Topological Learning and Its Application to Multimodal Brain Network Integration

Tananun Songdechakraiwut[1(✉)], Li Shen[2], and Moo Chung[1]

[1] University of Wisconsin–Madison, Madison, USA
songdechakra@wisc.edu
[2] University of Pennsylvania, Philadelphia, USA

Abstract. A long-standing challenge in multimodal brain network analyses is to integrate topologically different brain networks obtained from diffusion and functional MRI in a coherent statistical framework. Existing multimodal frameworks will inevitably destroy the topological difference of the networks. In this paper, we propose a novel topological learning framework that integrates networks of different topology through persistent homology. Such challenging task is made possible through the introduction of a new topological loss that bypasses intrinsic computational bottlenecks and thus enables us to perform various topological computations and optimizations with ease. We validate the topological loss in extensive statistical simulations with ground truth to assess its effectiveness of discriminating networks. Among many possible applications, we demonstrate the versatility of topological loss in the twin imaging study where we determine the extent to which brain networks are genetically heritable.

Keywords: Topological data analysis · Persistent homology · Wasserstein distance · Multimodal brain networks · Twin brain imaging study

1 Introduction

In standard brain network modeling, the whole brain is usually parcellated into a few hundred disjoint regions [7,17,27]. For instance, well established, widely used Automated Anatomical Labeling (AAL) parcellates the brain into 116 regions [27]. These disjoint regions form nodes in a brain network. Subsequently, functional or structural information is overlaid on top of the parcellation to obtain brain connectivity between the regions. Structural connectivity is obtained from diffusion MRI (dMRI), which traces the white matter fibers in the brain. Strength of the structural connectivity is determined by the number of fibers connecting the parcellations. Resting-state functional connectivity
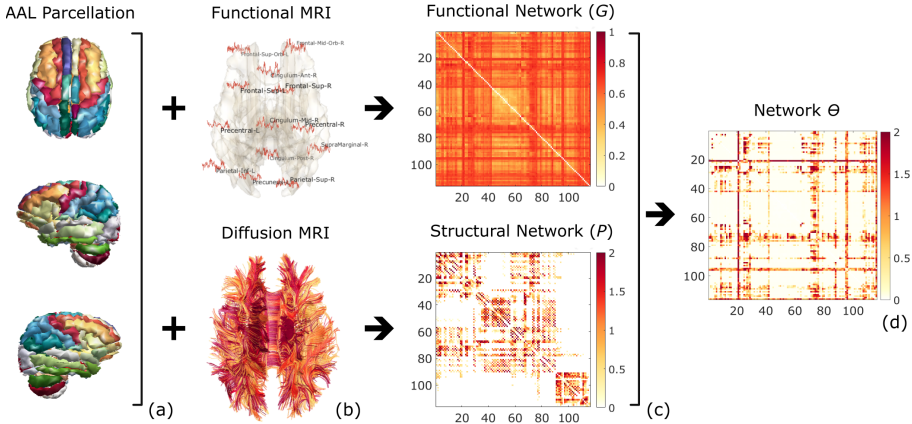

Fig. 1. Schematic of topological learning. (a) AAL partitions the human brain into 116 disjoint regions. (b, c) Functional network G is obtained from resting-state fMRI. The template structural network P is obtained from dMRI. The structural network P is sparse while the functional network G is densely connected with many cycles. (d) We learn network Θ that has the topological characteristics of both functional and structural networks.

obtained from functional MRI (fMRI) is often computed using the Pearson correlation coefficient between average fMRI time series in the parcellations [7]. While the structural connectivity provides information whether the brain regions are physically connected through the white matter fibers, the functional connectivity can exhibit relations between two regions without a direct neuroanatomical connection [14]. Thus, functional brain networks are often very dense with thousands of loops or cycles [7] while structural brain networks are expected to exhibit sparse topology without many cycles. Both the structural and functional brain networks provide topologically different information (Fig. 1). Nonetheless, not much research has been done thus far on integrating the brain networks at the localized connection level. Existing integration frameworks will inevitably destroy the topological difference in the process [16,30]. There is a need for a new multimodal brain network model that can integrate networks of different topology in a coherent statistical framework.

