

Topological Learning for Brain Networks

Tananun Songdechakraiwt, Moo K. Chung

Department of Biostatistics and Medical Informatics

University of Wisconsin–Madison

Abstract

This paper proposes a novel topological learning framework that can integrate brain networks of different sizes and topology through persistent homology. This is possible through the introduction of a new topological loss function that enables such challenging task. The use of the proposed loss function bypasses the intrinsic computational bottleneck associated with matching networks. We validate the method in extensive statistical simulations with ground truth to assess the effectiveness of the topological loss in discriminating networks with different topology. The method is further applied to a twin brain imaging study in determining if the brain network is genetically heritable. The challenge is in overlaying the topologically different functional brain networks obtained from the resting-state functional MRI (fMRI) onto the template structural brain network obtained through the diffusion MRI (dMRI).

Keywords– Topological data analysis, persistent homology, topological learning, brain network, twin imaging study

1 Introduction

The human brain is an extraordinarily complicated system of interconnected neurons. The brain is roughly estimated to have 10^{12} neurons with 10^{15} synapses (Johnston and Wu, 1994). In an ideal situation, a brain network should be modeled using neurons as network nodes with every synapse as an edge connecting the nodes. Due to the technical limitation of magnetic resonance imaging (MRI) scanners, it is not possible to do so yet. In the popular brain network modeling framework, the whole brain is parcellated into d disjoint regions, where d is usually a few hundreds (Tzourio-Mazoyer et al., 2002; Desikan et al., 2006; Hagmann et al., 2007; Shattuck et al., 2008; Gong et al., 2009; Fornito et al., 2010; Zalesky et al., 2010; Fan et al., 2016; Glasser et al., 2016; Fornito et al., 2016; Schaefer et al., 2017; Eickhoff et al., 2018; Arslan et al., 2018). Subsequently, functional or structural information is overlaid on top of the parcellation to obtain $d \times d$ connectivity matrices that measure the strength of connectivity between brain regions for standard network analysis. For instance, the Automated Anatomical Labeling (AAL) partitions the brain into 116 regions (Figure 1) (Tzourio-Mazoyer et al., 2002). These disjoint brain regions form nodes in the brain network. Although there has been a long-standing effort toward standardizing the parcellation, there is no single universally accepted approach.

Connectivity between brain regions that defines edges in the brain network is usually determined by the type of imaging modality (Ombao et al., 2016). Structural connectivity is obtained through diffusion MRI (dMRI), which can trace the white matter fibers connecting brain regions. The strength of structural connectivity between the brain regions is determined by the number of fibers passing through them (Chung et al., 2011, 2019c). The structural brain network is expected to exhibit sparse topology without many loops or cycles (Figure 1) (Gong et al., 2009; Chung et al., 2011; Zhang et al., 2018). On the other hand, functional connectivity obtained from the resting-state functional MRI (fMRI) is often computed as the Pearson correlation coefficient between brain regions (Chung, 2013; Bryant et al., 2017; Shappell et al., 2019; Chung et al., 2019b). While structural connectivity provides information whether the brain regions are physically connected through white matter fibers, functional connectivity exhibits connections between two regions without the direct neuroanatomical connections through additional intermediate connections (Honey et al., 2007, 2009). Thus, resting-state functional brain networks are often very dense with thousands of cycles. Both structural and functional brain networks provide topologically different information. Existing graph theory based brain network analyses have shown that there is some common topological profile that is conserved for both structural and functional brain networks (Bullmore and Sporns, 2009). However, due to the difficulty of integrating both networks in a coherent statistical

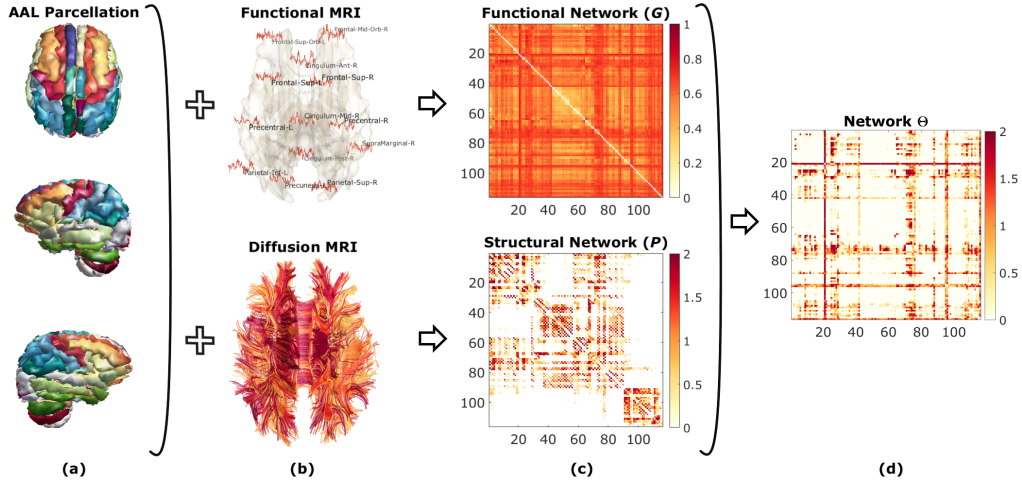


Figure 1: Schematic of topological learning on brain networks. (a) The Automated Anatomical Labeling (AAL) atlas obtained through structural-MRI is used to partition the human brain into 116 disjoint regions, which form the nodes of brain networks. In functional MRI, brain activity at each node is measured as a time series of changes associated with the relative blood oxygenation level (b-top). The functional connectivity between two nodes is given as the correlation between their fMRI time series, resulting in the functional network G through metric transform (c-top). The structural connectivity between two brain regions is measured by the number of white matter fiber tracts passing through them using dMRI (b-bottom). Structural connectivities over all subjects are then normalized and scaled, resulting in the structural network P (c-bottom) that serves as the template where statistical analysis can be performed. The structural network P is sparse while the functional network G is densely connected and has more cycles. Since both networks are topologically different, it is difficult to integrate them together in a coherent model. Simply overlaying functional brain networks on top of the structural network, as usually done in the field, will completely destroy 1D topology (cycles) of the functional networks (Zhu et al., 2014). (d) Using the proposed framework, we learn network Θ that has the topological characteristics of both functional and structural networks.

framework, not much research has been done on integrating such networks at the localized edge level. Many multimodal network researchers focus on comparing summary graph theory features across different networks (Bullmore and

Sporns, 2009; Ginestet et al., 2011; Karas et al., 2019). Although there are not many, some statistical studies have focused on fusing networks derived from both modalities *probabilistically*, which can easily destroy the aforementioned topological difference of the networks (Xue et al., 2015; Kang et al., 2017). Thus, there is a need for a new multimodal network model that can easily integrate networks of different topology at the localized connection level.

Topological Data Analysis (TDA) (Edelsbrunner et al., 2000; Wasserman, 2018), a general framework based on algebraic topology, can provide such novel solution to the long-standing multimodal brain network analysis challenge. Numerous TDA studies have been applied to increasingly diverse problems such as genetics (Chung et al., 2017b, 2019b), epileptic seizure detection (Wang et al., 2018), sexual dimorphism in the human brain (Songdechakraiut and Chung, 2020), analysis of brain arteries (Bendich et al., 2016), image segmentation (Clough et al., 2019), classification (Singh et al., 2014; Reininghaus et al., 2015; Chen et al., 2019), clinical predictive model (Crawford et al., 2020) and persistence-based clustering (Chazal et al., 2013). Persistent homology begins to emerge as a powerful mathematical representation to understand, characterize and quantify topology of brain networks (Lee et al., 2012; Chung et al., 2019b). In persistent homology, topological bases such as connected components and cycles are measured across different spatial resolutions. As the resolution changes, such features are born and die. Persistent homology associates the life-time to these features in the form of 1D interval from birth to death. Long-lived features persist over a long range of resolutions and are considered as signal while short-lived features are considered as noise (Carlsson, 2009). The collection of such intervals is summarized as a *barcode*, which completely characterizes the topology of underlying data (Ghrist, 2008).

Recently, it was proposed to penalize barcodes through *topological loss* for image segmentation (Hu et al., 2019). While the approach allows to incorporate topological information into segmentation problem, the method has been limited to 2D image segmentation with a small number of topological features due to its expensive optimization process involving $O(d^6)$ run-time (Edmonds and Karp, 1972; Kerber et al., 2017). Barcodes are typically computed at a finite set of pre-specified resolutions. A sufficient number of such resolutions is required to give a reasonably accurate estimation of barcodes, which quickly increases computational complexity when the size of data increases (Chung et al., 2019a; Hu et al., 2019). This is impractical in brain networks with far larger number of topological features involving hundreds of connected components and thousands of cycles. In this paper, motivated by (Cohen-Steiner et al., 2010; Hu et al., 2019), we propose a more principled approach that learns the topological structure of brain networks with large number of features in $O(d^2 \log d)$ run-time. Our proposed method bypasses the intrinsic computational bottleneck and thus

enables us to perform various topology computations and optimizations at every possible resolution.

We illustrate the method in studying the resting-state functional MRI networks of 194 twin pairs obtained from the Human Connectome Project (Van Essen et al., 2012, 2013). HCP twin brain imaging data is considered as the *gold standard*, where the zigosity is confirmed by the blood and saliva test (Gritsenko et al., 2020). Monozygotic (MZ) twins share 100% of genes while dizygotic (DZ) twins share 50% of genes (Falconer and Mackay, 1995; Chung et al., 2017b). MZ-twins are more similar or concordant than DZ-twins for cognitive aging, cognitive dysfunction and Alzheimer’s disease (Reynolds and Phillips, 2015). These genetic differences allow us to pull apart and examine genetic and environmental influences easily *in vivo*. The difference between MZ- and DZ-twins directly quantify the extent to which phenotypes are influenced by genetic factors. If MZ-twins show more similarity on a given trait compared to DZ-twins, this provides evidence that genes significantly influence that trait.

Previous twin brain imaging studies mainly used univariate imaging phenotypes such as cortical surface thickness (McKay et al., 2014), fractional anisotropy (Chiang et al., 2011), functional activation (Smit et al., 2008; Glahn et al., 2010; Blokland et al., 2011) in determining heritability in few regions of interest. Compared to existing studies on univariate imaging phenotypes, there are not many studies on the heritability of the whole brain functional networks (Blokland et al., 2011). Measures of network topology and features may be worth investigating as intermediate phenotypes that indicate the genetic risk for a neuropsychiatric disorder (Bullmore and Sporns, 2009). However, the brain network analysis has not yet been adapted for this purpose beyond a small number of regions. Determining the extent of heritability of the whole brain networks is the first necessary prerequisite for identifying network-based endophenotypes. Using the proposed method, we propose to determine the heritability of functional brain networks while integrating the structural brain network information. We demonstrate that our method increases the sensitivity in detecting subtle genetic signals.

2 Methods

2.1 Graph filtration

We consider a network, which is represented as a weighted graph $G = (V, w)$ comprising a set of nodes V and unique positive symmetric edge weights $w = (w_{ij})$. We assume the graph has no self-loops or parallel edges. The cardinality of sets is denoted using $|\cdot|$. The number of nodes and edges are then denoted as $|V|$ and $|E|$. Since G is a weighted complete graph, we have $|E| = |V|(|V| - 1)/2$. The binary graph $G_\epsilon = (V, w_\epsilon)$ of G is a graph consisting of node set V and

binary edge weight w_ϵ given by

$$w_\epsilon = (w_{\epsilon,ij}) = \begin{cases} 1 & \text{if } w_{ij} > \epsilon; \\ 0 & \text{otherwise.} \end{cases}$$

G_0 is a complete graph while G_∞ is the node set V . The binary network G_ϵ is the 1-skeleton, which is a simplicial complex consisting of nodes and edges only. In 1-skeleton, 0-dimensional (0D) holes are connected components and 1-dimensional (1D) holes are cycles (Chung et al., 2019b). Unlike Rips complex, which is the usual basic building block of persistent homology, there are no more higher dimensional topological features to compute beyond 1D in 1-skeleton. The number of connected components and the number of *independent* cycles in the network are referred to as the 0-th Betti number and the 1-st Betti number, respectively. The i -th Betti number of network G_ϵ is denoted as $\beta_i(G_\epsilon)$.

A graph filtration of G is defined as a collection of nested binary networks (Lee et al., 2012; Chung et al., 2017a,b, 2019b):

$$G_{\epsilon_0} \supset G_{\epsilon_1} \supset \cdots \supset G_{\epsilon_k},$$

where $\epsilon_0 < \epsilon_1 < \cdots < \epsilon_k$ are filtration values. Since there are many possible graph filtrations over the different choice of filtration values, we will simply take edge weights as filtration values and make the graph filtration unique.

By increasing ϵ , we are thresholding at higher connectivity, resulting in more edges being removed. Each removed edge either increments the number of connected components (β_0) or decrements the number of cycles (β_1) *at most by one* (Chung et al., 2019b). Figure 2-a displays an example of the graph filtration on a four-node network. Note G_0 is a complete graph consisting of a single connected component, while G_∞ is a node set. β_0 is monotonically increasing from $\beta_0(G_0) = 1$ to $\beta_0(G_\infty) = |V|$ and the increment is at most by one. Similarly, β_1 is monotonically decreasing from $\beta_1(G_0)$ to $\beta_1(G_\infty)$ and the decrement is at most by one. $\beta_1(G_\infty)$ is trivially zero. How many cycles are there in the complete graph G_0 ? The Euler characteristic χ of network G_0 is given by (Chung et al., 2019b)

$$\chi(G_0) = |V| - |E| = |V| - \frac{|V|(|V| - 1)}{2},$$

which can also be written as $\chi(G_0) = \beta_0(G_0) - \beta_1(G_0)$ (Adler et al., 2010). Thus, there are

$$\beta_1(G_0) = \beta_0(G_0) - \chi(G_0) = 1 + \frac{|V|(|V| - 3)}{2}$$

number of cycles. In Figure 2-a, G_0 is a tetrahedron with 4 faces, which are cycles. However, there exists only 3 independent cycles and $\beta_1(G_0) = 3$.

