Topological Distances between Brain Networks

Moo K. Chung[1], Hyekyoung Lee[2], Victor Solo[3], Richard J. Davidson[1], Seth D. Pollak[1]
[1]*University of Wisconsin-Madison* [2]*Seoul National University,* [3]*University of New South Wales*
email:mkchung@wisc.edu

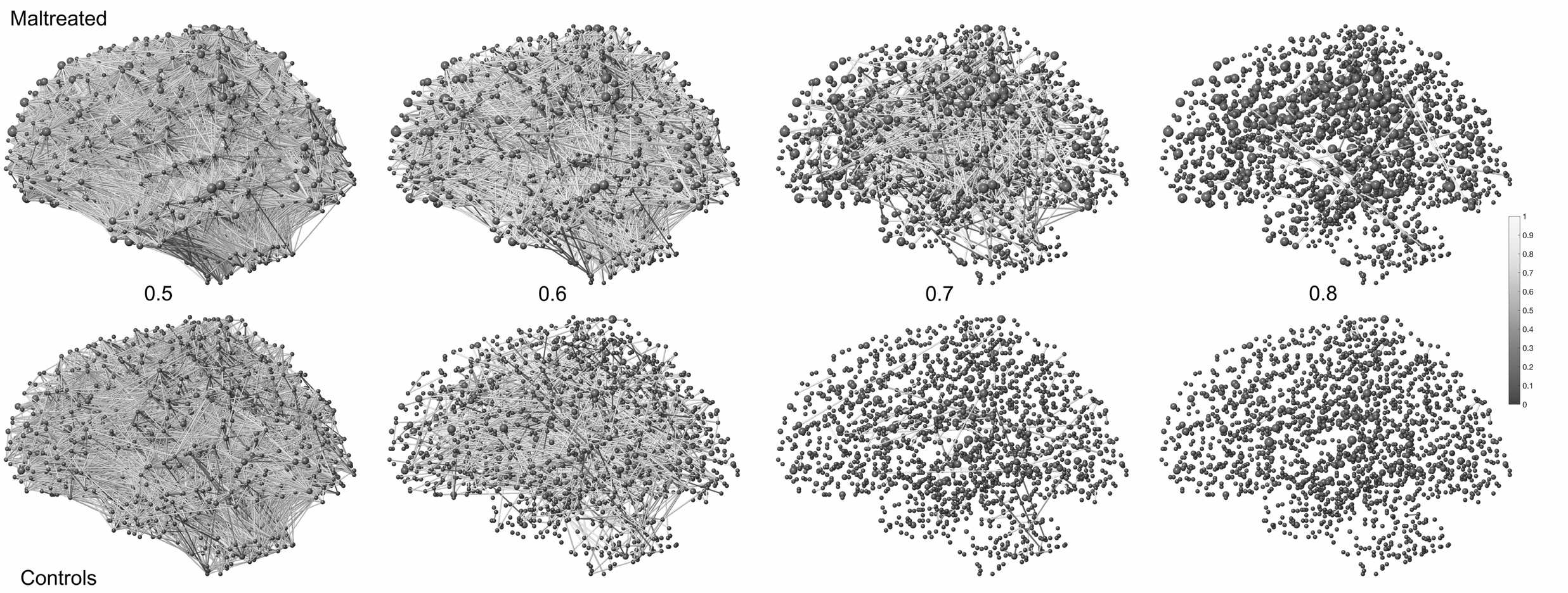

Figure 1. Graph filtration over correlations of FA-values on 1856 nodes. The maltreated children (top) has denser homogeneous network compared to normal controls (bottom). The size of nodes corresponds to the sum of correlations over edges.

Introduction

Many existing brain network distances are based on matrix norms. The element-wise differences may fail to capture underlying topological differences. Further, matrix norms are sensitive to outliers. A few extreme edge weights may severely affect the distance. There is a need to develop network distances that recognize topology.

We introduce Gromov-Hausdorff (GH) and Kolmogorov-Smirnov (KS) distances. GH-distance is often used in persistent homology based brain network models. The superior performance of KS-distance is contrasted against matrix norms and GH-distance in simulations with the ground truths. The KS-distance is then applied in characterizing the multimodal MRI and DTI study of maltreated children.

