cluster. The results obtained on real data were compared with two other methods from the literature.

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** 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** makes it also a relevant candidate for further studies and extensions to larger populations.

7 Funding sources

The data used in the preparation of this article were obtained from OASIS: Cross-Sectional: Principal Investigators: D. Marcus, R, Buckner, J, Csernansky J. Morris; P50 AG05681, P01 AG03991, P01 AG026276, R01 AG021910, P20 MH071616, U24 RR021382. The project leading to this publication has received funding from Excellence Initiative of Aix-Marseille University - A*MIDEX, a french "Investissements d'Avenir" programme (AMX-19-IET-002). The research leading to these results has also been supported by the ANR SulcalGRIDS Project, Grant ANR-19-CE45-0014 and the ERA-NET NEURON MULTI-FACT Project, Grant ANR-21-NEU2-0005 funded by the French National Research Agency.

8 Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

9 Author contributions

We report individual contributions to the paper using the relevant CRediT roles:

R.Yadav:Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Software; Visualization; Writing - original draft; Writing - review & editing.

F.-X.Dupé: Conceptualization; Formal analysis; Funding acquisition; Investigation; Methodology; Supervision; Writing - original draft; Writing - review & editing.

S.Takerkart: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Software; Supervision; Writing - original draft; Writing - review & editing.

G.Auzias: Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Resources; Software; Supervision; Validation; Visualization; Writing - original draft; Writing - review & editing.

References

Armstrong, E., Schleicher, A., Omran, H., Curtis, M., & Zilles, K. (1995). The ontogeny of human gyrification. *Cerebral cortex*, *5*(1), 56–63.

Auzias, G., Brun, L., Deruelle, C., & Coulon, O. (2015, May). Deep sulcal landmarks: Algorithmic and conceptual improvements in the definition and extraction of sulcal pits. *NeuroImage*, *111*, 12–25. Retrieved 2020-02-17, from `https://linkinghub.elsevier.com/retrieve/pii/S1053811915001007` doi: 10.1016/j.neuroimage.2015.02.008

Auzias, G., Viellard, M., Takerkart, S., Villeneuve, N., Poinso, F., Da Fonséca, D., ... Deruelle, C. (2014). Atypical sulcal anatomy in young children with autism spectrum disorder. *NeuroImage: Clinical*, *4*, 593–603.

Awate, S. P., Yushkevich, P. a., Song, Z., Licht, D. J., & Gee, J. C. (2010). Cerebral cortical folding analysis with multivariate modeling and testing: Studies on gender differences and neonatal development. , *53*(2), 450–459. (ISBN: 1095-9572 (Electronic)\r1053-8119 (Linking) Publisher: Elsevier Inc.) doi: 10.1016/j.neuroimage.2010.06.072

Behnke, K., Rettmann, M. E., Pham, D. L., Shen, D., Resnick, S. M., Davatzikos, C., & Prince, J. L. (2003). Automatic classification of sulcal regions of the human brain cortex using pattern recognition. In *Proceedings of SPIE* (Vol. 5032, pp. 1499–1510).

Bernard, F., Thunberg, J., Swoboda, P., & Theobalt, C. (2019). Hippi: Higher-order projected power iterations for scalable multi-matching. In *Proceedings of the ieee/cvf international conference on computer vision* (pp. 10284–10293).

Blaser, R., Fryzlewicz, P., Blaser, R., & Fryzlewicz, P. (2016). Random rotation ensembles.

Borne, L., Rivière, D., Mancip, M., & Mangin, J.-F. (2020). Automatic labeling of cortical sulci using patch-or cnn-based segmentation techniques combined with bottom-up geometric constraints. *Medical Image Analysis*, *62*, 101651.

Buskulic, N., Dupé, F.-X., Takerkart, S., & Auzias, G. (2021). Labelling sulcal graphs across indiviuals using multigraph matching. In *2021 ieee 18th international symposium on biomedical imaging (isbi)* (p. 1486-1490). doi: 10.1109/ISBI48211.2021.9434035

Cabral, R., De la Torre, F., Costeira, J. P., & Bernardino, A. (2013). Unifying nuclear norm and bilinear factorization approaches for low-rank matrix decomposition. In *Proceedings of the ieee international conference on computer vision* (pp. 2488–2495).

Cachia, A., Paillère-Martinot, M.-L., Galinowski, A., Januel, D., de Beaurepaire, R., Bellivier, F., ... others (2008). Cortical folding abnormalities in schizophrenia patients with resistant auditory hallucinations. *Neuroimage*, *39*(3), 927–935.

