Computing the Shape of Brain Networks using Graph Filtration and Gromov-Hausdorff Metric

Hyekyoung Lee [1,2,3], Moo K. Chung [2,6,7], Hyejin Kang [1,3],

Boong-Nyun Kim [5], Dong Soo Lee [1,3,4]

[1] Department of Nuclear Medicine

[2] Department of Brain and Cognitive Sciences

[3] Institute of Radiation Medicine, Medical Research Center

[4] WCU Department of Molecular Medicine and Biopharmaceutical Sciences

[5] Department of Neuropsychiatry

Seoul National University , Korea

[6] Department of Biostatistics and Medical Informatics

[7] Waisman Laboratory for Brain Imaging and Behavior

University of Wisconsin, Madison, WI 53705, USA

Email: mkchung@wisc.edu

March 13, 2011

1

Abstract

The difference between networks has been often assessed by the difference of global topological measures such as the clustering coefficient, degree distribution and modularity. In this paper, we introduce a new framework for measuring the network difference using the Gromov-Hausdorff (GH) distance, which is often used in shape analysis. In order to apply the GH distance, we need to know the shape of the network which is not well defined. Thus, we define the *shape of the brain network* by piecing together the patches of locally connected nearest neighbors using the *graph filtration*. The shape of the network is then transformed to an algebraic form called the single linkage matrix. The single linkage matrix is subsequently used in measuring network differences using the GH distance. As an illustration, we apply the proposed framework to compare the FDG-PET based functional brain networks out of 24 attention deficit hyperactivity disorder (ADHD) children, 26 autism spectrum disorder (ASD) children and 11 pediatric control subjects.

1 Introduction

The functional and anatomical connectivity studies based on graph theory have provided new understanding of human brain [1, 2, 3, 4]. The characteristic of the brain network is quantified by the global topological measures such as clustering coefficient, characteristic path length and modularity [1, 5, 6]. The network comparison is then performed by determining the difference between these topological measures. Each measure reflects different topological characteristic of the brain network. For example, the clustering coefficient and characteristic path length are related with the small-worldness, the assortativity and betweenness are related with the scale-freeness and the modularity is related with the community structure [7, 1]. These measures give us a clue for whether the networks have similar topological properties. However, it is unclear which measure is appropriate for network comparison.

Instead of trying to find one particular characteristic of network at a given scale, one can also look at the overall change of topological features through persistent homology [8, 9].

2

In the persistent homology, the topological features such as the connected components and circles of the network are tabulated in terms of the algebraic form known as Betti numbers. The network difference is then often measured using the bottleneck distance which basically ignores the geometric information of network nodes.

In this paper, we propose a radically different computational framework for determining network difference. Instead of trying to model the topological features of networks, we first define the *shape of network* using the topological concept called the *graph filtration*. The graph filtration is a new graph simplification technique that iteratively build a nested subgraphs of the original graph. The algorithm simplifies a complex graph by piecing together the patches of locally connected nearest nodes. The process of graph filtration can be shown to be mathematically equivalent to the single linkage hierarchical clustering and dendrogram construction.

Once the shape of network is defined, we transform the shape into an algebraic form called the *single linkage matrix*. The single linkage matrix is subsequently used in computing the network difference using the Gromov-Hausdorff (GH) metric. The GH metric is a deformation-invariant dissimilarity measure often used in matching deformable shapes [10, 11]. The GH metric was never used in measuring the distance between brain networks before.

The proposed method is applied in differentiating functional brain networks with 96 regions of interest (ROIs) extracted from FDG-PET data for 24 attention-deficit hyperactivity disorder (ADHD), 26 autism spectrum disorder (ASD) and 11 pediatric controls (PedCon). Numerical experiments show that the graph filtration framework can differentiate the populations better than most known graph theoretic approaches and the recently popular persistent homology framework.

The methodological contributions of this paper are:

(1) We propose a new geometric framework for defining the shape of networks using graph filtration. We introduce the concept of graph filtration and show that it is equivalent

to the single linkage hierarchical clustering and dendrogram construction. This implies that there is a mapping from any complex networks to dendrograms.