Matrix norms

Consider a weighted graph with node set $V = \{1,\dots,p\}$ and edge weights w_{ij} between nodes i and j. The measurement vector $\mathbf{x}_i = (x_{1i},\cdots,x_{ni})^\top \in \mathbb{R}^n$ is given at node i. The Pearson correlation between $\mathbf{x}_i$ and $\mathbf{x}_j$ is denoted as $\text{corr}(\mathbf{x}_i,\mathbf{x}_j)$. For weights $w_{ij} = \sqrt{1-\text{corr}(\mathbf{x}_i,\mathbf{x}_j)}$, it can be shown that $\mathcal{X} = (V,w_{ij})$ forms a metric space.

Given two networks $\mathcal{X}^1 = (V,w_{ij}^1)$ and $\mathcal{X}^2 = (V,w_{ij}^2)$, the L_l-norm of network difference is given by

$$D_l(\mathcal{X}^1,\mathcal{X}^2) = \Big(\sum_{i,j}|w_{ij}^1 - w_{ij}^2|^l\Big)^{1/l}.$$

When $l = \infty$, L_∞-distance is written as

$$D_\infty(\mathcal{X}^1,\mathcal{X}^2) = \max_{\forall i,j}|w_{ij}^1 - w_{ij}^2|.$$

The element-wise differences may not capture additional higher order similarity. Also L_1 and L_2-distances usually surfer the problem of outliers. Few outlying extreme edge weights may severely affect the distance. Further, these distances ignore the underlying topological structures. Thus, there is a need to define distances that are more topological.

Gromov-Hausdorff distance

GH-distance for brain networks was first used in Lee et al. (2012). The distance s_{ij} between the closest nodes in the two disjoint connected components in a graph is the single linkage distance (SLD). GH-distance between networks is then defined through GH-distance between corresponding dendrograms. Given two dendrograms $\mathcal{D}^1 = (V,s_{ij}^1)$ and $\mathcal{D}^2 = (V,s_{ij}^2)$ with single linkage distances s_{ij}^1 and s_{ij}^2,

$$D_{GH}(\mathcal{D}^1,\mathcal{D}^2) = \frac{1}{2}\max_{\forall i,j}|s_{ij}^1 - s_{ij}^2|.$$

Kolmogorov-Smirnov distance

Given network $\mathcal{X} = (V,w_{ij})$, its binary network $\mathcal{B}_\epsilon(\mathcal{X}) = (V,\mathcal{B}_\epsilon(w_{ij}))$ is a graph with edge weights $\mathcal{B}_\epsilon(w_{ij}) = 1$ if $w_{ij} \le \epsilon$ and 0 otherwise. It can be shown that

$$\mathcal{B}_{\epsilon_0}(\mathcal{X}) \subset \mathcal{B}_{\epsilon_1}(\mathcal{X}) \subset \mathcal{B}_{\epsilon_2}(\mathcal{X}) \subset \cdots$$

for $0 = \epsilon_0 \le \epsilon_1 \le \epsilon_2 \cdots$. The sequence of such nested graphs is called the *graph filtration* (Fig. 1) and $\epsilon_0,\epsilon_1,\epsilon_2\cdots$ are called the *filtration values* (Lee et al., 2011, 2012; Chung et al., 2017). The graph filtration can be quantified using monotonic function f satisfying

$$f \circ \mathcal{B}_{\epsilon_j}(\mathcal{X}) \ge f \circ \mathcal{B}_{\epsilon_{j+1}}(\mathcal{X})$$

for $\epsilon_j \le \epsilon_{j+1}$.

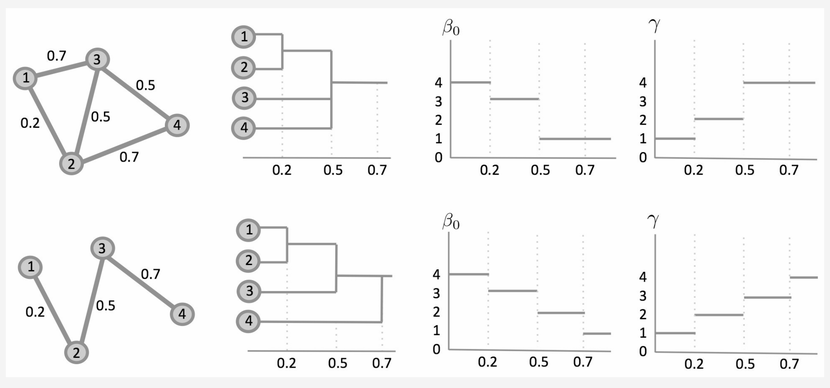

Figure 2. Toy networks, its dendrgrams, The number of connected components (β_0) and the size of the largest cluster (γ) are plotted over filtration values. Uncorrected edge should be treated as having ∞ weights.