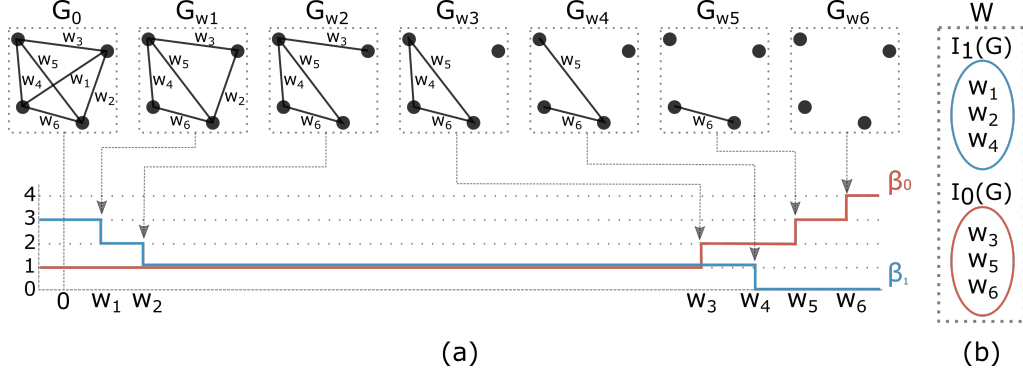


Figure 2: (a) Graph filtration on four-node network G . β_0 is monotonically increasing while β_1 is monotonically decreasing over the filtration. Connected components are born at the edge weights w_3, w_5, w_6 while cycles die at the edge weights w_1, w_2, w_4 . A cycle consisting of w_4, w_5, w_6 persists longer than any other 1D features and considered as topological signal. 0D barcode $P_0 = \{(-\infty, \infty), (w_3, \infty), (w_5, \infty), (w_6, \infty)\}$ is represented using the birth values as $I_0(G) = \{w_3, w_5, w_6\}$. 1D barcode $P_1 = \{(-\infty, w_1), (-\infty, w_2), (-\infty, w_4)\}$ is represented using the death values as $I_1(G) = \{w_1, w_2, w_4\}$. (b) We can prove that 0D and 1D barcodes uniquely partition the edge weight set W , i.e., $W = I_0(G) \cup I_1(G)$ with $I_0(G) \cap I_1(G) = \emptyset$.

Barcodes in graph filtration

Persistent homology keeps track of appearances (birth) and disappearances (death) of connected components and cycles over the whole range of filtration values ϵ , and associates their *persistence* (the life-time measured as the duration of birth to death) to them. Longer persistence indicates the presence of topological signal while shorter persistence is interpreted as topological noise (Edelsbrunner and Harer, 2008). One representation to capture the persistence of such topological features is in the form of barcodes (Ghrist, 2008; Adler et al., 2010; Lee et al., 2011). The 0D (or 1D) barcode corresponding to the network G is a collection P_0 (or P_1) of intervals $[\epsilon_b, \epsilon_d]$ such that each interval tabulates the life-time of a connected component (or a cycle) that appears at the first filtration value ϵ_b and vanishes at the last filtration value ϵ_d (Figure 2-a).

Since G_0 is a complete graph, a single connected component, we simply treat G_0 to be born at $-\infty$. By the monotonicity of β_0 (Chung et al., 2019b), the number of connected components is non-decreasing as ϵ increases. In other words, once a component is born, it never dies so every connected component has death value $\epsilon_d = \infty$, which we simply disregard. Ignoring one birth value corresponding

to the complete graph G_0 at $-\infty$ (Figure 2-a), the total number of finite birth values of connected components is then

$$m_0 = \beta_0(G_\infty) - 1 = |V| - 1.$$

Thus, we simply represent 0D barcode of the network only using the set of increasing birth values

$$I_0(G) : \epsilon_{b1} < \epsilon_{b2} < \cdots < \epsilon_{bm_0}.$$

Since G_0 is the complete graph, all cycles are considered to be born at $\epsilon_b = -\infty$. By the monotonicity of β_1 , the number of cycles in 1D barcode is non-increasing as ϵ increases. Thus, a cycle will die one at a time over increasing filtration values. The total number of death values of cycles is then equal to the number of cycles:

$$m_1 = \beta_1(G_0) = 1 + \frac{|V|(|V| - 3)}{2}.$$

Similarly, we can also represent 1D barcode only using the set of increasing death values

$$I_1(G) : \epsilon_{d1} < \epsilon_{d2} < \cdots < \epsilon_{dm_1}.$$

Figure 2-a illustrates how 0D and 1D barcodes corresponding to a four-node network are simplified to a collection of birth and death values, respectively.

Under the graph filtration, deleting edge w_{ij} in the network G must result in either the birth of a connected component or the death of a cycle. The birth of a component and the death of a cycle cannot possibly happen at the same time. Every edge weight must be in either 0D barcode or 1D barcode (Figure 2-b). Thus, we have

Theorem 1. *The set of 0D birth values $I_0(G)$ and 1D death values $I_1(G)$ partition the edge weight set W such that $W = I_0(G) \cup I_1(G)$ with $I_0(G) \cap I_1(G) = \emptyset$. The cardinalities of $I_0(G)$ and $I_1(G)$ are $|V| - 1$ and $1 + \frac{|V|(|V|-3)}{2}$, respectively.*

Theorem 1 is a non-trivial statement and used in development of the proposed topological learning framework.

Topological loss

Since a network is topologically completely characterized by 0D and 1D barcodes, the topological similarity between two networks can be measured using differences of such barcodes. We modified the Wasserstein distance to measure the differences in the barcodes for 1-skeleton (Cohen-Steiner et al., 2010; Rabin et al., 2011; Hu et al., 2019; Clough et al., 2019; Kolouri et al., 2019). Let $\Theta = (V^\Theta, w^\Theta)$ and $P = (V^P, w^P)$ be two given networks. For now, we will simply assume that the two networks have the same size, i.e., $|V^\Theta| = |V^P|$ but the

method works for $|V^\Theta| \neq |V^P|$ as well through the data augmentation (explained later). The topological loss $\mathcal{L}_{top}(\Theta, P)$ is defined as the optimal matching cost

$$\mathcal{L}_{top}(\Theta, P) = \min_{\tau} \left(\sum_{\epsilon_b \in I_0(\Theta)} [\epsilon_b - \tau(\epsilon_b)]^2 + \sum_{\epsilon_d \in I_1(\Theta)} [\epsilon_d - \tau(\epsilon_d)]^2 \right),$$

where τ is a *bijection* from $I_0(\Theta) \cup I_1(\Theta)$ to $I_0(P) \cup I_1(P)$. It is reasonable to match 0D to 0D persistences and 1D to 1D persistences. Thus, we further restrict the bijection to map from $I_0(\Theta)$ to $I_0(P)$ and $I_1(\Theta)$ to $I_1(P)$. Subsequently, it is equivalent to optimize separately as

$$\mathcal{L}_{top}(\Theta, P) = \min_{\tau_0} \sum_{\epsilon_b \in I_0(\Theta)} [\epsilon_b - \tau_0(\epsilon_b)]^2 + \min_{\tau_1} \sum_{\epsilon_d \in I_1(\Theta)} [\epsilon_d - \tau_1(\epsilon_d)]^2,$$

where τ_0 is a bijection from $I_0(\Theta)$ to $I_0(P)$ and τ_1 is a bijection from $I_1(\Theta)$ to $I_1(P)$. We will call the first term as 0D topological loss, which measures the topological similarity between the two networks Θ and P using the difference in 0D barcodes and is denoted by $\mathcal{L}_{0D}(\Theta, P)$. Similarly, we will call the second term as 1D topological loss, which measures the topological similarity through the difference of 1D barcodes and is denoted by $\mathcal{L}_{1D}(\Theta, P)$. The 0D topological loss \mathcal{L}_{0D} is a variation to the standard *assignment problem* and usually solved in a greedy fashion using Hungarian algorithm in $O(|I_0(\Theta)|^3)$, or equivalently $O(|V^\Theta|^3)$, in combinatorial optimization (Edmonds and Karp, 1972). However, for 1-skeletons, minimum matching is given exactly and can be numerically computed in $O(|I_0(\Theta)| \log |I_0(\Theta)|)$ time:

Theorem 2.

$$\min_{\tau_0} \sum_{\epsilon_b \in I_0(\Theta)} [\epsilon_b - \tau_0(\epsilon_b)]^2 = \sum_{\epsilon_b \in I_0(\Theta)} [\epsilon_b - \tau_0^*(\epsilon_b)]^2,$$

where τ_0^* maps the i -th smallest birth value in $I_0(\Theta)$ to the i -th smallest birth value in $I_0(P)$ for all i .

The minimization in Theorem 2 is equivalent to the following assignment problem. For monotonic sequences

$$a_1 < a_2 < \dots < a_n, \quad b_1 < b_2 < \dots < b_n,$$

we consider finding $\min_{\tau} \sum_{i=1}^n (a_i - \tau(a_i))^2$ over all possible bijections τ . The minimization is equivalent to maximizing the assignment cost $\sum_{i=1}^n a_i \tau(a_i)$. Then, we follow an inductive argument. For $n = 2$, there are two possible assignments: $a_1 b_1 + a_2 b_2$ or $a_1 b_2 + a_2 b_1$. Since $a_1 b_1 + a_2 b_2 > a_1 b_2 + a_2 b_1$, $\tau(a_i) = b_i$ is the

optimal matching. For $n = k$, we assume $\tau(a_i) = b_i$ is the optimal matching for all i . For $n = k + 1$, the cost is given by

$$\max_{\tau} \sum_{i=1}^{k+1} a_i \tau(a_i) \leq \max_{\tau} \sum_{i=1}^k a_i \tau(a_i) + \max_{\tau} a_{k+1} \tau(a_{k+1}).$$

The first term is maximized if $\tau(a_i) = b_i$. The second term is maximized if $\tau(a_{k+1}) = b_{k+1}$. Thus, we have proved the statement. The immediate consequence of the theorem is that the loss function \mathcal{L}_{0D} is symmetric: $\mathcal{L}_{0D}(\Theta, P) = \mathcal{L}_{0D}(P, \Theta)$.

Similarly for the 1D topological loss \mathcal{L}_{1D} , we also have

Theorem 3.

$$\min_{\tau_1} \sum_{\epsilon_d \in I_1(\Theta)} [\epsilon_d - \tau_1(\epsilon_d)]^2 = \sum_{\epsilon_d \in I_1(\Theta)} [\epsilon_d - \tau_1^*(\epsilon_d)]^2,$$

where τ_1^* maps the i -th smallest death value in $I_1(\Theta)$ to the i -th smallest death value in $I_1(P)$ for all i .

Subsequently, the loss function \mathcal{L}_{1D} is symmetric: $\mathcal{L}_{1D}(\Theta, P) = \mathcal{L}_{1D}(P, \Theta)$. Similar to τ_0^* , τ_1^* can be computed in $O(|I_1(\Theta)| \log |I_1(\Theta)|)$ time.

Augmentation: networks with different sizes

When networks have different sizes in terms of the number of nodes, we can still find the optimal matching through data augmentation on smaller networks. Similar data augmentations are done in matching various topological features of different sizes (Hu et al., 2019), matching trees of different sizes (Guo and Srivastava, 2020) or matching point sets of different sizes (Chung et al., 2019b).

Let $\Theta = (V^\Theta, w^\Theta)$ and $P = (V^P, w^P)$ be two networks of different sizes, i.e., $|V^\Theta| \neq |V^P|$. We have two cases as follows.

Case $|V^\Theta| > |V^P|$: Since $|I_0(\Theta)| > |I_0(P)|$, some birth values in Θ may not have corresponding matches in P . Connected components that are born later (larger birth values) have shorter persistences (life-times), which are considered as topological noise (Ghrist, 2008; Clough et al., 2019; Hu et al., 2019). Thus, we need to map any unmatched, short-lived connected components in Θ to the least persistent connected component in P and identify τ_0^* . In particular, we match the first $|I_0(P)|$ smallest birth values in Θ to all the birth values in P using Theorem 2. We then match the remaining unmatched, large birth values in Θ to the largest edge weight in P . Similarly, we match the first $|I_1(P)|$ largest death values in Θ to all the death values in P using Theorem 3. We then match

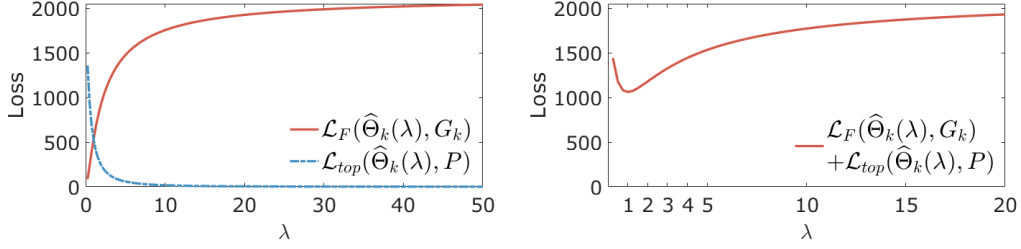


Figure 3: Topological learning using a functional brain network of a subject. When $\lambda = 0$, $\hat{\Theta}_k$ is simply equal to G_k . As λ increases, $\hat{\Theta}_k$ is deformed such that the topology of $\hat{\Theta}_k$ is closer to P . For this subject, the total loss function obtains the global minimum at $\lambda = 0.9997$. Other subjects show almost identical patterns.

any unmatched, small death values in Θ to the smallest edge weight in P and identify τ_1^* .

Case $|V^\Theta| < |V^P|$: Since $|I_0(\Theta)| < |I_0(P)|$, some birth and death values in P may not have corresponding matches in Θ . Similar to the case $|V^\Theta| > |V^P|$, any unmatched birth or death values in P are matched to the largest or smallest edge weights in Θ , respectively.

Topological Learning

Most learning tasks such as regression, classification and clustering can be performed by minimizing the topological loss. Let $G_1 = (V, w^1), \dots, G_n = (V, w^n)$ be the observed networks with identical node set V that will be used as *training networks*. Let $P = (V^P, w^P)$ be a network expressing a prior topological knowledge. In brain network analysis (Lv et al., 2010; Zhu et al., 2014; Kang et al., 2017), G_k can be the functional brain network of the k -th subject obtained from the resting-state fMRI and P can be the template structural brain network obtained through dMRI, where functional brain networks are often overlaid (Figure 1). The node sets V and V^P may differ. This can happen if we try to integrate brain networks obtained from different parcellations and studies.