(2) We determine the distance between networks using the Gromov-Hausdorff metric for the first time. The framework is then used in determining the brain network difference.

(3) We demonstrate that our framework outperforms most of graph theoretic measures and the recently popular persistent homology framework.

2   Main Ideas

The main problem we are trying to solve is to compare and quantify the brain network differences in ADHD, ASD and PedCon populations. We start with briefly introducing the correlation-based brain network construction.

2.1   Brain Network Construction.

Suppose FDG-PET measurements are obtained in p selected ROIs in n subjects. Each ROI serves as a node in the brain network. Let $X = \{x_1, \cdots, x_p\}$ be the collection of such nodes. Let f_i be the FDG-PET measurement at the node x_i modeled as a random variable. The measurement f_i are assumed to be distributed with mean zero and the covariance $\Sigma = [\sigma_{ij}] \in \mathbb{R}^{p \times p}$. The correlation between f_i and f_j is given by

$$\text{corr}(f_i, f_j) = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}.$$

We can define the metric between the nodes x_i and x_j through the correlation:

$$c_X(x_i, x_j) = 1 - \text{corr}(f_i, f_j).$$

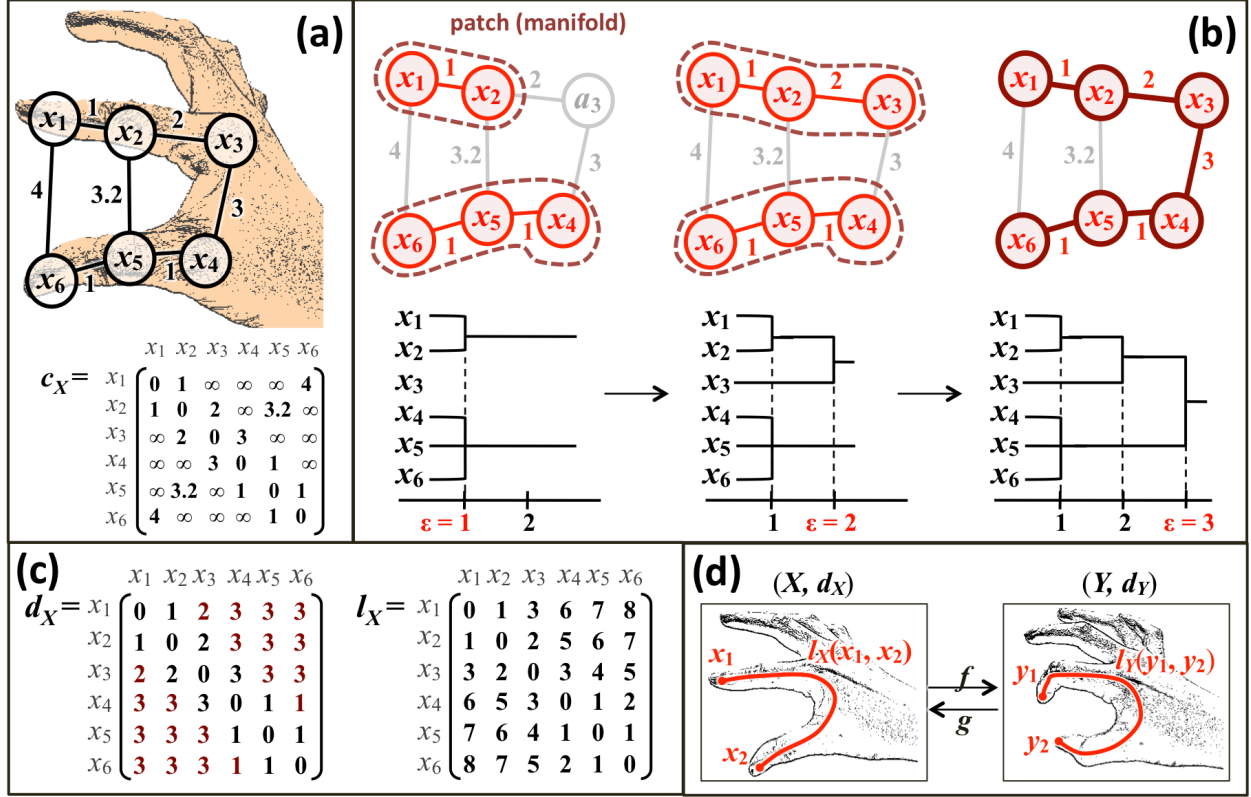Then the brain network can be represented as the metric space (X, c_X).