Persistent homology [7,9,15,17,25] provides a novel approach to the long-standing challenge in multimodal brain network analyses. In persistent homology, topological features such as connected components and cycles are measured across different spatial resolutions represented in the form of barcodes. It was recently proposed to penalize the barcodes as a loss function in image segmentation [15]. Though the method allows to incorporate topological information into the problem, it is limited to an image with a handful of topological features due to its expensive optimization process with $O(n^3)$. This is impractical in brain networks with a far larger number of topological features comprising hundreds of connected components and thousands of cycles. In this paper, we propose a more
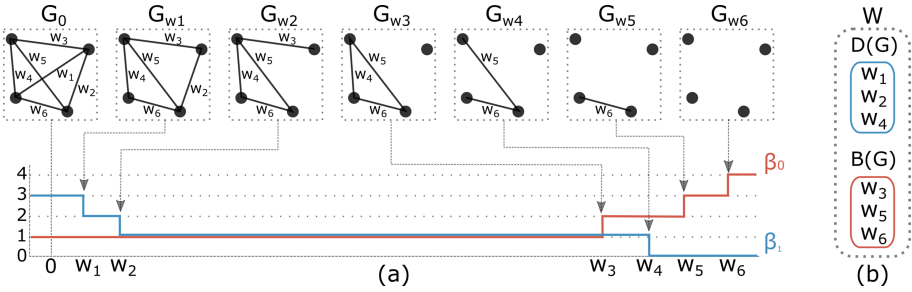
Fig. 2. (a) Graph filtration of G. β_0 is monotonically increasing while β_1 is monotonically decreasing over the graph filtration. Connected components are born at edge weights w_3, w_5, w_6 while cycles die at edge weights w_1, w_2, w_4. 0D barcode is represented by a set of birth values $B(G) = \{w_3, w_5, w_6\}$. 1D barcode is represented by a set of death values $D(G) = \{w_1, w_2, w_4\}$. (b) The weight set $W = \{w_1, ..., w_6\}$ is partitioned into 0D birth values and 1D death values: $W = B(G) \cup D(G)$.

principled and scalable *topological loss* with $O(n \log n)$. Our proposed method bypasses the intrinsic computational bottleneck and thus enables us to perform various topological computations and optimizations with ease.

Twin studies on brain imaging phenotypes provide a well established way to examine the extent to which brain networks are influenced by genetic factors. However, previous twin imaging studies have not been well adapted beyond determining heritability of a few brain regions of interest [2,4,12,21,24]. Measures of network topology are worth investigating as intermediate phenotypes that indicate the genetic risk for various neuropsychiatric disorders [3]. Determining heritability of the whole brain network is the first necessary prerequisite for identifying network based endophenotypes. With our topological loss, we propose a novel *topological learning* framework where we determine heritability of the functional brain networks while integrating the structural brain network information (Fig. 1). Our method increases statistical sensitivity to subtle topological differences, yielding more connections as genetic signals.

2   Method

Barcodes in Graph Filtration. Consider a network represented as a weighted graph $G = (V, w)$ comprising a node set V and edge weights $w = (w_{ij})$ with positive and unique weights. The number of nodes and edges are denoted by $|V|$ and $|E|$. Network G is a complete graph with $|E| = |V|(|V| - 1)/2$. The binary graph $G_\epsilon = (V, w_\epsilon)$ of G is defined as a graph consisting of the node set V and binary edge weight $w_{\epsilon,ij} = 1$ if $w_{ij} > \epsilon$ or 0 otherwise. The binary network G_ϵ is the 1-skeleton, a simplicial complex consisting of nodes and edges only [22]. In 1-skeleton, 0-dimensional (0D) holes are connected components and 1-dimensional (1D) holes are cycles [7]. The number of connected components and cycles in the binary network G_ϵ are referred to as the 0-th Betti number $\beta_0(G_\epsilon)$ and the 1-st Betti number $\beta_1(G_\epsilon)$. A *graph filtration* of G is defined as a collection of

nested binary networks [7,17]: $G_{\epsilon_0} \supset G_{\epsilon_1} \supset \cdots \supset G_{\epsilon_k}$, where $\epsilon_0 < \epsilon_1 < \cdots < \epsilon_k$ are filtration values. By increasing the filtration value ϵ, we are thresholding at higher connectivity resulting in more edges being removed, and thus the 0-th and 1-st Betti numbers change.