We are interested in learning the model $\Theta = (V^\Theta, w^\Theta)$ from the set of training data. At the subject level, we train the model using an individual network G_k by minimizing

$$\hat{\Theta}_k = \arg \min_{\Theta} \mathcal{L}_F(\Theta, G_k) + \lambda \mathcal{L}_{top}(\Theta, P), \quad (1)$$

where the squared Frobenius loss $\mathcal{L}_F(\Theta, G_k) = \|w^\Theta - w^k\|_F^2$ is the goodness-of-fit term between the model and the individual observation. The parameter λ controls the amount of topological information of network P we are introducing

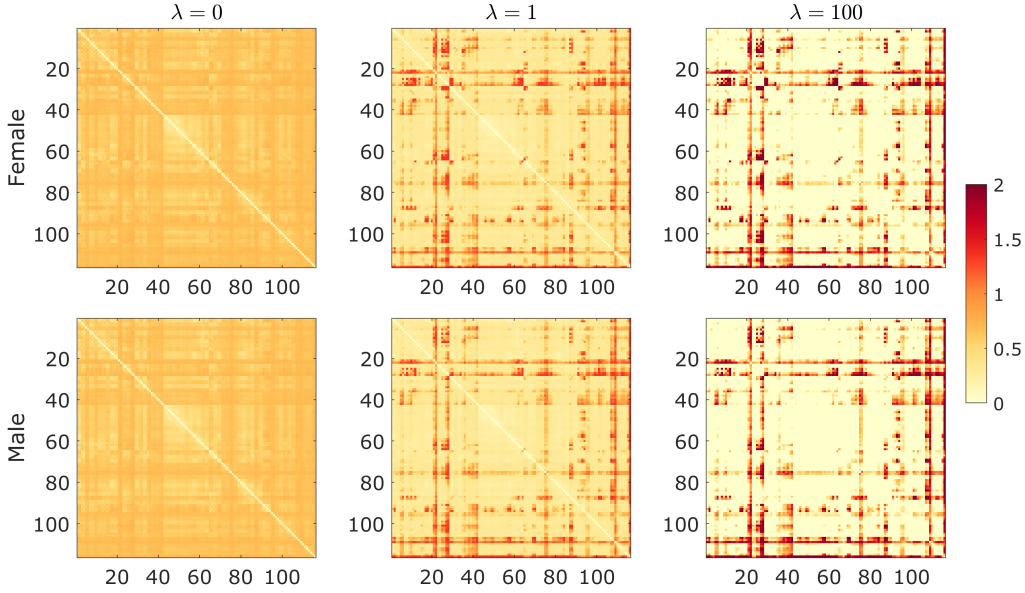


Figure 4: Group-level networks of female (top row) and male (bottom row) are estimated by minimizing the objective function (2) with different $\lambda = 0, 1$ and 100.

to the model. The larger the value of λ , the more we are learning toward P . If $\lambda = 0$, we no longer learn the topology of P but simply fit data to the individual network G_k . The total loss is the sum of quadratic losses and thus it obtains the global minimum. Figure 3 shows the total loss function with the global minimum at $\lambda = 0.9997$ for a subject. The average global minimum across all subjects is obtained at $\lambda = 1.0000 \pm 0.0002$ showing highly stable result. Thus, we simply use $\lambda = 1.0000$ for all subjects. Such stable result is not possible in non-topological loss functions. Similar to the stability results in persistent homology (Cohen-Steiner et al., 2010), we can prove that

$$\mathcal{L}_{top}(\Theta, P) \leq C \|w^\Theta - w^P\|_F^2$$

for some C providing the stability of topological loss. For real data we used in the study, we have the least upper bound $C = 0.7285 \pm 0.0455$ over all subjects.

Although the group-level learning is not the focus of this study, we can also learn network $\hat{\Theta}$ using all the training data such that

$$\hat{\Theta} = \arg \min_{\Theta} \frac{1}{n} \sum_{k=1}^n \mathcal{L}_F(\Theta, G_k) + \lambda \mathcal{L}_{top}(\Theta, P) \quad (2)$$

Figure 4 displays the average networks of females and males by minimizing the

objective function (2) with different $\lambda = 0, 1$ and 100. The larger the value of λ , the more we are reinforcing the topology of structural brain network that is sparse onto the functional brain network (Figure 5). Even though we will not show in this paper, the statistical significance of network differences between females and males can be determined by using the exact topological inference developed in our previous work (Chung et al., 2017b, 2019b).

Averaging networks of different sizes and topology

As an application of the proposed topological learning framework, it is also possible to average networks of different sizes and topology directly without aligning to the template network as in model (2). This might be useful in a situation when we do not have the template or it is not necessary to align networks to a template. Averaging networks of different topology is a difficult task using existing methods.

Given n networks $G_1 = (V_1, w^1), \dots, G_n = (V_n, w^n)$ with different node sets, we are interested in obtaining its average, which we will call the *topological mean*. Since the size and topology of the networks are different, we cannot simply average edge weight matrices w^1, \dots, w^n directly. Motivated by Fréchet mean (Le and Kume, 2000; Turner et al., 2014; Zemel and Panaretos, 2019), we obtain the topological mean $\hat{\Theta}$ by minimizing the sum of topological losses

$$\hat{\Theta} = \arg \min_{\Theta} \sum_{k=1}^n \mathcal{L}_{top}(\Theta, G_k) = \arg \min_{\Theta} \sum_{k=1}^n \left[\mathcal{L}_{0D}(\Theta, G_k) + \mathcal{L}_{1D}(\Theta, G_k) \right]. \quad (3)$$

$\hat{\Theta}$ is viewed as a network that is the topological centroid of n networks. The optimization can be done analytically as follows (Rabin et al., 2011).

For now, we assume the same number of nodes in the networks, which gives the same m_0 number of birth values. The 0D topological loss \mathcal{L}_{0D} depends on the birth values of G_1, \dots, G_n . Let $b_{k1} < b_{k2} < \dots < b_{km_0}$ be the birth values of network G_k . Let $\theta_1 < \theta_2 < \dots < \theta_{m_0}$ be the birth values of network Θ . By Theorem 2, the first term is equivalent to

$$\sum_{k=1}^n \mathcal{L}_{0D}(\Theta, G_k) = \sum_{k=1}^n (\theta_1 - b_{k1})^2 + \sum_{k=1}^n (\theta_2 - b_{k2})^2 + \dots + \sum_{k=1}^n (\theta_{m_0} - b_{km_0})^2.$$

This is a quadratic so we can simply find the minimum by setting its derivative equal to zero. This is given by $\hat{\theta}_j = \sum_{k=1}^n b_{kj} / n$. For the second term, it is similarly represented as the sum of squared differences of death values (Theorem 3). Thus, the i -th smallest birth (or death) value of the topological mean network $\hat{\Theta}$ is given by the mean of all the i -th smallest birth (or death) values of n networks. Given all the birth and death values of a network, we can completely

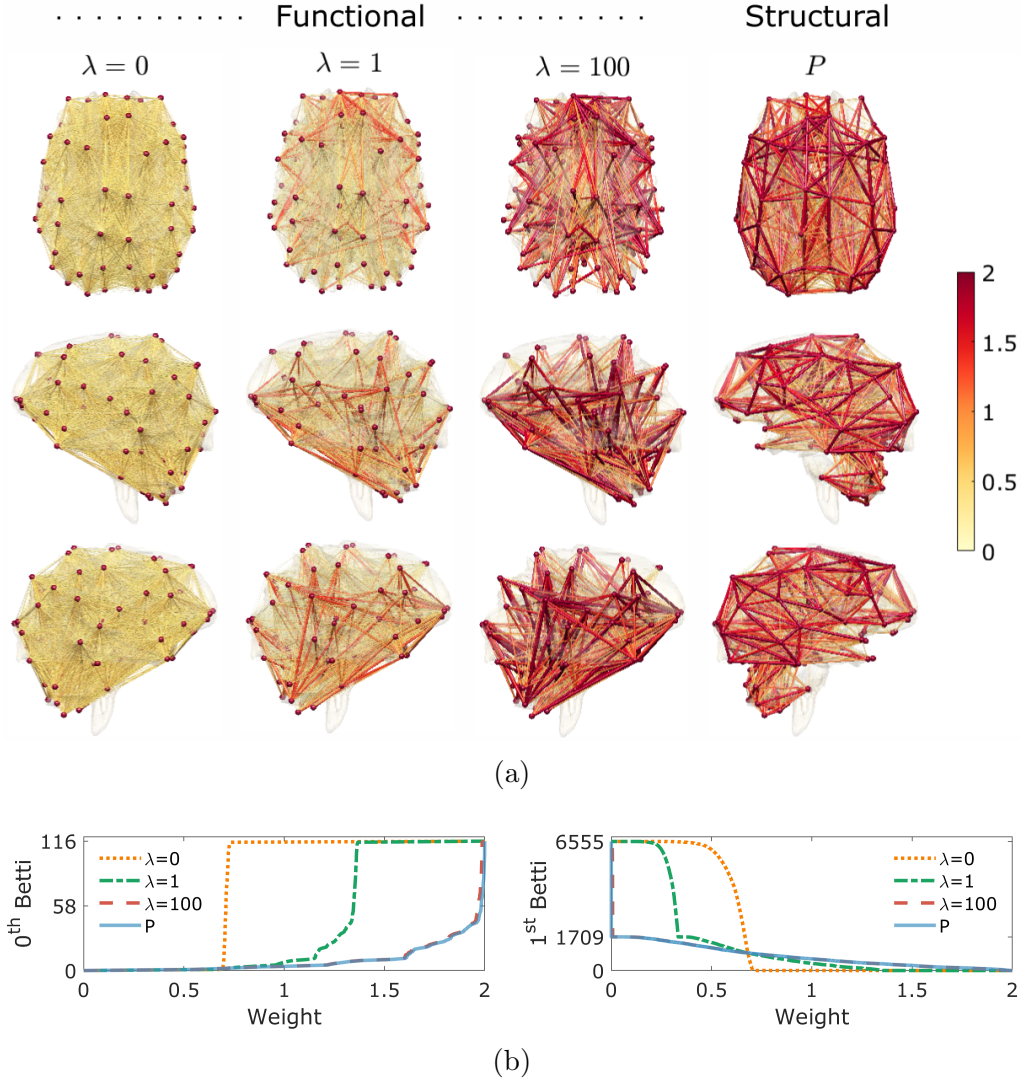


Figure 5: (a) The group-level networks learned by minimizing the objective function (2) over all subjects with different $\lambda = 0, 1$ and 100. The template structural network P is shown in the last column. (b) As λ increases, Betti-plots of the group-level network are adjusted toward that of P . β_0 -plot shows that the connected components in the structural network P are gradually born over a wide range of edge weights during graph filtration. β_1 -plot shows topological sparsity or lack of cycles in the structural network P . While the group-level functional network (when $\lambda = 0$) is densely connected with 6555 maximum number of cycles, the structural network is sparsely connected with only 1709 cycles.

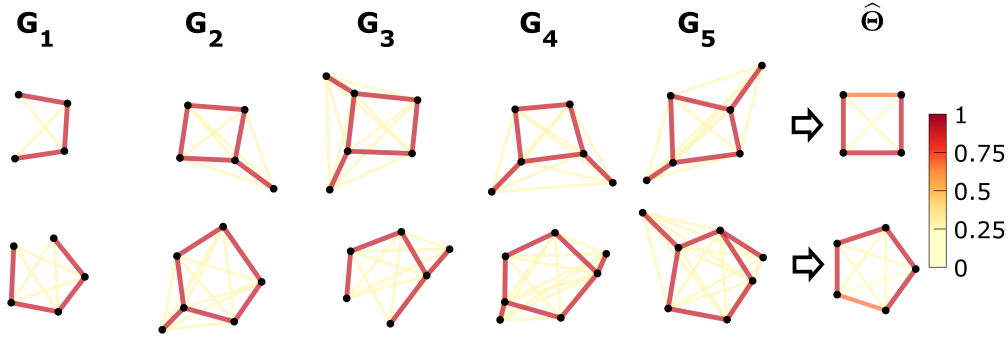


Figure 6: Examples of averaging networks of different sizes and topology using the proposed topological averaging. The topological mean network $\hat{\Theta}$ (right) is the topological centroid of five networks G_1, \dots, G_5 (left) showing the average topological pattern. The topological mean network $\hat{\Theta}$ is estimated by minimizing the sum of topological losses $\min_{\Theta} \sum_{k=1}^5 \mathcal{L}_{top}(\Theta, G_k)$. The topological mean network $\hat{\Theta}$ highlights topological characterization of the five networks. The existing methods will have difficulty averaging networks of different sizes and topology.

recover topology of the network. For networks with different number of nodes, we augment birth and death values. Figure 6 illustrates toy examples of averaging networks of different sizes and topology.

Numerical implementation: gradient descent

The topological learning (1) estimates $\Theta = (V^{\Theta}, w^{\Theta})$ iteratively through gradient descent (Bottou, 1998). The gradient of the topological loss can be computed efficiently without taking numerical derivatives and its computation mainly comprises the computation of barcodes I_0 and I_1 , and finding the optimal matchings through Theorems 2 and 3. The gradient of the topological loss $\nabla \mathcal{L}_{top}(\Theta, P)$ with respect to edge weight $w^{\Theta} = (w_{ij}^{\Theta})$ is given as a gradient matrix whose ij -th entry is

$$\frac{\partial \mathcal{L}_{top}(\Theta, P)}{\partial w_{ij}^{\Theta}} = \begin{cases} 2[w_{ij}^{\Theta} - \tau_0^*(w_{ij}^{\Theta})] & \text{if } w_{ij}^{\Theta} \in I_0(\Theta); \\ 2[w_{ij}^{\Theta} - \tau_1^*(w_{ij}^{\Theta})] & \text{if } w_{ij}^{\Theta} \in I_1(\Theta) \end{cases}$$

since $I_0(\Theta)$ and $I_1(\Theta)$ partition the weight set (Theorem 1). Intuitively, by slightly adjusting the edge weight w_{ij}^{Θ} , we have the slight adjustment of either a birth value in 0D barcode or a death value in 1D barcode, which slightly changes the topology of the network. During the estimation of Θ , we take steps in the

direction of negative gradient:

$$w_{ij}^\Theta \rightarrow w_{ij}^\Theta - 0.1 \left(2(w_{ij}^\Theta - w_{ij}^k) + \lambda \frac{\partial \mathcal{L}_{top}(\Theta, P)}{\partial w_{ij}^\Theta} \right),$$

where 0.1 is the learning rate. As w_{ij}^Θ are moved closer to its optimal match, the topology of the estimated network Θ gets closer to that of P while the Frobenius norm keeps the estimation Θ close to the observed network G_k .

Finding 0D birth values $I_0(G)$ is *equivalent* to finding edge weights comprising the *maximum spanning tree* (MST) of G (Lee et al., 2012). Once I_0 is computed, I_1 is simply given as the rest of the remaining edge weights (Theorem 1). Then, we can compute the optimal matchings τ_0^* and τ_1^* between Θ and P by simply sorting edge weights in the ascending order and matching them. The computational complexity for the topological loss gradient is dominated by the computation of the MST using the popular algorithms such as Prim’s and Kruskal’s. This takes $O(|E^\Theta| \log |V^\Theta|)$ if $|V^\Theta| \geq |V^P|$ or $O(|E^P| \log |V^P|)$ if $|V^\Theta| < |V^P|$.

3 Validation

For validation, we performed two simulation studies to assess the performance of the topological loss as a similarity measure between networks of different topology. Networks of the same size were simulated since many existing methods cannot handle networks of different sizes.