The number of connected components, the zeroth Betti number β_0, satisfies the monotonicity property (1). The size of the largest cluster, denoted as γ, satisfies a similar but opposite relation of monotonic increase (Fig. 2).

Kolmogorov-Smirnov (KS) distance between $\mathcal{X}^1$ and $\mathcal{X}^2$ is defined as (Chung et al., 2015, 2017)

$$D_{KS}(\mathcal{X}^1,\mathcal{X}^2) = \sup_{1\le j\le q}|f\circ\mathcal{B}_{\epsilon_j}(\mathcal{X}^1) - f\circ\mathcal{B}_{\epsilon_j}(\mathcal{X}^2)|.$$

The probability distribution of D_{KS} under the null hypothesis of no network difference is given by

$$\lim_{q\to\infty}\left(D_q/\sqrt{2q} \ge d\right) = 2\sum_{i=1}^{\infty}(-1)^{i-1}e^{-2i^2d^2}.$$

Comparisons

The simulations were performed 100 times and the average results were reported. The sample size was $n = 5$ in each group and the number of nodes was $p = 100$ (Fig. 3). The noise level $\sigma = 0.01$ was used. The results did change even if we increased the noise level to $\sigma = 0.1$. $\sqrt{1-\text{corr}}$ was used as edge weights.

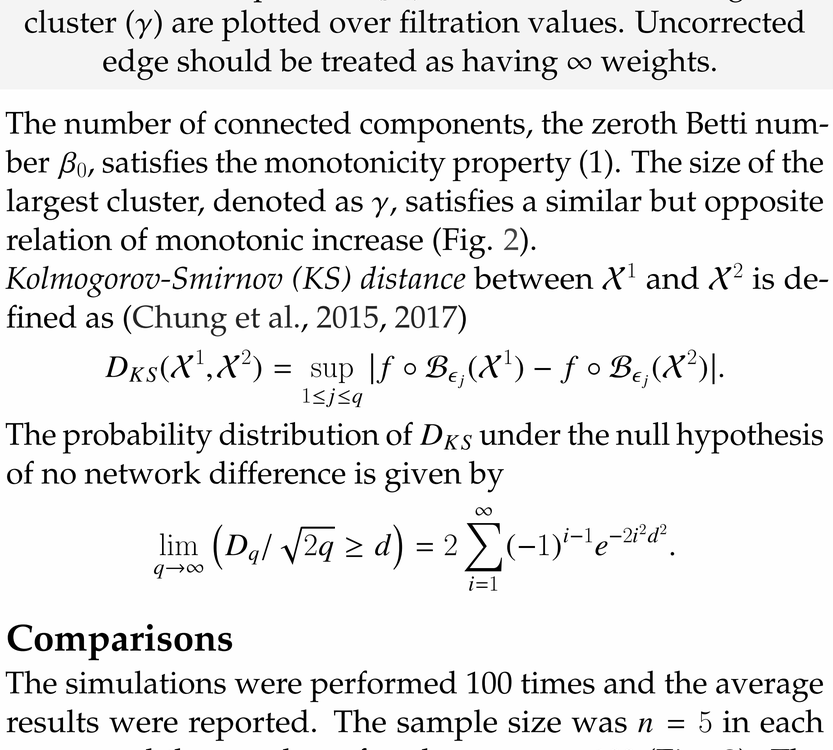

Figure 3. Randomly simulated correlation matrices with modular structures.

Group I. The measurement vector $\mathbf{x}_i$ was simulated as multivariate normal $N(0,I)$.
Group II. It was simulated as $\mathbf{y}_i = \mathbf{x}_i + N(0,\sigma^2 I_n)$.
Group III. It was simulated as $\mathbf{y}_i = 0.5\mathbf{x}_{ci+1} + N(0,\sigma^2 I_n)$. This introduce modules in the network. Each module consists of $c = p/k$ number of points ($k = 4,5,10$).
Group IV. The measurement vector $\mathbf{y}_i$ was simulated by adding noise to Group III: $\mathbf{z}_i = \mathbf{y}_i + N(0,\sigma^2 I_n)$.

No network difference. It was expected there was no network difference between Groups I and II (0 vs. 0) and III and IV (4 vs. 4). All the distances performed equally well and did not detect differences (Table 1).

Network difference. We compared 4 and 5 module networks, and 5 and 10 module networks. KS-distances performed extremely well compared with other distances.

Computation. The KS-distance method took about 20 seconds while all other distance methods took about 16 minutes. The code is available: http://www.stat.wisc.edu/~mchung/twins.