Persistent homology keeps track of appearances (birth) and disappearances (death) of connected components and cycles over filtration values ϵ, and associates their *persistence* (lifetimes measured as the duration of birth to death) to them. Long lifetimes indicate global topological features while short lifetimes indicate small-scale topological features [11,20,29]. The persistence is represented by *0D* and *1D barcodes* comprising a set of intervals $[b_i, d_i]$, each of which tabulates a lifetime of a connected component or a cycle that appears at the filtration value b_i and vanishes at d_i (Fig. 2). Since connected components are born one at a time over increasing filtration values [7], these connected components will never die once they are born. Thus, we simply ignore their death values at ∞ and represent 0D barcode as a set of only birth values $B(G) = \cup_i \{b_i\}$. Cycles are considered born at $-\infty$ and will die one at a time over the filtration. Ignoring the $-\infty$, we represent 1D barcode as a set of only death values $D(G) = \cup_i \{d_i\}$.

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

The proof is given in the supplementary material. Finding 0D birth values $B(G)$ is *equivalent* to finding edge weights of the *maximum spanning tree* (MST) of G using Prim's or Kruskal's algorithm [17]. Once B is computed, D is simply given as the remaining edge weights. Thus, the barcodes are computed efficiently in $O(|E| \log |V|)$.

Topological Loss. Since networks are topologically completely characterized by 0D and 1D barcodes, the topological dissimilarity between two networks can be measured through barcode differences. We adapt the Wasserstein distance to quantify the differences between the barcodes [9,15,23]. The Wasserstein distance measures the differences between underlying probability distributions on barcodes through the Dirac delta function [10]. Let $\Theta = (V, w^\Theta)$ and $P = (V, w^P)$ be two given networks. The *topological loss* $\mathcal{L}_{top}(\Theta, P)$ is defined as the optimal matching cost

$$\mathcal{L}_{top}(\Theta, P) = \min_{\tau_0} \sum_{b \in B(\Theta)} \left[b - \tau_0(b) \right]^2 + \min_{\tau_1} \sum_{d \in D(\Theta)} \left[d - \tau_1(d) \right]^2, \qquad (1)$$

where τ_0 is a bijection from $B(\Theta)$ to $B(P)$ and τ_1 is a bijection from $D(\Theta)$ to $D(P)$. By Theorem 1, the bijections τ_0 and τ_1 always exist. The first term measures how close two networks are in terms of 0D holes (connected components) and is referred to as 0D topological loss $\mathcal{L}_{0D}$. The second term measures how close two networks are in terms of 1D holes (cycles) and is called 1D topological

loss $\mathcal{L}_{1D}$. Connected components represent an integration of a brain network while cycles represent how strong the integration is [6]. The optimization can be done exactly as follows:

Theorem 2

$$\mathcal{L}_{0D} = \min_{\tau_0} \sum_{b \in B(\Theta)} \left[b - \tau_0(b) \right]^2 = \sum_{b \in B(\Theta)} \left[b - \tau_0^*(b) \right]^2, \tag{2}$$

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

$$\mathcal{L}_{1D} = \min_{\tau_1} \sum_{d \in D(\Theta)} \left[d - \tau_1(d) \right]^2 = \sum_{d \in D(\Theta)} \left[d - \tau_1^*(d) \right]^2, \tag{3}$$

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

The proof is given in the supplementary material. We can compute the optimal matchings τ_0^* and τ_1^* between Θ and P in $O\big(|B(\Theta)| \log |B(\Theta)|\big)$ and $O\big(|D(\Theta)| \log |D(\Theta)|\big)$ by sorting edge weights and matching them.