Study 1: Random network model with ground truth

Initial data vector b_i at node i was simulated as independent and identically distributed multivariate normal across n subjects, i.e., $b_i \sim \mathcal{N}(0, I_n)$ with the identity matrix I_n as the covariance matrix of size $n \times n$. The new data vector x_i at node i was then generated by introducing additional dependency structures to b_i through a mixed-effects model that partitions the covariance matrix of x_i into c blocks forming modular structures (Figure 7) (Snijders et al., 1995):

$$\begin{aligned} x_1, \dots, x_a &= b_1 + \mathcal{N}(0, \sigma^2 I_n), \\ x_{a+1}, \dots, x_{2a} &= b_{a+1} + \mathcal{N}(0, \sigma^2 I_n), \\ &\vdots \\ x_{(c-1)a+1}, \dots, x_{ca} &= b_{(c-1)a+1} + \mathcal{N}(0, \sigma^2 I_n), \end{aligned}$$

where a is the number of nodes in each module. In our simulation, we used $c = 2, 5, 10, 20$ and $a = 50, 20, 10, 5$ respectively, which give exactly $ca = 100$

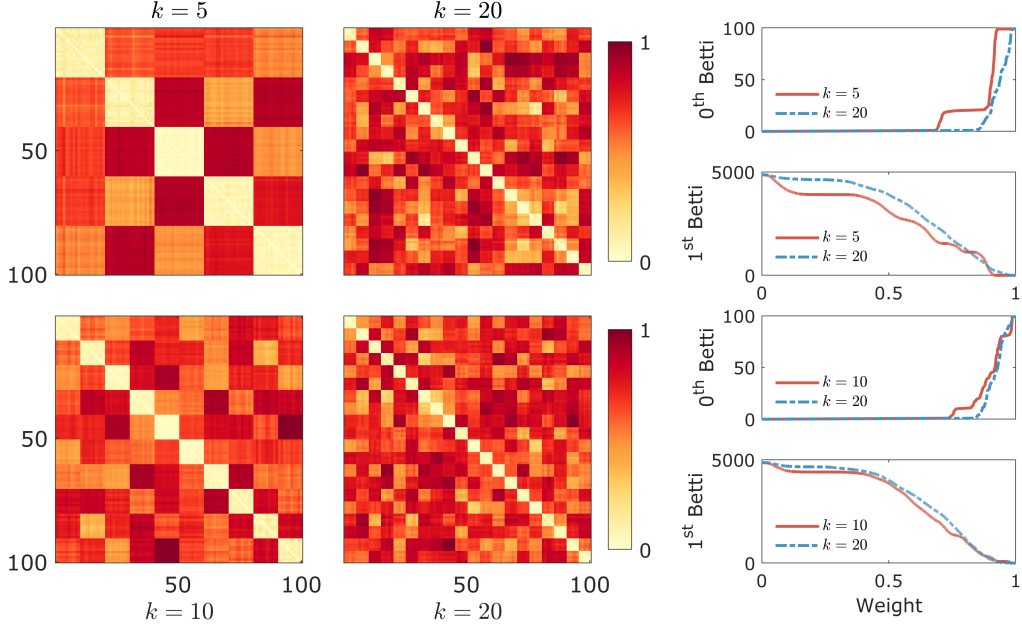


Figure 7: Study 1. Network difference. The comparison of networks with different topology: $k = 5$ vs. 20 (first row) and 10 vs. 20 (second row). The change of Betti numbers over filtration values clearly shows topology difference. The topological difference in the second row is subtle compared to the first row.

nodes across networks. $\sigma = 0.1$ was universally used as network variability. Then we computed Pearson correlation coefficient ρ_{ij}^x between x_i and x_j , which was then translated and scaled as $w_{ij}^x = \sqrt{(1 - \rho_{ij}^x)/2}$ as a metric (Chung et al., 2019b). This gives a block network $\mathcal{X} = (V, w^x)$. The mixed-effects model allows us to explicitly simulate the amount of statistical dependency between modules and nodes, providing the control over topological structures of connectedness.

Based on the statistical model above, we simulated two groups of networks consisting of $n = 7$ subjects in each group for two different studies. We then tested the performance of topological loss for network differences. For comparison, we tested the topological loss against widely-used Euclidean losses such as \mathcal{L}_1 -, \mathcal{L}_2 - and \mathcal{L}_∞ -norms. We also compared against other topological distances such as bottleneck, Gromov-Hausdorff (GH) and Kolmogorov-Smirnov (KS) distances (Cohen-Steiner et al., 2007; Chazal et al., 2009; Chung et al., 2017a). The bottleneck and GH-distance are two widely-used baseline distances in persistent homology often used in persistent diagrams and dendrograms (Carlsson and Mémoli, 2010), and brain networks (Lee et al., 2012). KS-distance based on

Table 1: Study 1. The performance results are summarized in terms of false negative rate (top rows) and false positive rate (bottom rows).

c	\mathcal{L}_1	\mathcal{L}_2	\mathcal{L}_∞	GH	Bottleneck	$\text{KS}(\beta_0)$	$\text{KS}(\beta_1)$	\mathcal{L}_{top}
2 vs. 10	0.02	0.00	0.02	0.30	0.40	0.00	0.00	0.02
5 vs. 20	0.06	0.00	0.02	0.20	0.32	0.02	0.00	0.00
10 vs. 20	1.00	0.86	0.34	0.32	0.20	0.24	0.00	0.08
2 vs. 2	0.00	0.00	0.00	0.00	0.00	0.54	0.88	0.00
5 vs. 5	0.00	0.00	0.00	0.00	0.00	0.10	0.42	0.00
10 vs. 10	0.00	0.00	0.00	0.00	0.00	0.02	0.12	0.00

β_0 and β_1 curves is later introduced as a more intuitive alternative that gives results that are easier to interpret (Chung et al., 2013). KS-distance was successfully applied to perform statistical inference on both β_0 and β_1 in a large-scale brain network studies (Chung et al., 2019b). For KS-distance, we computed its probability distribution exactly (Chung et al., 2017b). For all other distances, the permutation test was used. Since the sample size is small, the exact permutation test can be done by generating exactly $\binom{14}{7} = 3432$ number of every possible permutation by shuffling the group labels to determine the empirical distribution of the two-sample t -statistic under the null.

Network difference. We compared networks generated by varying parameter $c = 2$ vs. 10, 5 vs. 20 and 10 vs. 20. Since the networks had different topological structures, the distances were expected to detect the differences (Figure 7). The simulations were independently performed 50 times and the performance results were given in terms of the false negative rate computed as the fraction of 50 simulations that gave p -values above 0.05 (Table 1). As expected, topological loss performed very well in all settings while other distances were less sensitive to more subtle topological difference. Even though the KS-distance on cycles (β_1) performed better, it is known to be overly sensitive distance (Chung et al., 2019b) and often produces false positives as shown in the next simulation.

No network difference. We compared networks generated by unvarying parameter $c = 2$ vs. 2, 5 vs. 5 and 10 vs. 10, which should give networks of similar topology in each group. It was expected that the networks were not topologically different and we should not detect the network differences. The simulations were independently performed 50 times and the performance results were given in terms of the false positive rate computed as the fraction of 50 simulations that gave p -values below 0.05 (Table 1). Except KS-distance, all the distances performed well.

While all the loss functions performed well when there was no network dif-

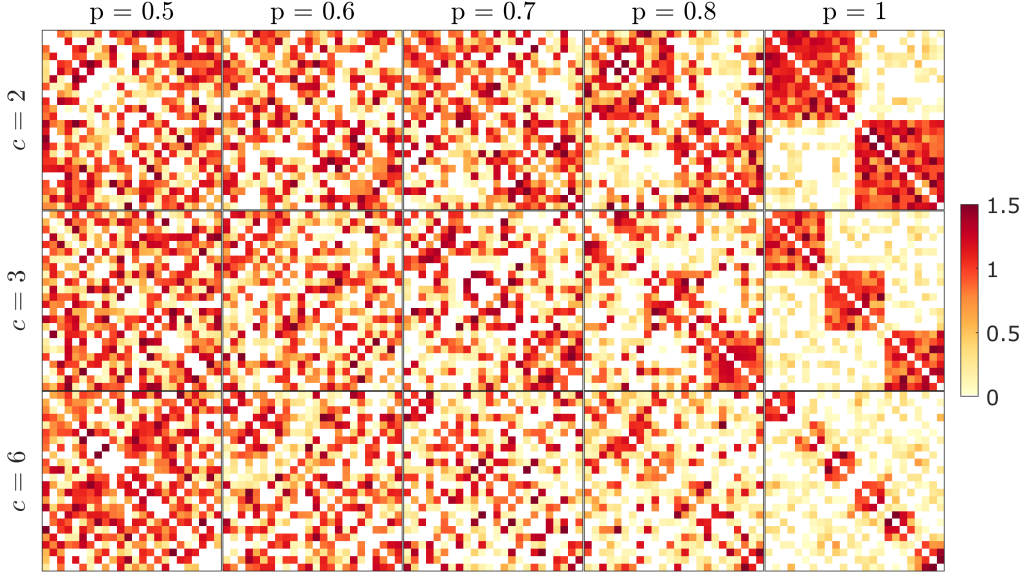


Figure 8: Study 2 simulation examples. Network modular structure varies as parameters p (probability of connection within modules) and c (the number of modules) change. The modular structure becomes more pronounced as p increases.

ferences, the Euclidean losses, bottleneck and GH-distance had the tendency to produce false negatives when there was network difference. On the other hand, KS-distance tended to produce false positives when there is no network difference. Overall, the proposed topological loss performed well.

Study 2: Comparison against graph matching

The aim of this simulation is to evaluate the performance of the proposed topological matching process against existing graph matching algorithms (Gold and Rangarajan, 1996; Leordeanu and Hebert, 2005; Leordeanu et al., 2009; Cho et al., 2010; Zhou and De la Torre, 2013) in differentiating networks of different topology. Graph matching algorithms are considered as the *baseline* for establishing correspondence between graphs and are usually done by penalizing network structures that cannot be exactly matched. Let $G_1 = (V_1, w^1)$ and $G_2 = (V_2, w^2)$ be two networks. For a graph matching problem modified for our brain network setting, where the edge weights are not binary but weighted, we need to find mapping τ_{gm} between nodes $i_1, j_1 \in V_1$ and $i_2, j_2 \in V_2$ that best preserves edge attributes between edge weights $w_{i_1 j_1}^1 \in w^1$ and $w_{i_2 j_2}^2 \in w^2$. In other words, we

seek τ_{gm} that maximizes the graph matching cost

$$J(\tau_{gm}) = \sum_{w_{i_1 j_1}^1, w_{i_2 j_2}^2} f(w_{i_1 j_1}^1, \tau_{gm}(w_{i_2 j_2}^2)),$$

where f measures the similarity between edge attributes and the summation is taken over all possible edge weights. The matching cost $J(\tau_{gm})$ quantifies similarity between two networks by taking large values for similar networks and values close to zero for dissimilar networks hence $J(\tau_{gm})$ is somewhat the inverse of distance metrics. We compared the proposed topological loss against four well-known graph matching algorithms: graduated assignment (GA) (Gold and Rangarajan, 1996), spectral matching (SM) (Leordeanu and Hebert, 2005), integer projected fixed point method (IPFP) (Leordeanu et al., 2009) and re-weighted random walk matching (RRWM) (Cho et al., 2010). Such graph matching methods are widely used as baseline algorithms in medical imaging, computer vision and machine learning studies (Cour et al., 2006; Tian et al., 2012; Zhou and De la Torre, 2013; Yu et al., 2018; Zhang et al., 2019; Wang et al., 2020). For all the baseline methods, we used existing implementation codes from authors' repository websites listed in the publication. We also used parameters recommended in the public code for each baseline algorithm *without* modification. Since we are dealing with weighted edges, the graph matching algorithms based on binary edge weight are excluded in the study (Babai and Luks, 1983; Zavlanos and Pappas, 2008; Guo and Srivastava, 2020).

In study 2, a different random network model from study 1 is used. We simulate random modular network \mathcal{X} with d number of nodes and c number of modules, where the nodes are evenly distributed among modules. Figure 8 displays modular networks with $d = 24$ nodes and $c = 2, 3, 6$ modules such that we have $d/c = 12, 8, 4$ number of nodes in each module respectively. Since time complexity of the aforementioned graph matching algorithms can be very demanding (Figure 9), we considered $d = 12, 18, 24$ and $c = 2, 3, 6$ in this simulation. Each edge connecting two nodes within the same module was then assigned a random weight following a normal distribution $\mathcal{N}(\mu, \sigma^2)$ with probability p or otherwise Gaussian noise $\mathcal{N}(0, \sigma^2)$ with probability $1 - p$. On the other hand, edge weights connecting nodes between different modules had probability $1 - p$ of being $\mathcal{N}(\mu, \sigma^2)$ and probability p of being $\mathcal{N}(0, \sigma^2)$. With larger value of within-module probability p , we have more pronounced modular structure. Any negative edge weights were set to zero. This gives the random network \mathcal{X} that exhibits topological structures of connectedness. Figure 8 illustrates changes of network modular structure as parameters p and c vary. We used $\mu = 1$ and $\sigma = 0.25$ universally throughout study 2.