Modules	L_1	L_2	L_∞	GH	KS (β_0)	KS (γ)
0 vs. 0	0.93	0.93	0.93	0.87	1.00	1.00
4 vs. 4	0.89	0.89	0.90	0.86	0.87	0.88
4 vs. 5	0.14	0.06	0.03	0.29	0.07^{**}	0.07^{**}
5 vs. 10	0.47	0.19	0.10	0.33	0.01	0.06^{*}

Table 1. Simulation results given in terms of p-values. In the case of no network differences (0 vs. 0 and 4 vs. 4), higher p-values are better. In the case of network differences (4 vs. 5 and 5 vs. 10), smaller p-values are better. * and ** indicates multiplying 10^{-3} and 10^{-4}.

Application

The methods were applied to multimodal MRI and DTI of 31 normal controls and 23 age-matched children who experienced maltreatment while living in post-institutional settings before being adopted by families in US. Ages range from 9 to 14 years. The average amount of time spend in institutional care was 2.5 ± 1.4 years. Children were on average 3.2 years when they were adapted. The detailed descrption of the study is in Chung et al. (2015).

For MRI, the Jacobian determinants of warping from the template to individual subjects were obtained. For DTI, fractional anisotropy (FA) were calculated for diffusion tensor volumes diffeomorphically registered to the study specific template. Jacobian determinants and FA-values are uniformly sampled at 1856 nodes along the white mater template boundary.

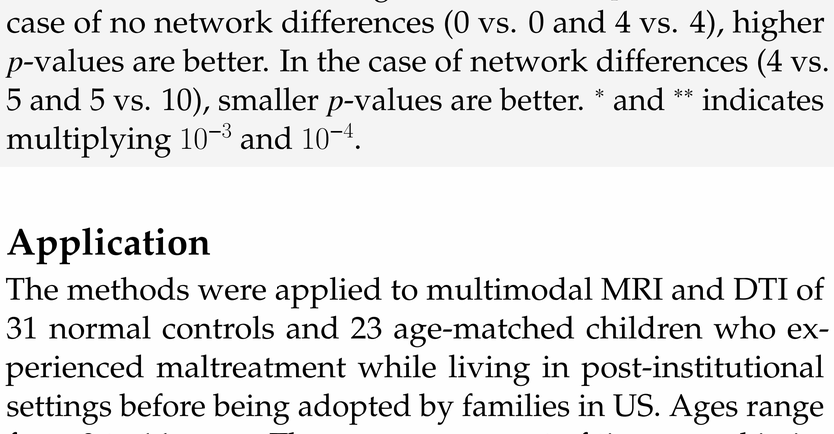

Figure 4. The plots of β_0 (left) and γ (right) over $\sqrt{1-\text{corr}}$. of FA-values. The plots clearly show the structural network differences between maltreated children (dotted red) and normal controls (solid black) on 1856 nodes. The maximum gap between the plots is KS-distance.

Correlation within modality. The correlations of the Jacobian determinant and FA-values were computed between nodes within each modality. Using KS-distance, we determined the statistical significance of differences β_0 and γ plots for each modality separately (Fig. 4). The statistical results in terms of p-values are all below 0.0001 indicating the very strong overall structural network differences in both MRI and DTI.

Cross-correlation across modalities. We also computed the cross-correlation between the Jacobian determinants and FA-values. The statistical significance of the cross-correlation matrix differences is then determined using KS-distance (p-value < 0.0001). The KS-distance method is robust under the change of node size and we also obtained the similar result when the node size changed to 548.

References

M.K. Chung, J.L. Hanson, J. Ye, R.J. Davidson, and S.D. Pollak. Persistent homology in sparse regression and its application to brain morphometry. *IEEE Transactions on Medical Imaging*, 34:1928–1939, 2015.

M.K. Chung, V. Vilalta-Gil, H. Lee, P.J. Rathouz, B.B. Lahey, and D.H. Zald. Exact topological inference for paired brain networks via persistent homology. In *Information Processing in Medical Imaging (IPMI), Lecture Notes in Computer Science*, volume 10265, pages 299–310, 2017.

H. Lee, M.K. Chung, H. Kang, B.-N. Kim, and D.S. Lee. Computing the shape of brain networks using graph filtration and Gromov-Hausdorff metric. *MICCAI, Lecture Notes in Computer Science*, 6892:302–309, 2011.

H. Lee, H. Kang, M.K. Chung, B.-N. Kim, and D.S Lee. Persistent brain network homology from the perspective of dendrogram. *IEEE Transactions on Medical Imaging*, 31:2267–2277, 2012.