Topological Learning. Let $G_1 = (V, w^1), \cdots, G_n = (V, w^n)$ be observed networks used for training a model. Let $P = (V, w^P)$ be a network expressing a prior topological knowledge. In brain network analysis, G_k can be a functional brain network of k-th subject obtained from resting-state fMRI, and P can be a template structural brain network obtained from dMRI. The functional networks can then overlay the template network (Fig. 1).

We are interested in learning the model $\Theta = (V, w^\Theta)$ using both the functional and structural brain networks. At the subject level, we train Θ using individual network G_k by optimizing

$$\widehat{\Theta}_k = \arg\min_{\Theta} \mathcal{L}_F(\Theta, G_k) + \lambda \mathcal{L}_{top}(\Theta, P), \tag{4}$$

where the squared Frobenius loss $\mathcal{L}_F(\Theta, G_k) = ||w^\Theta - w^k||_F^2$ measures the goodness of fit between the model and the individual network. The parameter λ controls the amount of topological information of network P that is introduced to G_k. The larger the value of λ, the more we are learning toward the topology of P. If $\lambda = 0$, we no longer learn the topology of P but simply fit the model Θ to the individual network G_k.

In numerical implementation, $\Theta = (V, w^\Theta)$ can be estimated iteratively through gradient descent efficiently by Theorems 1 and 2. The *topological gradient* with respect to edge weights $w^\Theta = (w_{ij}^\Theta)$ is given as

$$\frac{\partial \mathcal{L}_{top}(\Theta, P)}{\partial w_{ij}^\Theta} = \begin{cases} 2\big(w_{ij}^\Theta - \tau_0^*(w_{ij}^\Theta)\big) & \text{if } w_{ij}^\Theta \in B(\Theta); \\ 2\big(w_{ij}^\Theta - \tau_1^*(w_{ij}^\Theta)\big) & \text{if } w_{ij}^\Theta \in D(\Theta). \end{cases} \tag{5}$$
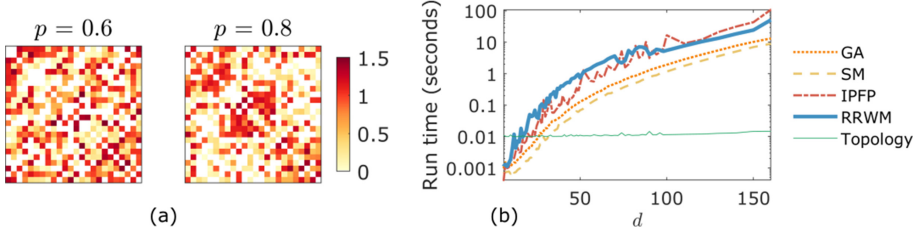
Fig. 3. (a) Two modular networks with $d = 24$ nodes and $c = 3$ modules were generated using $p = 0.6$ and 0.8. (b) The run time of graph matching cost between two modular networks of node size d plotted in the logarithmic scale. The run time of topological loss grows in a minuscule rate with the node size as opposed to the exponential run times of the graph matching algorithms.

By updating the edge weight w_{ij}^{Θ}, we adjust either a 0D birth value or a 1D death value, which changes topology of the model Θ. At each current iteration, we take a step in the direction of negative gradient with respect to an updated Θ from the previous iteration. As w_{ij}^{Θ} is moved through its optimal matching, the topology of Θ gets close to that of P while the Frobenius norm keeps Θ close to the observed network G_k. The time complexity of topological gradient is dominated by the computation of the MST with $O(|E| \log |V|)$.

3   Statistical Validation

We evaluated discriminative performance of the topological loss against four well-known graph matching algorithms: graduated assignment (GA) [13], spectral matching (SM) [18], integer projected fixed point method (IPFP) [19] and re-weighted random walk matching (RRWM) [5] using simulated networks.

We simulated random modular network $\mathcal{X}$ with d number of nodes and c number of modules where the nodes are evenly distributed among modules. We used $d = 12, 18, 24$ and $c = 2, 3, 6$. 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 Gaussian noise $\mathcal{N}(0, \sigma^2)$ with probability $1 - p$. Edge weights connecting nodes between *different* modules had probability $1 - p$ of being $\mathcal{N}(\mu, \sigma^2)$ and probability p of being $\mathcal{N}(0, \sigma^2)$. Any negative edge weights were set to zero. With larger value of within-module probability p, we have more pronounced modular structure (Fig. 3-a). The network $\mathcal{X}$ exhibits topological structures of connectedness. $\mu = 1$, $\sigma = 0.25$ and $p = 0.6$ were universally used as network variability.