Based on the statistical model above, we simulated two groups of random modular networks $\mathcal{X}_1, \dots, \mathcal{X}_m$ and $\mathcal{Y}_1, \dots, \mathcal{Y}_n$. If there is group difference, the

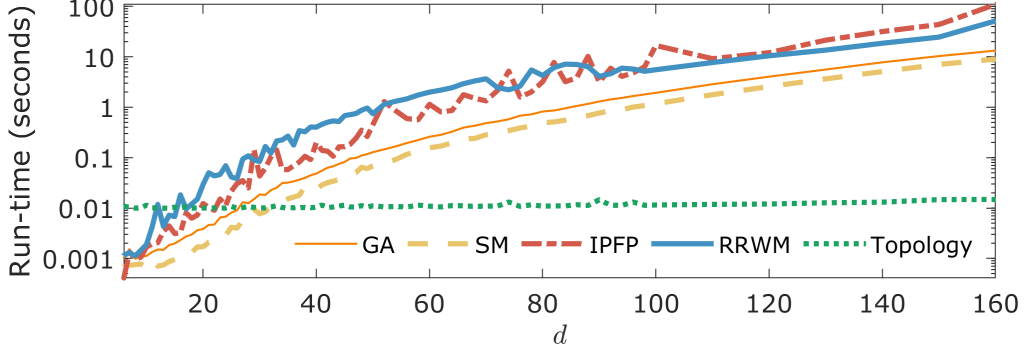


Figure 9: Study 2 run-time – the amount of time each algorithm takes to compute its matching cost between two modular networks of size d . Average run-times of topological loss and other baseline graph matching algorithms are plotted over d . The run-time performance of the baseline methods is consistent with [Cour et al. \(2006\)](#) for GA and SM, and [Zhang et al. \(2019\)](#) for IPFP and RRWM.

topological loss is expected to be relatively small within groups and relatively large between groups. The average topological loss within group given by

$$\bar{\mathcal{L}}_W = \frac{\sum_{i < j} \mathcal{L}(\mathcal{X}_i, \mathcal{X}_j) + \sum_{i < j} \mathcal{L}(\mathcal{Y}_i, \mathcal{Y}_j)}{\binom{m}{2} + \binom{n}{2}}$$

is expected to be smaller than the average topological loss between groups given by

$$\bar{\mathcal{L}}_B = \frac{\sum_{i=1}^m \sum_{j=1}^n \mathcal{L}(\mathcal{X}_i, \mathcal{Y}_j)}{mn}.$$

We measure the disparity between groups as the ratio statistic $\phi_{\mathcal{L}}$

$$\phi_{\mathcal{L}} = \bar{\mathcal{L}}_B / \bar{\mathcal{L}}_W.$$

If $\phi_{\mathcal{L}}$ is large, the groups differ significantly in network topology. On the other hand, if $\phi_{\mathcal{L}}$ is small, it is likely that there is no group differences. Similarly, we define the ratio statistic for graph matching cost J as

$$\phi_J = \bar{J}_W / \bar{J}_B,$$

where \bar{J}_W is the average graph matching cost within groups and \bar{J}_B is the average graph matching cost between groups. Since the distributions of the ratio statistics $\phi_{\mathcal{L}}$ and ϕ_J are unknown, the permutation test is used to determine the empirical

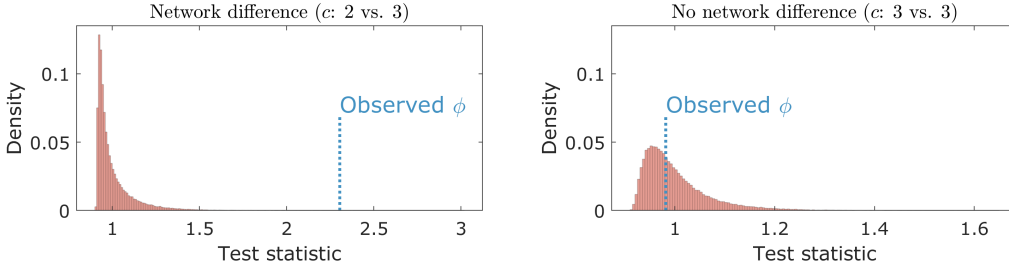


Figure 10: Empirical distributions in Study 2. The empirical distributions of the ratio statistic were generated by the permutation on two groups each consisting of 10 modular networks. Here we tested if there is group difference in networks with varying parameter $c = 2$ vs. 3 (left) and 3 vs. 3 (right). As expected, the test based on the topological loss rejected the null hypothesis when there is group difference.

distributions. Figure 10 displays the empirical distribution of $\phi_{\mathcal{L}}$. By comparing the observed group ratio $\phi_{\mathcal{L}}$ to this empirical distribution, we can determine the statistical significance of testing the group difference. However, when the sample size is large, existing matching algorithms are too slow for the permutation test. So we adapted for a scalable computation strategy as follows (Chung et al., 2019c).

Given two groups of networks, topological loss (or graph matching cost) for every pair of networks needs to be computed only once, which can then be arranged into a matrix whose rows and columns are networks, and the ij -entry is the loss between two networks corresponding to row i and column j (Figure 11). Once we obtain such matrix, the permutation process is equivalent to rearranging rows and columns based on permuted group labels. There are $\frac{1}{2} \binom{m+n}{m}$ total number of permutations excluding the symmetry of loss functions. Computing the ratio statistic over a permutation requires re-summing over all such losses, which is time consuming. Instead, we can perform the transposition procedure of swapping only one network per group and setting up iteration of how the ratio statistic changes over the transposition (Chung et al., 2019c).

Let $\mathcal{X} = (\mathcal{X}_1, \dots, \mathcal{X}_m)$ and $\mathcal{Y} = (\mathcal{Y}_1, \dots, \mathcal{Y}_n)$ be the two groups of networks. We *transpose* k -th and l -th networks between the groups as

$$\begin{aligned}\pi_{kl}(\mathcal{X}) &= (\mathcal{X}_1, \dots, \mathcal{X}_{k-1}, \mathcal{Y}_l, \mathcal{X}_{k+1}, \dots, \mathcal{X}_m), \\ \pi_{kl}(\mathcal{Y}) &= (\mathcal{Y}_1, \dots, \mathcal{Y}_{l-1}, \mathcal{X}_k, \mathcal{Y}_{l+1}, \dots, \mathcal{Y}_n).\end{aligned}$$

Over transposition π_{kl} , the ratio statistic is changed from $\phi_L(\mathcal{X}, \mathcal{Y})$ to $\phi_L(\pi_{kl}(\mathcal{X}), \pi_{kl}(\mathcal{Y}))$,

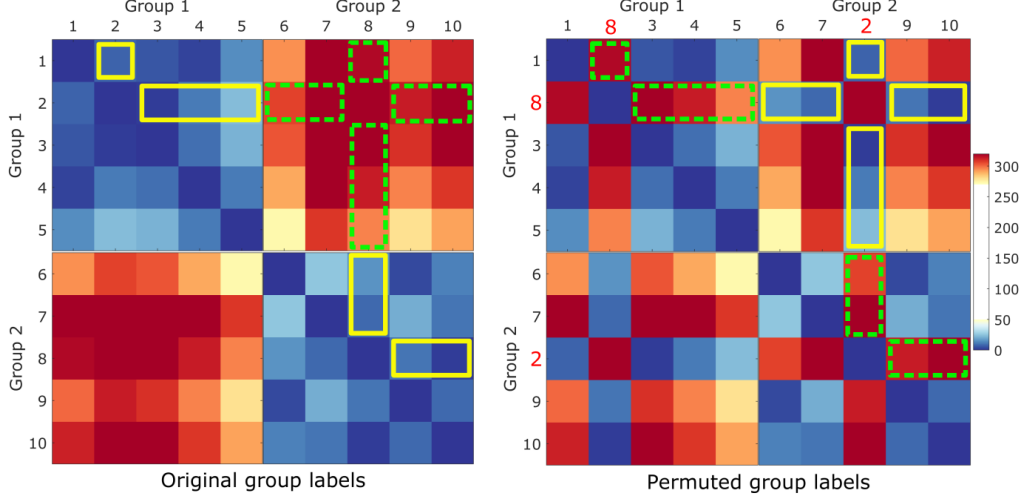


Figure 11: Study 2. Two groups each consisting of 5 modular networks simulated with parameters $c = 2$ vs. 3, resulting in small and large topological losses within groups and between groups respectively. Left: the matrix whose ij -th entry represents the loss between network i and j . The main diagonal consists of zero since topological loss between two identical networks is zero. Right: a transposition between the 2-nd network in Group 1 and the 8-th network in Group 2. We do not need to recompute all the pairwise losses again but just rearrange them. In particular, only the pairwise losses in the yellow solid lines and green dashed lines are rearranged. Thus, we simply need to figure out how the rearrangement of entries changes the ratio statistic in an iterative manner. This enables us to easily perform the permutation test in a scalable fashion. We compute δ in equation (4) by subtracting the sum of entries within the yellow lines from the sum of entries within the green lines.

which involves the following functions:

$$\begin{aligned}\nu(\mathcal{X}, \mathcal{Y}) &= \sum_{i < j} \mathcal{L}(\mathcal{X}_i, \mathcal{X}_j) + \sum_{i < j} \mathcal{L}(\mathcal{Y}_i, \mathcal{Y}_j), \\ \omega(\mathcal{X}, \mathcal{Y}) &= \sum_{i=1}^m \sum_{j=1}^n \mathcal{L}(\mathcal{X}_i, \mathcal{Y}_j),\end{aligned}$$

where ν is the total sum of within-group losses and ω is the total sum of between-group losses. Then we determine how ν and ω change over the transposition π_{kl} . As \mathcal{X}_k and \mathcal{Y}_l are swapped, the function ν is updated over the transposition π_{kl} as (Figure 11)

$$\nu(\pi_{kl}(\mathcal{X}), \pi_{kl}(\mathcal{Y})) = \nu(\mathcal{X}, \mathcal{Y}) + \delta(\mathcal{X}, \mathcal{Y})$$

with

$$\delta(\mathcal{X}, \mathcal{Y}) = \left(\sum_{i \neq k} \mathcal{L}(\mathcal{Y}_l, \mathcal{X}_i) - \sum_{i \neq k} \mathcal{L}(\mathcal{X}_k, \mathcal{X}_i) \right) + \left(\sum_{i \neq l} \mathcal{L}(\mathcal{X}_k, \mathcal{Y}_i) - \sum_{i \neq l} \mathcal{L}(\mathcal{Y}_l, \mathcal{Y}_i) \right). \quad (4)$$

Similarly, function ω are updated iteratively over the transposition π_{kl} as

$$\omega(\pi_{kl}(\mathcal{X}), \pi_{kl}(\mathcal{Y})) = \omega(\mathcal{X}, \mathcal{Y}) - \delta(\mathcal{X}, \mathcal{Y}).$$

The ratio statistic over the transposition is then computed as

$$\phi_{\mathcal{L}}(\pi_{kl}(\mathcal{X}), \pi_{kl}(\mathcal{Y})) = \frac{\omega(\pi_{kl}(\mathcal{X}), \pi_{kl}(\mathcal{Y}))}{\nu(\pi_{kl}(\mathcal{X}), \pi_{kl}(\mathcal{Y}))} \times \frac{\binom{m}{2} + \binom{n}{2}}{mn}.$$

For each transposition, we store information about the function values ν and ω , and update them sequentially. Each transposition requires manipulating $2(m + n - 2)$ terms as opposed to $\binom{m+n}{2}$ total number of terms over a random permutation. More transpositions than permutations can be generated given the same amount of run-time, which speed up the convergence of transposition procedure (Chung et al., 2019c). To further accelerate the rate of convergence and avoid possible bias, we introduce a full permutation to the sequence of transpositions. Figure 12 illustrates the convergence of transposition procedure.

In each simulation, we generated two groups each with 10 random modular networks. We then sequentially computed 200000 random transpositions while interjecting a random permutation for every 500 transpositions and obtained the p -value. This guarantees the convergence of p -value within 2 decimal places (within 0.01) in average. The simulations were independently performed 50 times and the average p -value across 50 simulations was reported.

Network difference. We compared two groups of networks generated by varying parameter $c = 2$ vs. 3, 2 vs. 6 and 3 vs. 6 each with $d = 12, 18, 24$ nodes and

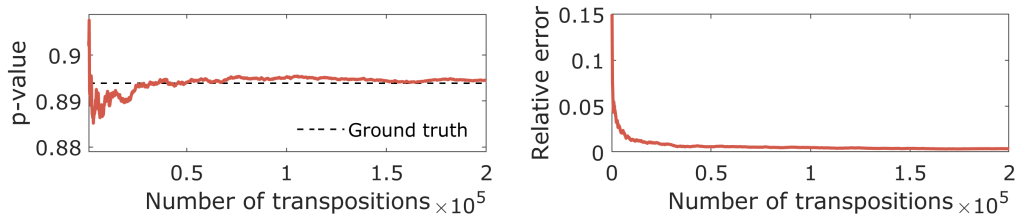


Figure 12: Transposition test is applied in determining the statistical significance of two groups each consisting of 10 simulated networks. To further speed up the convergence rate, a random permutation is intermixed for every sequence of 500 transpositions. The left panel shows the convergence of p -value over 50000 transpositions. For comparison, ground truth p -value is computed from the exact permutation test by enumerating every possible permutations exactly once. The left panel shows the average relative error against the ground truth across 100 independent simulations.

$p = 0.6, 0.8$ probability of connection within modules. Since each group exhibited different modular structure, topological loss and graph matching cost were expected to detect the group difference. Table 2 summarizes the performance results. Networks with $d = 12$ nodes might be too small to extract topologically distinct features used in each algorithm. Thus, all graph matching costs performed poorly while the topological loss performed reasonably well. When the number of nodes increases, all methods show overall performance improvement. In all settings, topological loss significantly outperforms other graph matching algorithms.

No network difference. We compared networks generated by unvarying parameter $c = 2$ vs. 2, 3 vs. 3 and 6 vs. 6 each with $d = 12, 18, 24$ nodes and $p = 0.6, 0.8$ probability of connection within modules. Since it was expected that there was no topological difference between networks generated using the same values for parameters c, d and p , topological loss and graph matching cost should not be able to detect the group difference. The performance result is summarized in Table 3. In all settings, all methods performed well when there was no group difference.

The baseline graph matching methods have low sensitivity over topological differences such as connected components and cycles in networks. They were unable to detect the network differences, especially subtle topological difference. While it might be possible to extend a graph matching algorithm to encode high-order geometrical relations, a small increment in the order of relations usually results in a combinatorial explosion of the amount of data needed to fit the model (Zhou and De la Torre, 2013). Thus, most of high-order graph matching

Table 2: Study 2. Network difference. The performance results are summarized as average p -values for various parameter settings of d (number of nodes), c (number of modules) and p (within-module probability).

d	c	p	GA	SM	RRWM	IPFP	\mathcal{L}_{top}
12 vs. 12	2 vs. 3	0.6	0.45 ± 0.27	0.48 ± 0.30	0.28 ± 0.31	0.34 ± 0.28	0.08 ± 0.16
		0.8	0.26 ± 0.24	0.30 ± 0.28	0.06 ± 0.12	0.28 ± 0.28	0.01 ± 0.03
	2 vs. 6	0.6	0.06 ± 0.10	0.17 ± 0.20	0.04 ± 0.13	0.23 ± 0.28	0.00 ± 0.00
		0.8	0.00 ± 0.01	0.01 ± 0.03	0.00 ± 0.00	0.02 ± 0.04	0.00 ± 0.00
	3 vs. 6	0.6	0.40 ± 0.29	0.35 ± 0.28	0.24 ± 0.26	0.35 ± 0.28	0.06 ± 0.13
		0.8	0.21 ± 0.23	0.28 ± 0.27	0.08 ± 0.14	0.26 ± 0.25	0.00 ± 0.01
	2 vs. 3	0.6	0.25 ± 0.23	0.41 ± 0.26	0.26 ± 0.24	0.42 ± 0.28	0.01 ± 0.02
		0.8	0.12 ± 0.17	0.19 ± 0.22	0.00 ± 0.00	0.04 ± 0.05	0.00 ± 0.00
18 vs. 18	2 vs. 6	0.6	0.02 ± 0.05	0.07 ± 0.17	0.00 ± 0.00	0.14 ± 0.20	0.00 ± 0.00
		0.8	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
	3 vs. 6	0.6	0.28 ± 0.24	0.37 ± 0.31	0.21 ± 0.24	0.37 ± 0.30	0.01 ± 0.01
		0.8	0.15 ± 0.22	0.13 ± 0.14	0.00 ± 0.01	0.16 ± 0.18	0.00 ± 0.00
24 vs. 24	2 vs. 3	0.6	0.23 ± 0.25	0.30 ± 0.26	0.14 ± 0.20	0.31 ± 0.28	0.00 ± 0.01
		0.8	0.06 ± 0.11	0.12 ± 0.19	0.00 ± 0.00	0.01 ± 0.05	0.00 ± 0.00
	2 vs. 6	0.6	0.00 ± 0.01	0.03 ± 0.06	0.00 ± 0.00	0.09 ± 0.13	0.00 ± 0.00
		0.8	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
	3 vs. 6	0.6	0.24 ± 0.26	0.29 ± 0.28	0.10 ± 0.13	0.37 ± 0.26	0.00 ± 0.00
		0.8	0.07 ± 0.12	0.13 ± 0.19	0.00 ± 0.01	0.12 ± 0.19	0.00 ± 0.00

methods are often limited to very sparse networks such as binary trees. They are not practical in dense functional brain networks with far larger number of cycles. On the other hand, the proposed topological loss is able to detect such subtle topological pattern differences with minimal amount of run-time.