Chen, Y., Guibas, L. J., & Huang, Q.-X. (2014). Near-optimal joint object matching via convex relaxation. *arXiv preprint arXiv:1402.1473*.

Cour, T., Srinivasan, P., & Shi, J. (2007). Balanced Graph Matching. In *Advances in Neural Information Processing Systems 19* (pp. 313–320). MIT Press. Retrieved 2020-03-20, from `http://papers.nips.cc/paper/2960-balanced-graph-matching.pdf`

Devlin, J. T., & Poldrack, R. A. (2007, October). In praise of tedious anatomy. *NeuroImage*,

37(4), 1033–41; discussion 1050–8. doi: 10.1016/j.neuroimage.2006.09.055

Duchesnay, E., Cachia, A., Roche, A., Rivière, D., Cointepas, Y., Papadopoulos-Orfanos, D., ... Mangin, J.-F. (2007, May). Classification based on cortical folding patterns. *IEEE transactions on medical imaging*, *26*(4), 553–65. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17427742 doi: 10.1109/TMI.2007.892501

Dupé, F.-X., Yadav, R., Auzias, G., & Takerkart, S. (2022). Kernelized multi-graph matching. In *Acml 2022.*

Eckstein, J., & Bertsekas, D. P. (1992). On the douglas—rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, *55*(1), 293–318.

Eickhoff, S. B., Yeo, B. T. T., & Genon, S. (2018). Imaging-based parcellations of the human brain. , *19*(11), 672–686. (Number: 11 Publisher: Nature Publishing Group) doi: 10.1038/s41583-018-0071-7

Faktor, A., & Irani, M. (2013). "clustering by composition"—unsupervised discovery of image categories. *IEEE transactions on pattern analysis and machine intelligence*, *36*(6), 1092–1106.

Fey, M., Lenssen, J. E., Morris, C., Masci, J., & Kriege, N. M. (2020). Deep graph matching consensus. In *International conference on learning representations.*

Fischl, B., Sereno, M. I., Tootell, R. B., & Dale, A. M. (1999). High-resolution intersubject averaging and a coordinate system for the cortical surface. *Human brain mapping*, *8*(4), 272–284.

Glasser, M. F., Coalson, T. S., Robinson, E. S. J., Hacker, C. D., Harwell, J., Yacoub, E., ... Andersson, J. (2016). A multi-modal parcellation of human cerebral cortex. , *536*(7615), 171–178. Retrieved from http://dx.doi.org/10.1038/nature18933http://www.nature.com/articles/nature18933 (ISBN: 0008-5472 (Print)\r0008-5472 (Linking) Publisher: Nature Publishing Group _eprint: NIHMS150003) doi: 10.1038/nature18933

Hastie, T., Mazumder, R., Lee, J. D., & Zadeh, R. (2015). Matrix completion and low-rank svd via fast alternating least squares. *The Journal of Machine Learning Research*, *16*(1), 3367–3402.

Hsu, H. H.-H., Shen, Y., & Cremers, D. (2022). A graph is more than its nodes: Towards structured uncertainty-aware learning on graphs. *arXiv preprint arXiv:2210.15575*.

Hu, N., Huang, Q., Thibert, B., & Guibas, L. J. (2018). Distributable consistent multi-object matching. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 2463–2471).

Hutchison, D., Kanade, T., Kittler, J., Kleinberg, J., Mattern, F., Mitchell, J., ... Lee, K. (2010). Reweighted Random Walks for Graph Matching. In *Computer Vision – ECCV 2010* (Vol. 6315, pp. 492–505). Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved 2020-03-10, from http://link.springer.com/10.1007/978-3-642-15555-0_36 (Lecture Notes in Computer Science) doi: 10.1007/978-3-642-15555-0_36

Im, K., Guimaraes, A., Kim, Y., Cottrill, E., Gagoski, B., Rollins, C., ... Grant, P. E. (2017, July). Quantitative Folding Pattern Analysis of Early Primary Sulci in Human Fetuses with Brain Abnormalities. *American Journal of Neuroradiology*, *38*(7), 1449–1455. Retrieved from http://www.ajnr.org/lookup/doi/10.3174/ajnr.A5217 doi: 10.3174/ajnr.A5217

Im, K., Jo, H. J., Mangin, J.-F., Evans, A. C., Kim, S. I., & Lee, J.-M. (2010, March). Spatial distribution of deep sulcal landmarks and hemispherical asymmetry on the cortical surface. *Cerebral cortex (New York, N.Y. : 1991)*, *20*(3), 602–11. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/19561060 doi: 10.1093/cercor/bhp127

Im, K., Lee, J.-M., Yoon, U., Shin, Y.-W., Hong, S. B., Kim, I. Y., ... Kim, S. I. (2006). Fractal dimension in human cortical surface: multiple regression analysis with cortical thickness, sulcal depth, and folding area. *Human brain mapping*, *27*(12), 994–1003.