We simulated two groups of random modular networks $\mathcal{X}_1, \cdots, \mathcal{X}_m$ and $\mathcal{Y}_1, \cdots, \mathcal{Y}_n$. If there is group difference in network topology, an average topological loss within group $\overline{\mathcal{L}}_W = \frac{\sum_{i<j} \mathcal{L}(\mathcal{X}_i, \mathcal{X}_j) + \sum_{i<j} \mathcal{L}(\mathcal{Y}_i, \mathcal{Y}_j)}{\binom{m}{2} + \binom{n}{2}}$ is expected to be smaller than an average topological loss between groups $\overline{\mathcal{L}}_B = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} \mathcal{L}(\mathcal{X}_i, \mathcal{Y}_j)}{mn}$. We measured the group disparity as the ratio statistic $\phi_{\mathcal{L}} = \overline{\mathcal{L}}_B \big/ \overline{\mathcal{L}}_W$. If $\phi_{\mathcal{L}}$ is large, the groups differ significantly in network topology. If $\phi_{\mathcal{L}}$ is small, it is likely

Table 1. Performance results are summarized as average p-values for various parameter settings of d (number of nodes) and c (number of modules).

d	c	GA	SM	RRWM	IPFP	$\mathcal{L}_{top}$
12 vs. 12	2 vs. 3	0.45 ± 0.27	0.48 ± 0.30	0.28 ± 0.31	0.34 ± 0.28	$\mathbf{0.08 \pm 0.16}$
	3 vs. 6	0.40 ± 0.29	0.35 ± 0.28	0.24 ± 0.26	0.35 ± 0.28	$\mathbf{0.06 \pm 0.13}$
18 vs. 18	2 vs. 3	0.25 ± 0.23	0.41 ± 0.26	0.26 ± 0.24	0.42 ± 0.28	$\mathbf{0.01 \pm 0.02}$
	3 vs. 6	0.28 ± 0.24	0.37 ± 0.31	0.21 ± 0.24	0.37 ± 0.30	$\mathbf{0.01 \pm 0.01}$
24 vs. 24	2 vs. 3	0.23 ± 0.25	0.30 ± 0.26	0.14 ± 0.20	0.31 ± 0.28	$\mathbf{0.00 \pm 0.01}$
	3 vs. 6	0.24 ± 0.26	0.29 ± 0.28	0.10 ± 0.13	0.37 ± 0.26	$\mathbf{0.00 \pm 0.00}$
12 vs. 12	2 vs. 2	0.49 ± 0.27	0.46 ± 0.30	0.51 ± 0.30	0.47 ± 0.28	0.53 ± 0.29
	3 vs. 3	0.45 ± 0.32	0.44 ± 0.26	0.47 ± 0.27	0.51 ± 0.30	0.46 ± 0.31
	6 vs. 6	0.57 ± 0.30	0.51 ± 0.28	0.56 ± 0.29	0.45 ± 0.26	0.58 ± 0.29

that there is no group difference. Similarly, we also defined the ratio statistic for the baseline algorithms. Since the distributions of the ratio statistics were unknown, the permutation test was used. In each simulation, we generated two groups with 10 modular networks each. We then computed 200000 permutations by shuffling group labels and obtained the p-values. The simulations were independently performed 50 times and the average p-value was reported.