4 Application

Dataset and preprocessing

The dataset is the resting-state fMRI of 412 subjects collected as part of the Human Connectome Project (HCP) twin study (Van Essen et al., 2012, 2013). The fMRI were collected over 14 minutes and 33 seconds using a gradient-echo-planar imaging (EPI) sequence with multiband factor 8, repetition time (TR) 720 ms, time echo (TE) 33.1 ms, flip angle 52° , 104×90 (RO \times PE) matrix size, 72 slices, 2 mm isotropic voxels, and 1200 time points. Subjects without fMRI or full 1200 time points were excluded. Additional details on the imaging protocol is given in <https://protocols.humanconnectome.org/HCP/3T/imaging-protocols.html>. During each scanning, participants were at rest with eyes open with relaxed fixation on a projected bright cross-hair on a dark back-

Table 3: Study 2. No network difference. The performance results are summarized as average p -values for various parameter settings of d (number of nodes), c (number of modules) and p (within-module probability).

d	c	p	GA	SM	RRWM	IPFP	\mathcal{L}_{top}
12 vs. 12	2 vs. 2	0.6	0.49 ± 0.27	0.46 ± 0.30	0.51 ± 0.30	0.47 ± 0.28	0.53 ± 0.29
		0.8	0.45 ± 0.25	0.47 ± 0.31	0.56 ± 0.29	0.47 ± 0.30	0.50 ± 0.30
	3 vs. 3	0.6	0.45 ± 0.32	0.44 ± 0.26	0.47 ± 0.27	0.51 ± 0.30	0.46 ± 0.31
		0.8	0.54 ± 0.31	0.51 ± 0.27	0.51 ± 0.29	0.52 ± 0.29	0.51 ± 0.30
	6 vs. 6	0.6	0.57 ± 0.30	0.51 ± 0.28	0.56 ± 0.29	0.45 ± 0.26	0.58 ± 0.29
		0.8	0.55 ± 0.29	0.48 ± 0.26	0.52 ± 0.27	0.54 ± 0.30	0.49 ± 0.27
18 vs. 18	2 vs. 2	0.6	0.48 ± 0.26	0.49 ± 0.32	0.54 ± 0.29	0.47 ± 0.30	0.54 ± 0.31
		0.8	0.52 ± 0.28	0.50 ± 0.28	0.46 ± 0.30	0.52 ± 0.25	0.50 ± 0.26
	3 vs. 3	0.6	0.49 ± 0.28	0.58 ± 0.31	0.43 ± 0.28	0.51 ± 0.27	0.53 ± 0.30
		0.8	0.46 ± 0.30	0.51 ± 0.27	0.52 ± 0.33	0.45 ± 0.29	0.53 ± 0.27
	6 vs. 6	0.6	0.53 ± 0.28	0.48 ± 0.30	0.51 ± 0.30	0.45 ± 0.29	0.44 ± 0.33
		0.8	0.54 ± 0.27	0.52 ± 0.30	0.48 ± 0.26	0.52 ± 0.31	0.43 ± 0.30
24 vs. 24	2 vs. 2	0.6	0.52 ± 0.28	0.49 ± 0.30	0.50 ± 0.30	0.48 ± 0.28	0.55 ± 0.26
		0.8	0.53 ± 0.27	0.56 ± 0.30	0.51 ± 0.30	0.56 ± 0.32	0.52 ± 0.30
	3 vs. 3	0.6	0.48 ± 0.29	0.54 ± 0.27	0.49 ± 0.26	0.49 ± 0.30	0.52 ± 0.30
		0.8	0.55 ± 0.29	0.49 ± 0.27	0.52 ± 0.28	0.49 ± 0.30	0.47 ± 0.26
	6 vs. 6	0.6	0.47 ± 0.30	0.45 ± 0.31	0.51 ± 0.29	0.56 ± 0.28	0.49 ± 0.29
		0.8	0.51 ± 0.30	0.47 ± 0.28	0.54 ± 0.28	0.56 ± 0.31	0.51 ± 0.31

ground (Van Essen et al., 2013). The standard minimal preprocessing pipelines (Glasser et al., 2013) were applied on the fMRI scans including spatial distortion removal (Andersson et al., 2003; Jovicich et al., 2006), motion correction (Jenkinson and Smith, 2001; Jenkinson et al., 2002), bias field reduction (Glasser and Van Essen, 2011), registration to the structural MNI template, and data masking using the brain mask obtained from FreeSurfer (Glasser et al., 2013). This resulted in the resting-state functional time series with $91 \times 109 \times 91$ 2-mm isotropic voxels at 1200 time points. The subject ranged from 22 to 36 years in age with average age 29.24 ± 3.39 years. There are 172 males and 240 females. Among them, there are 131 MZ-twin pairs and 75 same sex DZ-twin pairs.

Subsequently, we employed the Automated Anatomical Labeling (AAL) template to parcellate the brain volume into 116 non-overlapping anatomical regions (Tzourio-Mazoyer et al., 2002). We averaged fMRI across voxels within each brain parcellation, resulting in 116 average fMRI time series with 1200 time points for each subject. Previous studies reported that head movement produces spatial artifacts in functional connectivity (Power et al., 2012; Van Dijk et al., 2012; Satterthwaite et al., 2012; Caballero-Gaudes and Reynolds, 2017). Thus, we scrubbed the data to remove fMRI volumes with significant head motion (Power et al., 2012; Huang et al., 2020). We calculated the framewise displacement (FD)

from the three translational displacements and three rotational displacements at each time point to measure the head movement from one volume to the next. The volumes with FD larger than 0.5 mm and their neighbors (one back and two forward time points) were scrubbed (Van Dijk et al., 2012; Power et al., 2012; Huang et al., 2020). We excluded 12 subjects having excessive head movement, resulting in fMRI data of 400 subjects (168 males and 232 females). More than one third of 1200 volumes being scrubbed in the excluded 12 subjects. Among the remaining 400 subjects, there are 124 monozygotic (MZ) twin pairs and 70 same-sex dizygotic (DZ) twin pairs. The first 20 time points were removed from all subjects to avoid artifacts in the fMRI data, leaving 1180 time points per subject (Diedrichsen and Shadmehr, 2005; Shah et al., 2016).

For dMRI, the white matter fiber orientation information was extracted by multi-shell, multi-tissue constrained spherical deconvolution from different tissue types such as white matter and gray matter (Callaghan et al., 1988; Jeurissen et al., 2014). The fiber orientation distribution functions were estimated and apparent fiber densities were exploited to produce the reliable white and gray matter volume maps (Jeurissen et al., 2014; Christiaens et al., 2015). Subsequently, multiple random seeds were selected in each voxel to generate about 10 million initial streamlines per subject with the maximum fiber tract length at 250 mm and FA larger than 0.06 using MRtrix3 (<http://www.mrtrix.org>) (Tournier et al., 2012; Xie et al., 2018). The Spherical-Deconvolution Informed Filtering of Tractograms (SIFT2) technique making use of complete streamlines was subsequently applied to generate more biologically accurate brain connectivity, which results in about 1 million tracts per subject (Smith et al., 2015). Nonlinear diffeomorphic registration between subject images to the template was performed using ANTS (Avants et al., 2008, 2011). AAL was used to parcellate the brain into 116 regions (Tzourio-Mazoyer et al., 2002). The subject-level connectivity matrices were constructed by counting the number of tracts connecting between brain regions. The structural brain network P that serves as the template where all the functional networks are aligned is obtained by computing the one sample t -statistic map P over all the subjects and rescaling t -statistics between 0 to 2 using the hyperbolic tangent function \tanh , then adding 1.

Learning individual networks

Among 400 subjects, there are $p = 124$ monozygotic (MZ) twin pairs and $q = 70$ same-sex dizygotic (DZ) twin pairs. For subject k , we have resting-state fMRI time series $x = (x_1, x_2, \dots, x_{1180})$ for region i and $y = (y_1, y_2, \dots, y_{1180})$ for region j with 1180 time points. Correlation ρ_{ij}^k between regions i and j is computed as the Pearson correlation between x and y , as usually done in the field (Chung, 2013; Bryant et al., 2017; Shappell et al., 2019; Chung et al.,

2019b). This gives the correlation matrix $C_k = (\rho_{ij}^k)$, which is used as the baseline against the proposed method. We then translate and scale the correlation as $w_{ij}^k = \sqrt{(1 - \rho_{ij}^k)/2}$, which is a metric (Chung et al., 2019b). The subject-level functional brain network is given by $G_k = (V, w^k)$. The t -statistic map P is used as the template structural brain network where the functional network G_k is matched.

Given λ_k , we applied the topological learning to estimate the subject-level network $\Theta_k(\lambda_k)$ by minimizing the objective (1) using the individual network G_k and the structural network P :

$$\hat{\Theta}_k(\lambda_k) = \arg \min_{\Theta} \mathcal{L}_F(\Theta, G_k) + \lambda_k \mathcal{L}_{top}(\Theta, P).$$

Θ is initialized to G_k . For every subject, we globally used $\lambda = 1.0000$ where the average minimum is obtained (Figure 3).

Heritability in twins

We investigated in localizing what parts of brain networks are genetically heritable. In particular, we investigated if the estimated network $\hat{\Theta}_k$ is genetically heritable in twins as follows. At edge ij , let (a_{1l}, a_{2l}) be the l -th twin pair in MZ-twin and (b_{1l}, b_{2l}) be the l -th twin pair in DZ-twin. MZ-twin and DZ-twin pairs are then represented as

$$a = \begin{pmatrix} a_{11} & \cdots & a_{1p} \\ a_{21} & \cdots & a_{2p} \end{pmatrix}, \quad b = \begin{pmatrix} b_{11} & \cdots & b_{1q} \\ b_{21} & \cdots & b_{2q} \end{pmatrix}.$$

Let $a_r = (a_{r1}, a_{r2}, \dots, a_{rp})$ and $b_r = (b_{r1}, b_{r2}, \dots, b_{rq})$. Then MZ-correlation is computed as the Pearson correlation $\gamma^{MZ}(a_1, a_2)$ between a_1 and a_2 . Similarly for DZ-correlation $\gamma^{DZ}(b_1, b_2)$. In the widely-used ACE genetic model, the heritability index (HI) h , which determines the amount of variation caused by genetic factors in population, is estimated using Falconer’s formula (Falconer and Mackay, 1995; Chung et al., 2019b). Thus, HI h is given by $h(a, b) = 2(\gamma^{MZ} - \gamma^{DZ})$.

Since the order of the twins is interchangeable, we can *transpose* the l -th twin pair in MZ-twin as

$$\begin{aligned} \pi_l(a_1) &= (a_{11}, \dots, a_{1,l-1}, a_{2l}, a_{1,l+1}, \dots, a_{1p}), \\ \pi_l(a_2) &= (a_{21}, \dots, a_{2,l-1}, a_{1l}, a_{2,l+1}, \dots, a_{2p}) \end{aligned}$$

and obtain another MZ-correlation $\gamma^{MZ}(\pi_l(a_1), \pi_l(a_2))$. Likewise, we obtain many different correlations for DZ-twin. Similar to the transposition test used in the simulation study 2, we perform a sequence of random transpositions iteratively to estimate the twin correlations γ^{MZ} and γ^{DZ} sequentially as follows (Chung et al., 2019c).

Over the transposition π_l , the MZ-correlation is changed from $\gamma^{MZ}(a_1, a_2)$ to $\gamma^{MZ}(\pi_l(a_1), \pi_l(a_2))$, which involves the following functions:

$$\begin{aligned}\nu(a_r) &= \sum_{j=1}^p a_{rj}, \\ \omega(a_r, a_s) &= \sum_{j=1}^p (a_{rj} - \nu(a_r)/p)(a_{sj} - \nu(a_s)/p).\end{aligned}$$

The functions ν and ω are updated iteratively over the transposition π_l as

$$\begin{aligned}\nu(\pi_l(a_r)) &= \nu(a_r) - a_{rl} + a_{sl}, \\ \omega(\pi_l(a_r), \pi_l(a_s)) &= \omega(a_r, a_s) + (a_{rl} - a_{sl})^2/p - (a_{rl} - a_{sl})(\nu(a_r) - \nu(a_s))/p.\end{aligned}$$

Then MZ-correlation after the transposition is calculated by

$$\gamma^{MZ}(\pi_l(a_1), \pi_l(a_2)) = \frac{\omega(\pi_l(a_1), \pi_l(a_2))}{\sqrt{\omega(\pi_l(a_1), \pi_l(a_1))\omega(\pi_l(a_2), \pi_l(a_2))}}.$$

The time complexity for computing correlation iteratively is 33 operations per transposition, which is significantly more efficient than that of direct correlation computation per permutation. In numerical implementation, we sequentially perform random transpositions $\pi_{l_1}, \pi_{l_2}, \dots, \pi_{l_J}$, which result in J different twin correlations. Let

$$\kappa_1 = \pi_{l_1}, \quad \kappa_2 = \pi_{l_2} \circ \pi_{l_1}, \quad \dots, \quad \kappa_J = \pi_{l_J} \circ \dots \circ \pi_{l_2} \circ \pi_{l_1}.$$

The average MZ-correlation $\bar{\gamma}_J^{MZ}$ of J correlations is given by

$$\bar{\gamma}_J^{MZ} = \frac{1}{J} \sum_{j=1}^J \gamma^{MZ}(\kappa_j(a_1), \kappa_j(a_2)),$$