Im, K., Pienaar, R., Lee, J.-M., Seong, J.-K., Choi, Y. Y., Lee, K. H., & Grant, P. E. (2011). Quantitative comparison and analysis of sulcal patterns using sulcal graph matching: a twin study. *Neuroimage*, *57*(3), 1077–1086.

Kaltenmark, I., Deruelle, C., Brun, L., Lefèvre, J., Coulon, O., & Auzias, G. (2020). Cortical inter-subject correspondences with optimal group-wise parcellation and sulcal pits labeling. *Medical Image Analysis*.

Lawler, E. L. (1963). The quadratic assignment problem. *Management science*, *9*(4), 586–599.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436–444.

Lefèvre, J., Pepe, A., Muscato, J., De Guio, F., Girard, N., Auzias, G., & Germanaud, D. (2018). SPANOL (SPectral ANalysis of lobes): A spectral clustering framework for individual and group parcellation of cortical surfaces in lobes. , *12*, 1–14. doi: 10.3389/fnins.2018.00354

Le Guen, Y., Auzias, G., Leroy, F., Noulhiane, M., Dehaene-Lambertz, G., Duchesnay, E., ... Frouin, V. (2017). Genetic Influence on the Sulcal Pits: On the Origin of the First Cortical Folds. *Cerebral Cortex*(2015), 1–12. Retrieved from https://academic.oup.com/cercor/article-lookup/doi/10.1093/cercor/bhx098 (ISBN: 9781509011728) doi: 10.1093/cercor/bhx098

Leordeanu, M., & Hebert, M. (2005). A spectral technique for correspondence problems using pairwise constraints.

Leordeanu, M., Hebert, M., & Sukthankar, R. (2009). An Integer Projected Fixed Point Method for Graph Matching and MAP Inference. In *Advances in Neural Information Processing Systems 22* (pp. 1114–1122). Curran Associates, Inc. Retrieved 2020-03-20, from http://papers.nips.cc/paper/3756-an-integer-projected-fixed-point-method-for-graph-matching-and-map-inference.pdf

Leroy, F., Cai, Q., Bogart, S. L., Dubois, J., Coulon, O., Monzalvo, K., ... Dehaene-Lambertz, G. (2015). New human-specific brain landmark: The depth asymmetry of superior temporal sulcus. , *112*(4), 1208–1213. Retrieved 2021-08-31, from http://www.pnas.org/lookup/doi/10.1073/pnas.1412389112 doi: 10.1073/pnas.1412389112

Loiola, C. F., Silva, D. A. d., & Galati, E. A. B. (2007). 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*, *102*, 581–585.

Lyu, I., Bao, S., Hao, L., Yao, J., Miller, J. A., Voorhies, W., ... Landman, B. A. (2021). Labeling lateral prefrontal sulci using spherical data augmentation and context-aware training. *NeuroImage*, *229*, 117758. Retrieved from https://www.sciencedirect.com/science/article/pii/S1053811921000355 doi: https://doi.org/10.1016/j.neuroimage.2021.117758

Mangin, J.-F., Riviere, D., Cachia, A., Duchesnay, E., Cointepas, Y., Papadopoulos-Orfanos, D., ... Régis, J. (2004). Object-based morphometry of the cerebral cortex. *IEEE transactions on medical imaging*, *23*(8), 968–982.

Marcus, D. S., Wang, T. H., Parker, J., Csernansky, J. G., Morris, J. C., & Buckner, R. L. (2007). Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented older adults. *Journal of cognitive neuroscience*, *19*(9), 1498–1507.

Meng, Y., Li, G., Wang, L., Lin, W., Gilmore, J. H., & Shen, D. (2018). Discovering cortical sulcal folding patterns in neonates using large-scale dataset. *Human brain mapping*, *39*(9), 3625–3635.