The baseline graph matching algorithms are of polynomial time and not scalable compared to our method. For networks with $d = 100$ nodes, the run times of all the baselines are more than 100 times longer than that of topological loss (Fig. 3-b). When there is *network difference* (first three rows in Table 1), small p-value indicates that a method performs well at discriminating networks. In all the parameter settings, topological loss outperformed the other graph matching algorithms. Topological loss also consistently outperformed the baseline algorithms for other values of c, d and p. In the case of *no network difference* (last row in Table 1), small p-value indicates a method falsely detects the network difference when there is none. Since p-values of all the methods were not statistically significant, they all performed well. We also get similar results for other values of c, d and p. The graph matching algorithms are unable to detect topological differences while topological loss is able to easily detect such differences in subtle topological patterns with the minimal amount of run time. The MATLAB code for the simulation study is available at https://topolearn.github.io/topo-loss. The SM algorithm used in this simulation study and methods proposed in [1,26] rely on the same spectral graph theory and are expected to show analogous performance results.

4   Application to a Twin Imaging Study

Dataset and Preprocessing. dMRI and resting-state fMRI data were obtained from the Human Connectome Project [28]. fMRI went through further preprocessing including motion correction, scrubbing, bandpass filtering and
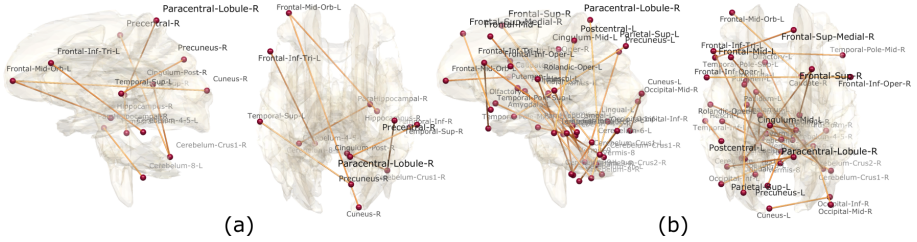
Fig. 4. Most heritable connections with 100% heritability using (a) Pearson correlation matrices and (b) topologically learned networks.

outlier removal among others. AAL was used to parcellate the brain into 116 regions [27]. fMRI were spatially averaged across voxels within each brain region resulting in 116 average fMRI time series per subject. There are 124 monozygotic (MZ) twin pairs and 70 same-sex dizygotic (DZ) twin pairs. For dMRI, about one million fiber tracts per subject were generated to compute biologically accurate brain connectivity [8]. AAL was used to parcellate the brain into 116 regions. The subject-level connectivity matrices were constructed by counting the number of tracts between the regions. The template structural network P was obtained by computing one sample t-statistic map over all the subjects and rescaling the t-statistic between 0 to 2 through the hyperbolic tangent function then adding 1 (Fig. 1). The t-statistic map from [8] is made publicly available at http://stat.wisc.edu/~mchung/softwares/dti.

Genetic Heritability. For the k-th subject, functional connectivity ρ_{ij}^k between regions i and j was computed using the Pearson correlation between time series. We converted the correlation into a metric through $w^k = (w_{ij}^k)$, where $w_{ij}^k = \sqrt{(1 - \rho_{ij}^k)/2}$ and obtained a subject-level functional brain network $G_k = (V, w^k)$ [7]. The topological learning was applied to estimate the subject-level model Θ_k by minimizing the objective function (4) using the individual network G_k and the template structural network P. The model Θ_k was initialized to G_k. To determine an optimal subject-level λ, we searched over different λ's to find a value that minimized the total loss $\mathcal{L}_F + \mathcal{L}_{top}$ for each subject. The average of the optimal λ's across all the subjects was $\lambda = 1.0000 \pm 0.0002$, a highly stable result. Thus, we globally used $\lambda = 1$ for all the subjects. We then investigated if the learned networks $\widehat{\Theta}_k$ are genetically heritable. We used the ACE model where the heritability index (HI) is estimated using Falconer's formula [6].

Results and Discussion. We computed the HI using the initial Pearson correlation matrices as a baseline versus the topologically learned networks. Figure 4, which displays resulting HI thresholded at 100% heritability, shows far more connections for the learned networks as opposed to the Pearson correlation matrices. The learned networks are expected to inherit sparse topology without many

cycles from the template network P (Fig. 1). This suggests that short-lived cycles were removed from the initial functional networks, improving the statistical sensitivity. For the learned networks, the connection with the highest HI is between left superior parietal lobule and left amygdala among many other connections with 100% heritability, suggesting that genes influence the development of these connections. Our findings can be used as a baseline for studying more complex relations between brain networks and other phenotypes.