Figure 1: (a) An example of shape representation using a network of nodes $X = \{x_1, \ldots, x_6\}$ and the distance c_X. The pair (X, c_X) defines the hand. (b) Graph filtration algorithm for representing the graph (X, c_X). (c) The resulting shape can be equivalently represented as the single linkage matrix d_X and the geodesic distance matrix l_X. d_X and l_X are deformation-invariant. (d) A deformable hand where d_X and l_X are invariant.

2.2 Shape of Brain Network.

One can characterize the deformable shapes in images using a collection of nodes and the mapping between the deforming nodes. In deformation-invariant shape matching frameworks [10, 11], we can identify an open and bended hands as equivalent by establishing the correspondence between nodes (Fig. 1 (d)).

Unlike shapes in images, the shape of brain network is difficult to define and visualize since it is not determined by the Euclidean distance between the nodes, but the correlation

5

between measurements on the nodes. In this paper, we define the *shape of the network* by piecing together patches of locally connected nearest neighbor nodes in an iterative fashion as illustrated in Fig. 1 (b).

The brain network can be viewed as the weighted graph (X, c_X) consisting of the collection of nodes X and the distance c_X. We start with $\epsilon = 0$ and increase the ϵ value at each iteration. The value of ϵ is taken discretely from the smallest $c_X(x_i, x_j)$ to the largest $c_X(x_i, x_j)$.

We connect two nodes x_i and x_j if the distance $c_X(x_i, x_j)$ is smaller than ϵ. By increasing ϵ, more connected edges are allowed and larger patches are generated. If two nodes are already connected directly or indirectly *via* other intermediate nodes in smaller ϵ values, we do not connect them. For example, in Fig. 1 (b), we do not connect x_2 and x_3 at $\epsilon = 3.2$ since they were already connected through other nodes at $\epsilon = 3$. When ϵ is larger than any distance $c_X(x_i, x_j)$, the iteration terminates since the graph does not change anymore.

Suppose G_j corresponds the graph obtained at the j-th iteration with $\epsilon = \epsilon_j$. Then for

$$\epsilon_1 < \epsilon_2 < \epsilon_3 < \cdots ,$$

the algorithm generates the sequence of graphs

$$G_1 \subset G_2 \subset G_3 \subset \cdots .$$

Such the sequence of the nested graph is called the *graph filtration* in algebraic topology[8, 9]. In this fashion, we define the shapes of the brain network as a sequence of nested subgraphs (Fig. 2).

2.3   Single Linkage Matrix.

The graph filtration exactly corresponds to the single linkage hierarchical clustering as demonstrated in Fig. 1 (b). The equivalence to the graph filtration and the dendrogram is self-evident. The linking of nodes x_i and x_j corresponds to the linking of leaves in the dendrogram. Increasing the ϵ value in the graph filtration corresponds to increasing the height of the dendrogram.