Morton, S. U., Maleyeff, L., Wypij, D., Yun, H. J., Newburger, J. W., Bellinger, D. C., ... Im, K. (2019, June). Abnormal Left-Hemispheric Sulcal Patterns Correlate with Neurodevelopmental Outcomes in Subjects with Single Ventricular Congenital Heart Disease. *Cerebral Cortex*, 1–12. Retrieved from https://academic.oup.com/cercor/advance-article/doi/10.1093/cercor/bhz101/5520589 doi: 10.1093/cercor/bhz101

Pachauri, D., Kondor, R., & Singh, V. (2013). Solving the multi-way matching problem by permutation synchronization. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 26* (pp. 1860–1868). Curran Associates, Inc. Retrieved 2020-02-20, from http://papers.nips.cc/paper/4987-solving-the-multi-way-matching-problem-by-permutation-synchronization.pdf

Pizzagalli, F., Auzias, G., Yang, Q., Mathias, S. R., Faskowitz, J., Boyd, J. D., ... Jahanshad, N. (2020, September). The reliability and heritability of cortical folds and their genetic correlations across hemispheres. *Communications Biology*, *3*(1), 1–12. Retrieved 2021-02-08, from https://www.nature.com/articles/s42003-020-01163-1 (Number: 1 Publisher: Nature Publishing Group) doi: 10.1038/s42003-020-01163-1

Rabiei, H., Richard, F., Coulon, O., & Lefèvre, J. (2017). Local spectral analysis of the cerebral cortex: New gyrification indices. , *36*(3), 838–848. doi: 10.1109/TMI.2016.2633393

Rivière, D., Mangin, J.-F., Papadopoulos-Orfanos, D., Martinez, J.-M., Frouin, V., &

Régis, J. (2002, July). Automatic recognition of cortical sulci of the human brain using a congregation of neural networks. *Medical image analysis*, *6*(2), 77–92. Retrieved from http://www.sciencedirect.com/science/article/pii/S136184150200052Xhttp://www.ncbi.nlm.nih.gov/pubmed/12044997

Rolínek, M., Swoboda, P., Zietlow, D., Paulus, A., Musil, V., & Martius, G. (2020). Deep graph matching via blackbox differentiation of combinatorial solvers. In *European conference on computer vision* (pp. 407–424).

Rousseeuw, P. J. (1987, November). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65. Retrieved 2021-10-04, from https://www.sciencedirect.com/science/article/pii/0377042787901257 doi: 10.1016/0377-0427(87)90125-7

Rubinstein, M., Joulin, A., Kopf, J., & Liu, C. (2013). Unsupervised joint object discovery and segmentation in internet images. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 1939–1946).

Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., & Monfardini, G. (2009). The graph neural network model. *IEEE Transactions on Neural Networks*, *20*(1), 61-80. doi: 10.1109/TNN.2008.2005605

Sprung-Much, T., & Petrides, M. (2020). Morphology and spatial probability maps of the horizontal ascending ramus of the lateral fissure. , *30*(3), 1586–1602. Retrieved 2021-03-23, from https://doi.org/10.1093/cercor/bhz189 doi: 10.1093/cercor/bhz189

Takerkart, S., Auzias, G., Brun, L., & Coulon, O. (2017, January). Structural graph-based morphometry: A multiscale searchlight framework based on sulcal pits. *Medical Image Analysis*, *35*, 32–45. Retrieved 2020-02-17, from https://linkinghub.elsevier.com/retrieve/pii/S1361841516300251 doi: 10.1016/j.media.2016.04.011

Tron, R., Zhou, X., Esteves, C., & Daniilidis, K. (2017, October). Fast Multi-image Matching via Density-Based Clustering. In *2017 IEEE International Conference on Computer Vision (ICCV)* (pp. 4077–4086). Venice: IEEE. Retrieved 2020-02-21, from http://ieeexplore.ieee.org/document/8237699/ doi: 10.1109/ICCV.2017.437

Van Essen, D. C., & Dierker, D. L. (2007, October). On navigating the human cerebral cortex: response to 'in praise of tedious anatomy'. *NeuroImage*, *37*(4), 1050–4; discussion 1066–8. doi: 10.1016/j.neuroimage.2007.02.021

Van Essen, D. C., Glasser, M. F., Dierker, D. L., Harwell, J. W., & Coalson, T. S. (2012). Parcellations and hemispheric asymmetries of human cerebral cortex analyzed on surface-based atlases. , *22*(10), 2241–62. doi: 10.1093/cercor/bhr291

Von Mises, R. (1964). *Mathematical theory of probability and statistics*. Academic press.

Voorhies, W. I., Miller, J. A., Yao, J. K., Bunge, S. A., & Weiner, K. S. (2021). Cognitive insights from tertiary sulci in prefrontal cortex. , *12*(1), 5122. Retrieved 2021-09-01, from https://www.nature.com/articles/s41467-021-25162-w doi: 10.1038/s41467-021-25162-w

Wang, B., Mezlini, A. M., Demir, F., Fiume, M., Tu, Z., Brudno, M., ... Goldenberg, A. (2014). Similarity network fusion for aggregating data types on a genomic scale. *Nature methods*, *11*(3), 333–337.