References

1. Becker, C.O., et al.: Spectral mapping of brain functional connectivity from diffusion imaging. Sci. Rep. **8**(1), 1–15 (2018)
2. Blokland, G., McMahon, K., Thompson, P., Martin, N., de Zubicaray, G., Wright, M.: Heritability of working memory brain activation. J. Neurosci. **31**, 10882–10890 (2011)
3. Bullmore, E., Sporns, O.: Complex brain networks: graph theoretical analysis of structural and functional systems. Nat. Rev. Neurosci. **10**(3), 186–198 (2009)
4. Chiang, M.C., et al.: Genetics of white matter development: a DTI study of 705 twins and their siblings aged 12 to 29. NeuroImage **54**, 2308–2317 (2011)
5. Cho, M., Lee, J., Lee, K.M.: Reweighted random walks for graph matching. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010. LNCS, vol. 6315, pp. 492–505. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15555-0_36
6. Chung, M.K., Huang, S.G., Gritsenko, A., Shen, L., Lee, H.: Statistical inference on the number of cycles in brain networks. In: 16th International Symposium on Biomedical Imaging, pp. 113–116. IEEE (2019)
7. Chung, M.K., Lee, H., DiChristofano, A., Ombao, H., Solo, V.: Exact topological inference of the resting-state brain networks in twins. Network Neurosci. **3**, 674 (2019)
8. Chung, M.K., Xie, L., Huang, S.-G., Wang, Y., Yan, J., Shen, L.: Rapid acceleration of the permutation test via transpositions. In: Schirmer, M.D., Venkataraman, A., Rekik, I., Kim, M., Chung, A.W. (eds.) CNI 2019. LNCS, vol. 11848, pp. 42–53. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32391-2_5
9. Clough, J.R., Oksuz, I., Byrne, N., Schnabel, J.A., King, A.P.: Explicit topological priors for deep-learning based image segmentation using persistent homology. In: Chung, A.C.S., Gee, J.C., Yushkevich, P.A., Bao, S. (eds.) IPMI 2019. LNCS, vol. 11492, pp. 16–28. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-20351-1_2
10. Cohen-Steiner, D., Edelsbrunner, H., Harer, J., Mileyko, Y.: Lipschitz functions have Lp-stable persistence. Found. Comput. Math. **10**, 127 (2010). https://doi.org/10.1007/s10208-010-9060-6