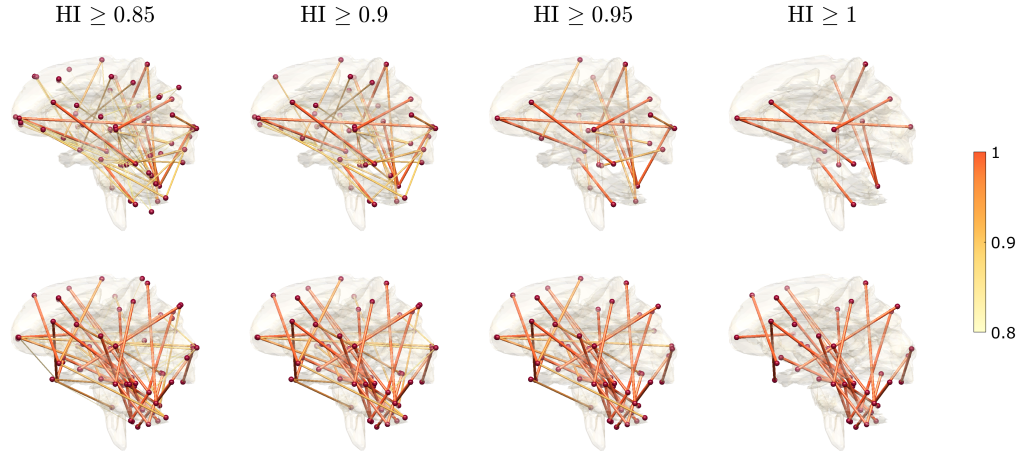
which is iteratively updated as

$$\bar{\gamma}_J^{MZ} = \frac{J-1}{J} \bar{\gamma}_{J-1}^{MZ} + \frac{1}{J} \gamma^{MZ}(\kappa_J(a_1), \kappa_J(a_2)).$$

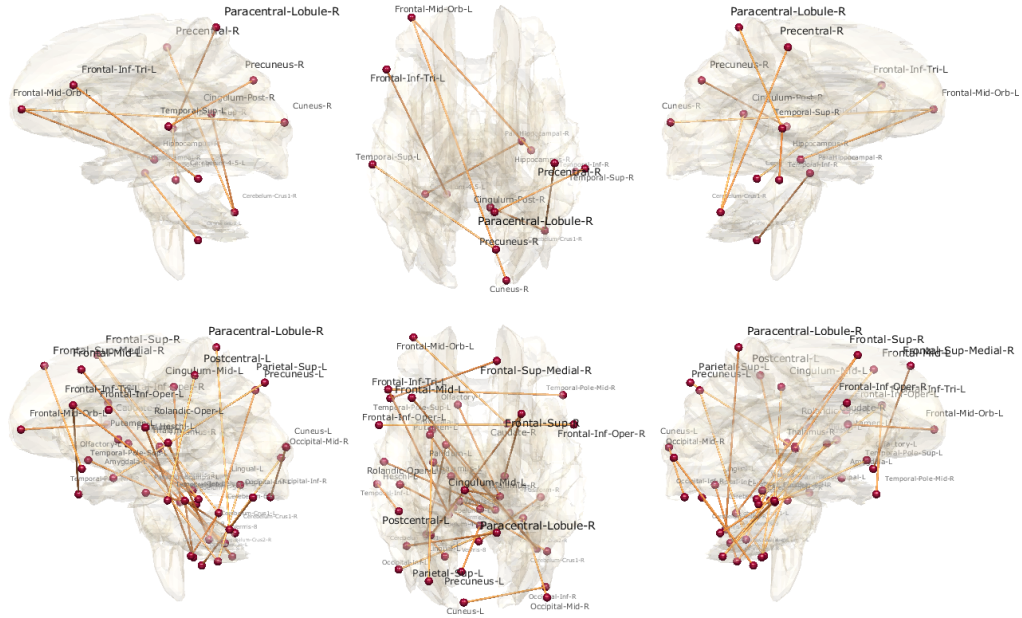
The average correlation $\bar{\gamma}_J^{MZ}$ converges to the true underlying twin correlation γ^{MZ} for sufficiently large J . Similarly, DZ-correlation γ^{DZ} is estimated.

Results

Using the transposition method, we randomly transposed a twin and updated the correlation for 50000 times. This process was repeated 100 times and the total



(a)



(b)

Figure 13: (a) The HI maps for the original Pearson correlation matrix (top) and topologically learned network (bottom) are thresholded at different $HI \geq 0.85, 0.9, 0.95$ and 1 . (b) Most highly heritable connections above $HI \geq 1$ using the original Pearson correlation matrix (top) and the topologically learned network (bottom).

Table 4: Top ten most heritable connections with $HI \geq 1$ using the topologically learned network.

Most heritable connections

Left superior parietal lobule – Left amygdala
Left lobule VIIIB of cerebellar hemisphere – Left globus pallidus
Left lobule III of cerebellar hemisphere – Right crus I of cerebellar hemisphere
Left – Right opercular part of inferior frontal gyrus
Left lobule IV, V of cerebellar hemisphere – Left thalamus
Left lobule IX of cerebellar hemisphere – Left lobule VI of cerebellar hemisphere
Right thalamus – Right superior frontal gyrus, dorsolateral
Left middle frontal gyrus, orbital part – Right caudate nucleus
Right crus II of cerebellar hemisphere – Left globus pallidus
Lobule VIII of vermis – Right fusiform gyrus

50000×100 correlations were used in estimating the underlying MZ- and DZ-correlations. At every edge, the standard deviation of the average correlations from 100 results was smaller than 0.01, which guarantees the convergence of the estimate within two decimal places in average.

We computed the HI-maps using the original correlation matrix C_k and the proposed topologically learned network $\hat{\Theta}_k$. Figures 13 displays far more connections with 100% heritability for the topologically learned network $\hat{\Theta}_k$ compared to the original Pearson correlation matrix C_k . As demonstrated in the simulation studies as well, the topological approach boosts subtle topological signals. The networks $\hat{\Theta}_k$ are expected to inherit topological sparsity from the template structural brain network P that has sparse topology with less cycles (Figure 5). This suggests that noisy, short-lived cycles were removed from the functional networks, improving the statistical sensitivity. Comparing HI-maps from the two methods, there are overlaps but the topological approach detects more connections with higher heritability. For network $\hat{\Theta}_k$, left superior parietal lobule and left amygdala connection shows the strongest heritability among many other connections (Table 4).

5 Conclusion

We have presented a novel topological learning framework that can analyze networks of different sizes and topology through a new loss function. Among many different learning tasks, the method is illustrated with averaging and regression. The method is briefly shown on how to average networks of different sizes and topology, which is not easy with existing methods. We have also applied the method to twin brain imaging data in analyzing functional and structural brain

networks together. Our topological learning framework is more sensitive on detecting subtle network genetic signals than the baseline method.

The topological loss can be easily adapted to other learning tasks such as clustering. It is well known that the Euclidean-distance based clustering such as k -means clustering does not perform well against more geometric clustering method such as spectral clustering (Ng et al., 2002; Kriegel et al., 2009). It may be possible that a new clustering method utilizing the proposed loss function might perform better than k -means or spectral clustering. This is left as a future study.

Acknowledgements

We thank Gary Shiu of University of Wisconsin-Madison for discussion on stability theorems. We thank Li Shen of University of Pennsylvania for providing the t -statistic map of structural brain networks used in this study. The t -statistic map came from study Chung et al. (2019c). We also like to thank Shih-Gu Huang of National University of Singapore for providing support for fMRI processing. This study is funded by NIH R01 EB022856, EB02875, NSF MDS-2010778.

References

- Adler, R., Bobrowski, O., Borman, M., Subag, E., Weinberger, S., 2010. Persistent homology for random fields and complexes. In: *Borrowing strength: theory powering applications—a Festschrift for Lawrence D. Brown*. Institute of Mathematical Statistics, pp. 124–143.
- Andersson, J. L., Skare, S., Ashburner, J., 2003. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *NeuroImage* 20 (2), 870–888.
- Arslan, S., Ktena, S., Makropoulos, A., Robinson, E., Rueckert, D., Parisot, S., 2018. Human brain mapping: A systematic comparison of parcellation methods for the human cerebral cortex. *NeuroImage* 170, 5–30.
- Avants, B., Epstein, C., Grossman, M., Gee, J., 2008. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis* 12, 26–41.
- Avants, B., Tustison, N., Song, G., Cook, P., Klein, A., Gee, J., 2011. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *NeuroImage* 54, 2033–2044.

- Babai, L., Luks, E., 1983. Canonical labeling of graphs. In: Proceedings of the fifteenth annual ACM symposium on Theory of computing. pp. 171–183.
- Bendich, P., Marron, J. S., Miller, E., Pieloch, A., Skwerer, S., 2016. Persistent homology analysis of brain artery trees. *The Annals of Applied Statistics* 10 (1), 198.
- Blokland, G., McMahon, K., Thompson, P., Martin, N., de Zubicaray, G., Wright, M., 2011. Heritability of working memory brain activation. *The Journal of Neuroscience* 31, 10882–10890.
- Bottou, L., 1998. Online learning and stochastic approximations. *On-Line Learning in Neural Networks* 17 (9), 142.
- Bryant, C., Zhu, H., Ahn, M., Ibrahim, J., 2017. LCN: a random graph mixture model for community detection in functional brain networks. *Statistics and Its Interface* 10 (3), 369.
- Bullmore, E., Sporns, O., 2009. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience* 10 (3), 186–198.
- Caballero-Gaudes, C., Reynolds, R., 2017. Methods for cleaning the BOLD fMRI signal. *NeuroImage* 154, 128–149.
- Callaghan, P., Eccles, C., Xia, Y., 1988. NMR microscopy of dynamic displacements: k-space and q-space imaging. *Journal of Physics E: Scientific Instruments* 21, 820.
- Carlsson, G., 2009. Topology and data. *Bulletin of the American Mathematical Society* 46 (2), 255–308.
- Carlsson, G., Mémoli, F., 2010. Characterization, stability and convergence of hierarchical clustering methods. *The Journal of Machine Learning Research* 11, 1425–1470.
- Chazal, F., Cohen-Steiner, D., Guibas, L. J., Mémoli, F., Oudot, S. Y., 2009. Gromov-Hausdorff stable signatures for shapes using persistence. In: *Computer Graphics Forum*. Vol. 28. Wiley Online Library, pp. 1393–1403.
- Chazal, F., Guibas, L. J., Oudot, S. Y., Skraba, P., 2013. Persistence-based clustering in riemannian manifolds. *Journal of the ACM (JACM)* 60 (6), 1–38.

- Chen, C., Ni, X., Bai, Q., Wang, Y., 2019. A topological regularizer for classifiers via persistent homology. In: The 22nd International Conference on Artificial Intelligence and Statistics. pp. 2573–2582.
- Chiang, M.-C., McMahon, K., de Zubicaray, G., Martin, N., Hickie, I., Toga, A., Wright, M., Thompson, P., 2011. Genetics of white matter development: a DTI study of 705 twins and their siblings aged 12 to 29. *NeuroImage* 54, 2308–2317.
- Cho, M., Lee, J., Lee, K. M., 2010. Reweighted random walks for graph matching. In: European Conference on Computer Vision. Springer, pp. 492–505.
- Christiaens, D., Reisert, M., Dhollander, T., Sunaert, S., Suetens, P., Maes, F., 2015. Global tractography of multi-shell diffusion-weighted imaging data using a multi-tissue model. *NeuroImage* 123, 89–101.
- Chung, M., Adluru, N., Dalton, K., Alexander, A., Davidson, R., 2011. Scalable brain network construction on white matter fibers. In: Proc. of SPIE. Vol. 7962. p. 79624G.
- Chung, M. K., 2013. Statistical and computational methods in brain image analysis. CRC Press.
- Chung, M. K., Hanson, J. L., Lee, H., Adluru, N., Alexander, A. L., Davidson, R. J., Pollak, S. D., 2013. Persistent homological sparse network approach to detecting white matter abnormality in maltreated children: MRI and DTI multimodal study. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 300–307.
- Chung, M. K., Huang, S.-G., Gritsenko, A., Shen, L., Lee, H., 2019a. Statistical inference on the number of cycles in brain networks. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). IEEE, pp. 113–116.
- Chung, M. K., Lee, H., DiChristofano, A., Ombao, H., Solo, V., 2019b. Exact topological inference of the resting-state brain networks in twins. *Network Neuroscience* 3 (3), 674–694.
- Chung, M. K., Lee, H., Solo, V., Davidson, R. J., Pollak, S. D., 2017a. Topological distances between brain networks. In: International Workshop on Connectomics in NeuroImaging. Springer, pp. 161–170.
- Chung, M. K., Villalta-Gil, V., Lee, H., Rathouz, P. J., Lahey, B. B., Zald, D. H., 2017b. Exact topological inference for paired brain networks via persistent

- homology. In: International Conference on Information Processing in Medical Imaging. Springer, pp. 299–310.
- Chung, M. K., Xie, L., Huang, S.-G., Wang, Y., Yan, J., Shen, L., 2019c. Rapid acceleration of the permutation test via transpositions. In: International Workshop on Connectomics in NeuroImaging. Springer, pp. 42–53.
- Clough, J. R., Oksuz, I., Byrne, N., Schnabel, J. A., King, A. P., 2019. Explicit topological priors for deep-learning based image segmentation using persistent homology. In: International Conference on Information Processing in Medical Imaging. Springer, pp. 16–28.
- Cohen-Steiner, D., Edelsbrunner, H., Harer, J., 2007. Stability of persistence diagrams. *Discrete & Computational Geometry* 37 (1), 103–120.
- Cohen-Steiner, D., Edelsbrunner, H., Harer, J., Mileyko, Y., 2010. Lipschitz functions have L_p -stable persistence. *Foundations of Computational Mathematics* 10 (2), 127–139.
- Cour, T., Srinivasan, P., Shi, J., 2006. Balanced graph matching. *Advances in Neural Information Processing Systems* 19, 313–320.
- Crawford, L., Monod, A., Chen, A. X., Mukherjee, S., Rabadán, R., 2020. Predicting clinical outcomes in glioblastoma: an application of topological and functional data analysis. *Journal of the American Statistical Association* 115 (531), 1139–1150.
- Desikan, R., Ségonne, F., Fischl, B., Quinn, B., Dickerson, B., Blacker, D., Buckner, R., Dale, A., Maguire, R., Hyman, B., Marilyn, S., Ronald, J., 2006. An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *NeuroImage* 31, 968–980.
- Diedrichsen, J., Shadmehr, R., 2005. Detecting and adjusting for artifacts in fMRI time series data. *NeuroImage* 27, 624–634.
- Edelsbrunner, H., Harer, J., 2008. Persistent homology - a survey. *Contemporary Mathematics* 453, 257–282.
- Edelsbrunner, H., Letscher, D., Zomorodian, A., 2000. Topological persistence and simplification. In: *Proceedings 41st Annual Symposium on Foundations of Computer Science. IEEE*, pp. 454–463.
- Edmonds, J., Karp, R. M., 1972. Theoretical improvements in algorithmic efficiency for network flow problems. *Journal of the ACM (JACM)* 19 (2), 248–264.