Wang, Q., Zhou, X., & Daniilidis, K. (2018). Multi-image semantic matching by mining consistent features. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 685–694).

Wang, R., Yan, J., & Yang, X. (2019). Learning combinatorial embedding networks for deep graph matching. In *Proceedings of the ieee/cvf international conference on computer vision* (pp. 3056–3065).

Wang, R., Yan, J., & Yang, X. (2020a). Combinatorial learning of robust deep graph matching: an embedding based approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Wang, R., Yan, J., & Yang, X. (2020b). Graduated assignment for joint multi-graph matching and clustering with application to unsupervised graph matching network learning. *Advances in Neural Information Processing Systems*, *33*, 19908–19919.

Wang, R., Yan, J., & Yang, X. (2021). 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*.

Willbrand, E. H., Parker, B. J., Voorhies, W. I., Miller, J. A., Lyu, I., Hallock, T., ... Weiner, K. S. (2022). Uncovering a tripartite landmark in posterior cingulate cortex. , *8*(36), eabn9516. Retrieved 2023-01-16, from `https://www.science.org/doi/full/10.1126/sciadv.abn9516` (Publisher: American Association for the Advancement of Science) doi: 10.1126/sciadv.abn9516

Xu, K., Hu, W., Leskovec, J., & Jegelka, S. (2019). How powerful are graph neural networks? In *International conference on learning representations.*

Yadav, R., Dupé, F. X., Takerkart, S., & Auzias, G. (2022). On the relevance of multi-graph matching for sulcal graphs. In *2022 ieee international conference on image processing (icip)* (p. 2536-2540). doi: 10.1109/ICIP46576.2022.9897185

Yan, J., Cho, M., Zha, H., Yang, X., & Chu, S. M. (2016). Multi-graph matching via affinity optimization with graduated consistency regularization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *38*(6), 1228-1242. doi: 10.1109/TPAMI.2015.2477832

Yan, J., Li, Y., Liu, W., Zha, H., Yang, X., & Chu, S. M. (2014). Graduated consistency-regularized optimization for multi-graph matching. In *European conference on computer vision* (pp. 407–422).

Yan, J., Tian, Y., Zha, H., Yang, X., Zhang, Y., & Chu, S. M. (2013). Joint optimization for consistent multiple graph matching. In *Proceedings of the ieee international conference on computer vision* (pp. 1649–1656).

Yan, J., Wang, J., Zha, H., Yang, X., & Chu, S. (2015, March). Consistency-Driven Alternating Optimization for Multigraph Matching: A Unified Approach. *IEEE Transactions on Image Processing*, *24*(3), 994–1009. Retrieved 2020-02-20, from `http://`

ieeexplore.ieee.org/document/7001592/ doi: 10.1109/TIP.2014.2387386

Yan, J., Yin, X.-C., Lin, W., Deng, C., Zha, H., & Yang, X. (2016). A short survey of recent advances in graph matching. In *Proceedings of the 2016 acm on international conference on multimedia retrieval* (pp. 167–174).

Yu, T., Wang, R., Yan, J., & Li, B. (2019). Learning deep graph matching with channel-independent embedding and hungarian attention. In *International conference on learning representations.*

Yu, T., Wang, R., Yan, J., & Li, B. (2021, 18–24 Jul). Deep latent graph matching. In M. Meila & T. Zhang (Eds.), *Proceedings of the 38th international conference on machine learning* (Vol. 139, pp. 12187–12197). PMLR. Retrieved from https://proceedings.mlr.press/v139/yu21d.html

Zanfir, A., & Sminchisescu, C. (2018). Deep learning of graph matching. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 2684–2693).

Zhang, Z., Xiang, Y., Wu, L., Xue, B., & Nehorai, A. (2019). KerGM: Kernelized Graph Matching. In *Advances in Neural Information Processing Systems 32* (pp. 3335–3346). Retrieved 2020-03-04, from http://papers.nips.cc/paper/8595-kergm-kernelized-graph-matching.pdf

Zhou, F., & De la Torre, F. (2015). Factorized graph matching. *IEEE transactions on pattern analysis and machine intelligence*, *38*(9), 1774–1789.

Zhou, T., Jae Lee, Y., Yu, S. X., & Efros, A. A. (2015). Flowweb: Joint image set alignment by weaving consistent, pixel-wise correspondences. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 1191–1200).