11. Ghrist, R.: Barcodes: the persistent topology of data. Bull. Am. Math. Soc. **45**(1), 61–75 (2008)
12. Glahn, D., et al.: Genetic control over the resting brain. Proc. Nat. Acad. Sci. **107**, 1223–1228 (2010)
13. Gold, S., Rangarajan, A.: A graduated assignment algorithm for graph matching. IEEE Trans. Pattern Anal. Mach. Intell. **18**(4), 377–388 (1996)
14. Honey, C.J., Kötter, R., Breakspear, M., Sporns, O.: Network structure of cerebral cortex shapes functional connectivity on multiple time scales. Proc. Nat. Acad. Sci. **104**(24), 10240–10245 (2007)
15. Hu, X., Li, F., Samaras, D., Chen, C.: Topology-preserving deep image segmentation. In: Advances in Neural Information Processing Systems (2019)
16. Kang, H., Ombao, H., Fonnesbeck, C., Ding, Z., Morgan, V.L.: A Bayesian double fusion model for resting-state brain connectivity using joint functional and structural data. Brain Connectivity **7**(4), 219–227 (2017)
17. Lee, H., Kang, H., Chung, M.K., Kim, B.N., Lee, D.S.: Persistent brain network homology from the perspective of dendrogram. IEEE Trans. Med. Imag. **31**(12), 2267–2277 (2012)
18. Leordeanu, M., Hebert, M.: A spectral technique for correspondence problems using pairwise constraints. In: International Conference on Computer Vision, pp. 1482–1489. IEEE (2005)
19. Leordeanu, M., Hebert, M., Sukthankar, R.: An integer projected fixed point method for graph matching and map inference. In: Advances in Neural Information Processing Systems, pp. 1114–1122 (2009)
20. Marchese, A., Maroulas, V.: Signal classification with a point process distance on the space of persistence diagrams. Adv. Data Anal. Classification **12**(3), 657–682 (2017). https://doi.org/10.1007/s11634-017-0294-x
21. McKay, D., et al.: Influence of age, sex and genetic factors on the human brain. Brain Imag. Behav. **8**(2), 143–152 (2013). https://doi.org/10.1007/s11682-013-9277-5
22. Munkres, J.R.: Elements of Algebraic Topology. CRC Press, Boca Raton (2018)
23. Rabin, J., Peyré, G., Delon, J., Bernot, M.: Wasserstein barycenter and its application to texture mixing. In: Bruckstein, A.M., ter Haar Romeny, B.M., Bronstein, A.M., Bronstein, M.M. (eds.) SSVM 2011. LNCS, vol. 6667, pp. 435–446. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-24785-9_37
24. Smit, D., Stam, C., Posthuma, D., Boomsma, D., De Geus, E.: Heritability of small-world networks in the brain: a graph theoretical analysis of resting-state EEG functional connectivity. Hum. Brain Map. **29**, 1368–1378 (2008)
25. Songdechakraiwut, T., Chung, M.K.: Dynamic topological data analysis for functional brain signals. In: 17th International Symposium on Biomedical Imaging Workshops, pp. 1–4. IEEE (2020)
26. Surampudi, S.G., Naik, S., Surampudi, R.B., Jirsa, V.K., Sharma, A., Roy, D.: Multiple kernel learning model for relating structural and functional connectivity in the brain. Sci. Rep. **8**(1), 1–14 (2018)
27. Tzourio-Mazoyer, N., et al.: Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. NeuroImage **15**, 273–289 (2002)
28. Van Essen, D.C., et al.: The human connectome project: a data acquisition perspective. NeuroImage **62**(4), 2222–2231 (2012)

29. Xia, K., Wei, G.W.: Persistent homology analysis of protein structure, flexibility, and folding. Int. J. Numerical Methods Biomed. Eng. **30**(8), 814–844 (2014)
30. Xue, W., Bowman, F.D., Pileggi, A.V., Mayer, A.R.: A multimodal approach for determining brain networks by jointly modeling functional and structural connectivity. Front. Comput. Neurosci. **9**, 22 (2015)

Supplementary material for the paper "Topological Learning and Its Application to Multimodal Brain Network Integration"

A Proof: Theorem 1

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

Proof. Let $G = (V, w)$ be a network comprising a node set V and edge weights $w = (w_{ij})$ with positive and unique weights. The number of nodes and edges are denoted as $|V|$ and $|E|$. We have $|E| = |V|(|V| - 1)/2$. During the graph filtration, the deletion of every edge w_{ij} must result in either a birth of a connected component or a death of a cycle. The birth of a component and the death of a cycle cannot possibly happen simultaneously. Thus, every edge weight must be in either the set of birth values $B(G)$ or the set of death values $D(G)$. The binary graph G_0 is a complete graph consisting of a single connected component, while G_∞ is a node set. During the filtration, β_0 is monotonically increasing from $\beta_0(G_0) = 1$ to $\beta_0(G_\infty) = |V|$. There are $\beta_0(G_\infty) - \beta_0(G_0) = |V| - 1$ number of connected components that are born over the filtration. Out of $|V|(|V| - 1)/2$ total number of edge weights in the complete graph, there are $|V| - 1$ weights corresponding to the birth values of connected components. Thus, there are $|V|(|V| - 1)/2 - (|V| - 1) = 1 + \frac{|V|(|V|-3)}{2}$ number of cycles. ∎

B Proof: Theorem 2

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

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

we consider finding $\min_\tau \sum_{i=1}^n (a_i - \tau(a_i))^2$ over all possible bijections τ.

Proof. The minimization is equivalent to maximizing the assignment cost $\sum_{i=1}^n a_i \tau(a_i)$. We then 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.