- Eickhoff, S., Yeo, B., Genon, S., 2018. Imaging-based parcellations of the human brain. *Nature Reviews Neuroscience* 19, 672–686.
- Falconer, D., Mackay, T., 1995. *Introduction to Quantitative Genetics*, 4th ed. Longman.
- Fan, L., Li, H., Zhuo, J., Zhang, Y., Wang, J., Chen, L., Yang, Z., Chu, C., Xie, S., Laird, A., Fox, P., Eickhoff, S., Yu, C., Jiang, T., 2016. The human brain-netome atlas: a new brain atlas based on connectonal architecture. *Cerebral Cortex* 26, 3508–3526.
- Fornito, A., Zalesky, A., Bullmore, E., 2010. Network scaling effects in graph analytic studies of human resting-state fMRI data. *Frontiers in Systems Neuroscience* 4, 1–16.
- Fornito, A., Zalesky, A., Bullmore, E., 2016. *Fundamentals of Brain Network Analysis*. Academic Press, New York.
- Ghrist, R., 2008. Barcodes: the persistent topology of data. *Bulletin of the American Mathematical Society* 45 (1), 61–75.
- Ginestet, C. E., Nichols, T. E., Bullmore, E. T., Simmons, A., 2011. Brain network analysis: separating cost from topology using cost-integration. *PLoS One* 6 (7), e21570.
- Glahn, D., Winkler, A., Kochunov, P., Almasy, L., Duggirala, R., Carless, M., Curran, J., Olvera, R., Laird, A., Smith, S., Beckmann, C., Fox, P., Blangero, J., 2010. Genetic control over the resting brain. *Proceedings of the National Academy of Sciences* 107, 1223–1228.
- Glasser, M., Smith, S., Marcus, D., Andersson, J., Auerbach, E., Behrens, T., Coalson, T., Harms, M., Jenkinson, M., Moeller, S., 2016. The human connectome project’s neuroimaging approach. *Nature Neuroscience* 19, 1175.
- Glasser, M., Van Essen, D., 2011. Mapping human cortical areas in vivo based on myelin content as revealed by T1-and T2-weighted MRI. *Journal of Neuroscience* 31, 11597–11616.
- Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J. R., et al., 2013. The minimal preprocessing pipelines for the human connectome project. *NeuroImage* 80, 105–124.

- Gold, S., Rangarajan, A., 1996. A graduated assignment algorithm for graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 18 (4), 377–388.
- Gong, G., He, Y., Concha, L., Lebel, C., Gross, D., Evans, A., Beaulieu, C., 2009. Mapping anatomical connectivity patterns of human cerebral cortex using in vivo diffusion tensor imaging tractography. *Cerebral Cortex* 19, 524–536.
- Gritsenko, A., Lindquist, M., Chung, M., 2020. Twin classification in resting-state brain connectivity. *IEEE International Symposium on Biomedical Imaging (ISBI)*, arXiv:1807.00244.
- Guo, X., Srivastava, A., 2020. Representations, metrics and statistics for shape analysis of elastic graphs. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. pp. 832–833.
- Hagmann, P., Kurant, M., Gigandet, X., Thiran, P., Wedeen, V., Meuli, R., Thiran, J., 2007. Mapping human whole-brain structural networks with diffusion MRI. *PLoS One* 2 (7), e597.
- Honey, C. J., Kötter, R., Breakspear, M., Sporns, O., 2007. Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *Proceedings of the National Academy of Sciences* 104 (24), 10240–10245.
- Honey, C. J., Sporns, O., Cammoun, L., Gigandet, X., Thiran, J.-P., Meuli, R., Hagmann, P., 2009. Predicting human resting-state functional connectivity from structural connectivity. *Proceedings of the National Academy of Sciences* 106 (6), 2035–2040.
- Hu, X., Li, F., Samaras, D., Chen, C., 2019. Topology-preserving deep image segmentation. In: *Advances in Neural Information Processing Systems*. pp. 5657–5668.
- Huang, S.-G., Samdin, S.-T., Ting, C., Ombao, H., Chung, M., 2020. Statistical model for dynamically-changing correlation matrices with application to brain connectivity. *Journal of Neuroscience Methods* 331, 108480.
- Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage* 17 (2), 825–841.
- Jenkinson, M., Smith, S., 2001. A global optimisation method for robust affine registration of brain images. *Medical Image Analysis* 5 (2), 143–156.

- Jeurissen, B., Tournier, J.-D., Dhollander, T., Connelly, A., Sijbers, J., 2014. Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion MRI data. *NeuroImage* 103, 411–426.
- Johnston, D., Wu, S. M.-S., 1994. *Foundations of cellular neurophysiology*. MIT press.
- Jovicich, J., Czanner, S., Greve, D., Haley, E., van Der Kouwe, A., Gollub, R., Kennedy, D., Schmitt, F., Brown, G., MacFall, J., et al., 2006. Reliability in multi-site structural mri studies: effects of gradient non-linearity correction on phantom and human data. *NeuroImage* 30 (2), 436–443.
- Kang, H., Ombao, H., Fonnesbeck, C., Ding, Z., Morgan, V., 2017. A Bayesian double fusion model for resting-state brain connectivity using joint functional and structural data. *Brain connectivity* 7, 219–227.
- Karas, M., Brzyski, D., Dziedzic, M., Goñi, J., Kareken, D. A., Randolph, T. W., Harezlak, J., 2019. Brain connectivity-informed regularization methods for regression. *Statistics in Biosciences* 11 (1), 47–90.
- Kerber, M., Morozov, D., Nigmatov, A., 2017. Geometry helps to compare persistence diagrams. *Journal of Experimental Algorithmics (JEA)* 22, 1–20.
- Kolouri, S., Nadjahi, K., Simsekli, U., Badeau, R., Rohde, G., 2019. Generalized sliced wasserstein distances. In: *Advances in Neural Information Processing Systems*. pp. 261–272.
- Kriegel, H.-P., Kröger, P., Zimek, A., 2009. Clustering high-dimensional data: A survey on subspace clustering, pattern-based clustering, and correlation clustering. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 3 (1), 1–58.
- Le, H., Kume, A., 2000. The fréchet mean shape and the shape of the means. *Advances in Applied Probability*, 101–113.
- Lee, H., Chung, M., Kang, H., Kim, B.-N., Lee, D., 2011. Discriminative persistent homology of brain networks. In: *IEEE International Symposium on Biomedical Imaging (ISBI)*. pp. 841–844.
- Lee, H., Kang, H., Chung, M. K., Kim, B.-N., Lee, D. S., 2012. Persistent brain network homology from the perspective of dendrogram. *IEEE Transactions on Medical Imaging* 31 (12), 2267–2277.

- Leordeanu, M., Hebert, M., 2005. A spectral technique for correspondence problems using pairwise constraints. In: Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1. Vol. 2. IEEE, pp. 1482–1489.
- 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. pp. 1114–1122.
- Lv, J., Guo, L., Hu, X., Zhang, T., Li, K., Zhang, D., Yang, J., Liu, T., 2010. Fiber-centered analysis of brain connectivities using DTI and resting state fMRI data. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). Springer, pp. 143–150.
- McKay, D., Knowles, E., Winkler, A., Sprooten, E., Kochunov, P., Olvera, R., Curran, J., Kent Jr., J., Carless, M., Göring, H., Dyer, T., Duggirala, R., Almasy, L., Fox, P., Blangero, J., Glahn, D., 2014. Influence of age, sex and genetic factors on the human brain. *Brain Imaging and Behavior* 8, 143–152.
- Ng, A., Jordan, M., Weiss, Y., 2002. On spectral clustering: analysis and an algorithm. In: Advances in Neural Information Processing Systems. pp. 849–856.
- Ombao, H., Lindquist, M., Thompson, W., Aston, J., 2016. *Handbook of Neuroimaging Data Analysis*. CRC Press.
- Power, J., Barnes, K., Snyder, A., Schlaggar, B., Petersen, S., 2012. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage* 59, 2142–2154.
- Rabin, J., Peyré, G., Delon, J., Bernot, M., 2011. Wasserstein barycenter and its application to texture mixing. In: International Conference on Scale Space and Variational Methods in Computer Vision. Springer, pp. 435–446.
- Reininghaus, J., Huber, S., Bauer, U., Kwitt, R., 2015. A stable multi-scale kernel for topological machine learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4741–4748.
- Reynolds, C., Phillips, D., 2015. Genetics of brain aging–twin aging.
- Satterthwaite, T., Wolf, D., Loughhead, J., Ruparel, K., Elliott, M., Hakonarson, H., Gur, R., Gur, R., 2012. Impact of in-scanner head motion on multiple measures of functional connectivity: relevance for studies of neurodevelopment in youth. *NeuroImage* 60, 623–632.

- Schaefer, A., Kong, R., Gordon, E., Laumann, T., Zuo, X.-N., Holmes, A., Eickhoff, S., Yeo, B., 2017. Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cerebral Cortex* 28, 3095–3114.
- Shah, L., Cramer, J., Ferguson, M., Birn, R., Anderson, J., 2016. Reliability and reproducibility of individual differences in functional connectivity acquired during task and resting state. *Brain and Behavior* 6, e00456.
- Shappell, H., Caffo, B., Pekar, J., Lindquist, M., 2019. Improved state change estimation in dynamic functional connectivity using hidden semi-markov models. *bioRxiv*, 519868.
- Shattuck, D., Mirza, M., Adisetiyo, V., Hojatkashani, C., Salamon, G., Narr, K., Poldrack, R., Bilder, R., Toga, A., 2008. Construction of a 3D probabilistic atlas of human cortical structures. *NeuroImage* 39, 1064–1080.
- Singh, N., Couture, H. D., Marron, J., Perou, C., Niethammer, M., 2014. Topological descriptors of histology images. In: *International Workshop on Machine Learning in Medical Imaging*. Springer, pp. 231–239.
- Smit, D., Stam, C., Posthuma, D., Boomsma, D., De Geus, E., 2008. Heritability of small-world networks in the brain: a graph theoretical analysis of resting-state EEG functional connectivity. *Human Brain Mapping* 29, 1368–1378.
- Smith, R., Tournier, J.-D., Calamante, F., Connelly, A., 2015. SIFT2: enabling dense quantitative assessment of brain white matter connectivity using streamlines tractography. *NeuroImage* 119, 338–351.
- Snijders, T., Sreen, M., Zwaagstra, R., 1995. The use of multilevel modeling for analysing personal networks: Networks of cocaine users in an urban area. *Journal of Quantitative Anthropology* 5 (2), 85–105.
- Songdechakraiut, T., Chung, M. K., 2020. Dynamic topological data analysis for functional brain signals. In: *2020 IEEE 17th International Symposium on Biomedical Imaging Workshops (ISBI Workshops)*. IEEE, pp. 1–4.
- Tian, Y., Yan, J., Zhang, H., Zhang, Y., Yang, X., Zha, H., 2012. On the convergence of graph matching: Graduated assignment revisited. In: *European Conference on Computer Vision*. Springer, pp. 821–835.
- Tournier, J., Calamante, F., Connelly, A., et al., 2012. Mrtrix: diffusion tractography in crossing fiber regions. *International Journal of Imaging Systems and Technology* 22, 53–66.

- Turner, K., Mileyko, Y., Mukherjee, S., Harer, J., 2014. Fréchet means for distributions of persistence diagrams. *Discrete & Computational Geometry* 52, 44–70.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage* 15, 273–289.
- Van Dijk, K., Sabuncu, M., Buckner, R., 2012. The influence of head motion on intrinsic functional connectivity MRI. *NeuroImage* 59, 431–438.
- Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., Ugurbil, K., Consortium, W.-M. H., et al., 2013. The WU-Minn human connectome project: an overview. *NeuroImage* 80, 62–79.
- Van Essen, D. C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T., Bucholz, R., Chang, A., Chen, L., Corbetta, M., Curtiss, S. W., et al., 2012. The human connectome project: a data acquisition perspective. *NeuroImage* 62 (4), 2222–2231.
- Wang, T., Liu, H., Li, Y., Jin, Y., Hou, X., Ling, H., 2020. Learning combinatorial solver for graph matching. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 7568–7577.
- Wang, Y., Ombao, H., Chung, M. K., 2018. Topological data analysis of single-trial electroencephalographic signals. *The Annals of Applied Statistics* 12 (3), 1506.
- Wasserman, L., 2018. Topological data analysis. *Annual Review of Statistics and Its Application* 5, 501–532.
- Xie, L., Amico, E., Salama, P., Wu, Y.-C., Fang, S., Sporns, O., Saykin, A., Goñi, J., Yan, J., Shen, L., 2018. Heritability estimation of reliable connectomic features. In: *International Workshop on Connectomics in Neuroimaging, Lecture Notes in Computer Science*. Vol. 11083. pp. 58–66.
- Xue, W., Bowman, F. D., Pileggi, A. V., Mayer, A. R., 2015. A multimodal approach for determining brain networks by jointly modeling functional and structural connectivity. *Frontiers in Computational Neuroscience* 9, 22.
- Yu, T., Yan, J., Wang, Y., Liu, W., et al., 2018. Generalizing graph matching beyond quadratic assignment model. In: *Advances in Neural Information Processing Systems*. pp. 853–863.

- Zalesky, A., Fornito, A., Harding, I., Cocchi, L., Yücel, M., Pantelis, C., Bullmore, E., 2010. Whole-brain anatomical networks: Does the choice of nodes matter? *NeuroImage* 50, 970–983.
- Zavlanos, M., Pappas, G., 2008. A dynamical systems approach to weighted graph matching. *Automatica* 44, 2817–2824.
- Zemel, Y., Panaretos, V., 2019. Fréchet means and procrustes analysis in Wasserstein space. *Bernoulli* 25, 932–976.
- Zhang, Z., Descoteaux, M., Zhang, J., Girard, G., Chamberland, M., Dunson, D., Srivastava, A., Zhu, H., 2018. Mapping population-based structural connectomes. *NeuroImage* 172, 130–145.
- Zhang, Z., Xiang, Y., Wu, L., Xue, B., Nehorai, A., 2019. KerGM: Kernelized graph matching. In: *Advances in Neural Information Processing Systems*. pp. 3335–3346.
- Zhou, F., De la Torre, F., 2013. Deformable graph matching. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 2922–2929.
- Zhu, D., Zhang, T., Jiang, X., Hu, X., Chen, H., Yang, N., Lv, J., Han, J., Guo, L., Liu, T., 2014. Fusing DTI and fMRI data: a survey of methods and applications. *NeuroImage* 102, 184–191.