Zhou, X., Zhu, M., & Daniilidis, K. (2015). Multi-image matching via fast alternating minimization. In *Proceedings of the ieee international conference on computer vision* (pp. 4032–4040).

Appendices

A Additional description of the β-binomial distribution

In the following section we provide additional description of the formulation of the β-binomial distribution.

The β-binomial distribution is parameterized by ν, α and β. The parameter ν defines the size of support (in our case, maximum number of n_o/n_s). The setting of ν can impact the skewness of the distribution but the shape will be Gaussian as long as the value is sufficiently large. We show in figure A.1 the β-binomial distributions for different values of ν, with $\alpha = 7.15$ and $\beta = 28.62$. In this work, we set $\nu = 30$ which is sufficient to get a distribution close to Gaussian for all combinations of α and β parameters that are relevant in our context.

Although α and β are not trivial to calibrate, they can be related to μ and σ of a Gaussian

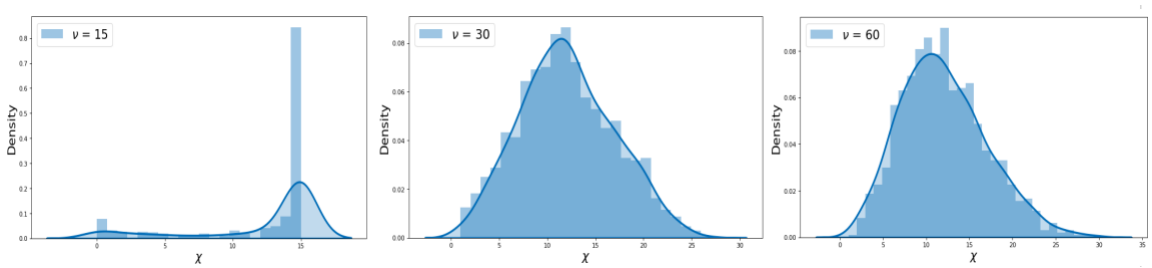

Figure A.1: Effect on β-binomial mass function for different values of ν fixing $\alpha = 7.15$ and $\beta = 28.62$

distribution using the following formulae:

$$\rho = \frac{\nu - \mu}{\mu}, \qquad \alpha = \frac{(1+\rho)^2\sigma - n^2\rho}{(\nu\rho(1+\rho) - \sigma(1+\rho)^3}, \qquad \beta = \frac{\nu - \mu}{\mu}\alpha \tag{A.1}$$

which allows us to control for μ and σ, i.e. the amount of nodes to suppress(n_s) and outliers to add(n_o). Figure A.1 illustrates how we can obtain distributions close to two Gaussian with identical μ but different σ value, by controlling α and β.

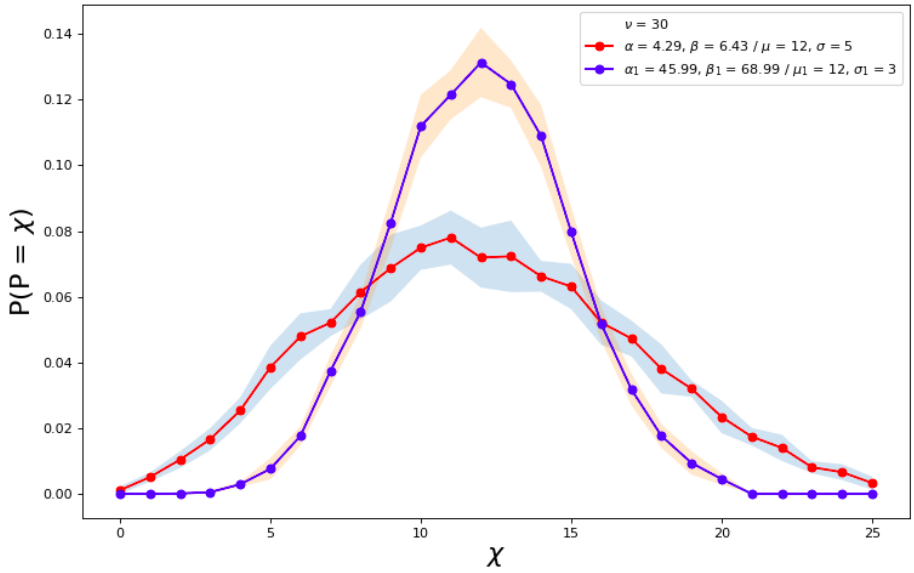

Figure A.1: β-binomial distributions for identical mean: $\mu, \mu_1 = 12$ but different standard deviations: $\sigma = 3, \sigma_1 = 5$. The dotted lines signifies the mean of the distribution where as the shaded area is the standard deviation across 5 trials.

B Supplementary data showing the fit between simulated and real graphs

As stated in section in 4.3.1 we empirically set the simulation parameters $\mu_{pert}, \sigma_{pert}$ and $p = 10\%$ such that our simulated graphs follow the intrinsic properties of real graphs. With the choice of $\mu_{pert}, \sigma_{pert}$ we estimate the corresponding α and β of β-binomial distribution

as described in Appendix A. This allows us to generate graphs with similar mean and standard deviation of number of nodes as in the real data. This control on the number of nodes is independent from the other types of perturbations we induce. In particular, we show on Figure B.1 the match between simulated and real graphs for various values of the parameter controlling for the perturbation of the coordinates of the nodes, κ. This figure shows the density distribution for the number of nodes in the simulated graphs for different values of κ, compared to number of nodes in the real population. The largely overlapping distributions confirm the match of the number of nodes, for any value of κ.

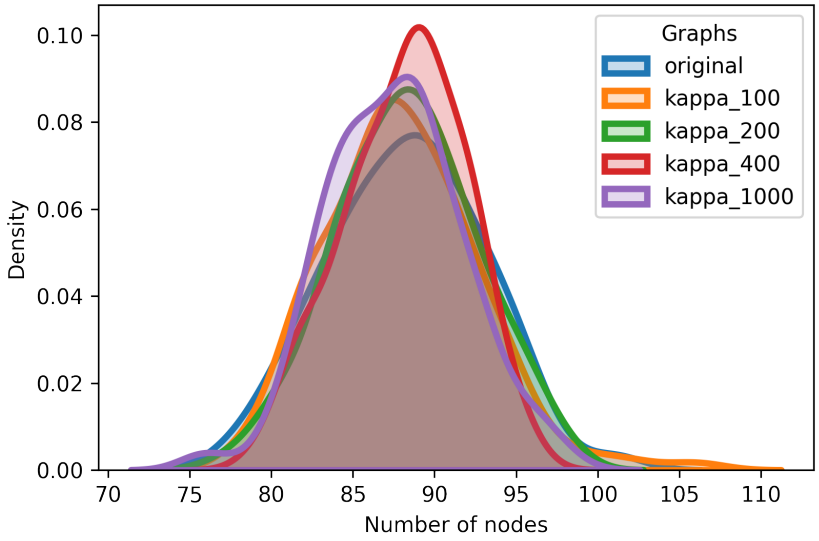

Figure B.1: Distribution for number of node in the simulated population corresponding to different κ value with the distribution for number of nodes in the real sulcal graphs.

In addition, we also compare the distributions of the geodesic length of the edges which serves as the feature on the edges. Figure B.2 shows the distributions of mean geodesic distance across a populations of real data and simulated graphs for different values of κ. The shaded area surrounding each curve shows the standard deviation across a population of 137 graphs in both real and simulated population. As stated in section 3.1 the distance between the nodes in the real graphs are larger than a minimum distance, which is illustrated by the flat portion of the blue curve for low geodesic distances. Our simulations do not reproduce this feature, as expected from the uniform sampling of the location of outliers nodes that can get close to previous nodes (figure 3.b,c). Note that the fit is good for larger geodesic distance values.

Finally, we show on Figure B.3 the distribution of the degree of nodes for simulated and real graphs. The degree corresponds to the number of neighbors of each node, and is thus indicative of the local topology of the graphs. This figure confirms the good match between simulated and real data, independently of κ that controls the perturbation level.

Overall, all our measures confirmed a good match between simulated and real graphs, for any value of κ.

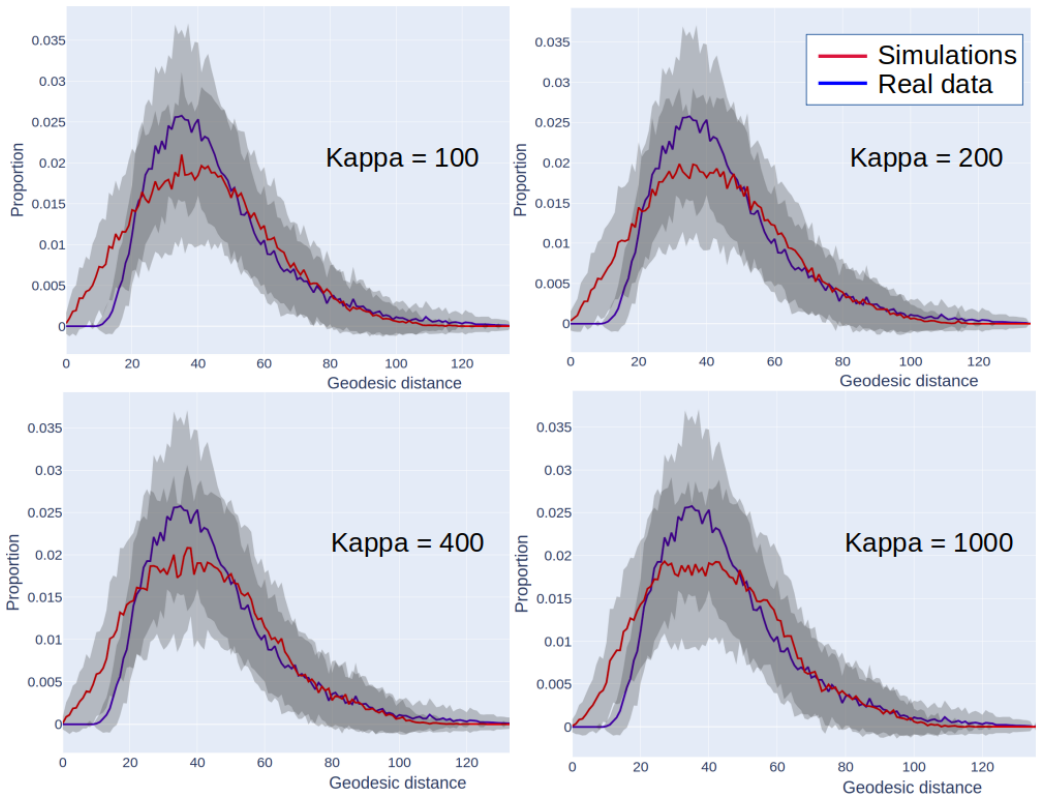

Figure B.2: Distribution for geodesic distance in the simulated and real population of 137 graphs. The shaded region corresponds to standard deviations across graphs in the population.

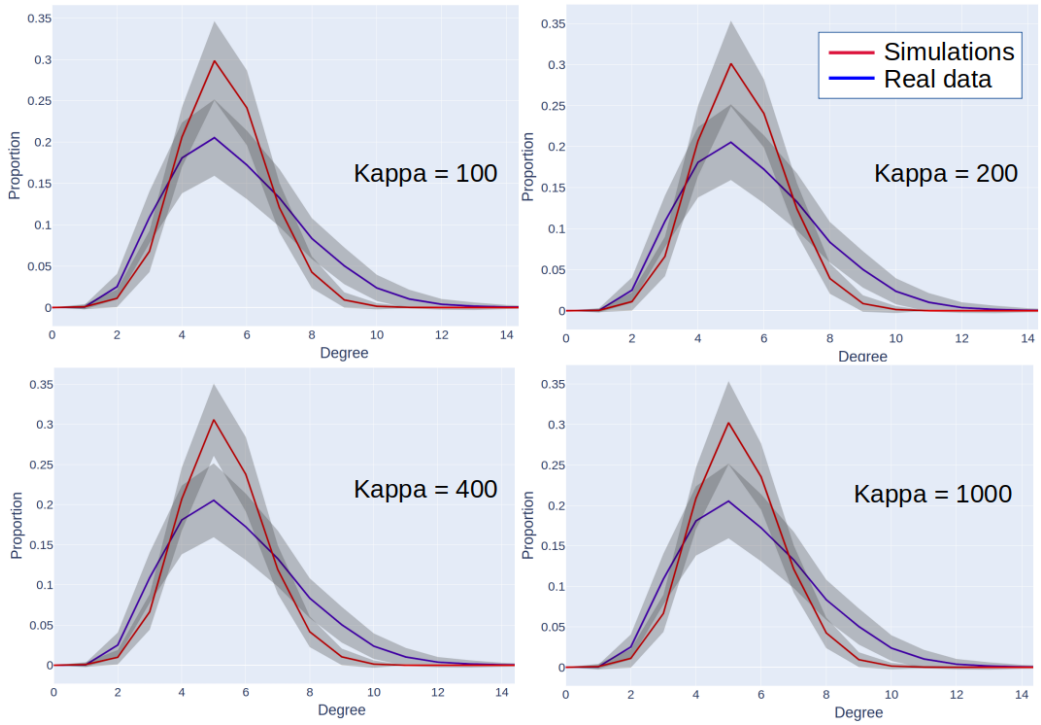

Figure B.3: Degree distribution in the simulated and real population of 137 graphs. The shaded region corresponds to standard deviations across graphs in the population.