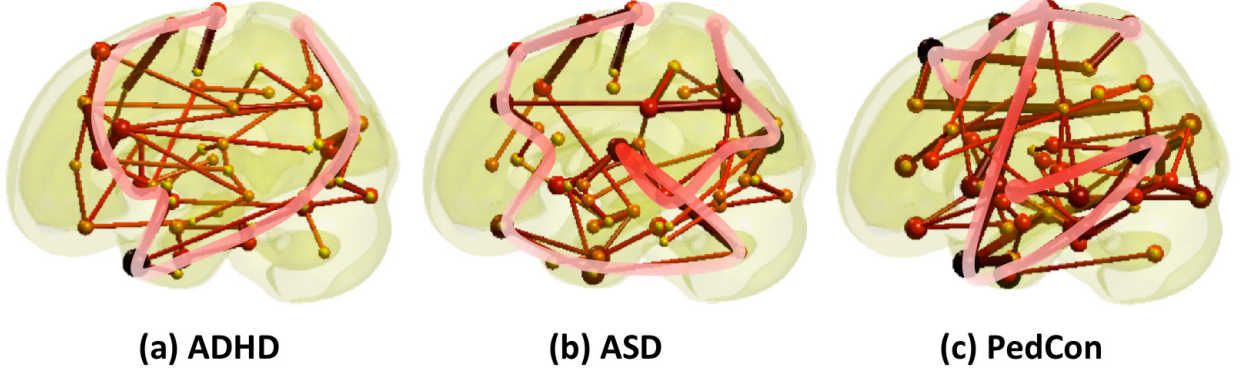**(a) ADHD**   **(b) ASD**   **(c) PedCon**

Figure 2: The shapes of brain networks at the end of the graph filtration (a) ADHD, (b) ASD and (c) PedCon.

In the hierarchical clustering, the distance between patches of nodes $\mathcal{C}_1$ and $\mathcal{C}_2$ is given by the distance between the closest members in $\mathcal{C}_1$ and $\mathcal{C}_2$:

$$d_X(\mathcal{C}_1, \mathcal{C}_2) = \min_{x_1 \in \mathcal{C}_1} \min_{x_2 \in \mathcal{C}_2} c_X(x_1, x_2).$$

For example, when $\epsilon = 3$, the distance between the two patches $\{x_1, x_2, x_3\}$ and $\{x_4, x_5, x_6\}$ is given by the distance between x_3 and x_4. Thus, we can represent the shape of brain network as the *single linkage matrix*, where the elements are the single linkage distances between nodes.

Instead of using the single linkage matrix, we can use the geodesic distance matrix. The geodesic distance l_X is defined as the sum of edge weights along the shortest path in the graph. For example, in Fig. 1 (b), the geodesic distance $l_X(x_2, x_5)$ is the sum of $c_X(x_2, x_3)$, $c_X(x_3, x_4)$ and $c_X(x_4, x_5)$. We will not discuss the geodesic distance formulation in this paper.

2.4   Gromov-Hausdorff Distance.

After representing the shapes of brain networks, we need to compute the distance between the networks for quantification. Given two metric spaces (X, d_X) and (Y, d_Y), the Gromov-

Hausdorff Distance (GH) distance between X and Y is defined as [12, 10]:

$$d_{GH}(X,Y) = \inf_{\substack{f:X\to Y \\ g:Y\to X}} \frac{1}{2} \max\left(\mathcal{F}(f), \mathcal{G}(g), \mathcal{H}(f,g)\right), \tag{1}$$

where

$$\begin{aligned}
\mathcal{F}(f) &= \sup_{x_1,x_2\in X} |d_X(x_1,x_2) - d_Y(f(x_1), f(x_2))|, \\
\mathcal{G}(g) &= \sup_{y_1,y_2\in Y} |d_X(g(y_1), g(y_2)) - d_Y(y_1, y_2)|, \\
\mathcal{H}(f,g) &= \sup_{x\in X, y\in Y} |d_X(x, g(y)) - d_Y(f(x), y)|.
\end{aligned}$$

We used the single linkage distance for d_X and d_Y. Note that the single linkage distance does not satisfy the triangle inequality but satisfies [13]

$$\max(d_X(x_1,x_2), d_X(x_2,x_3)) \geq d_X(x_1,x_3).$$

In computing the GH distance, we need to determine the the isometric embedding functions f and g, which is not straightforward in general [11, 14, 15]. In our problem, all the nodes in X and Y are in the fixed locations, thus, the mapping functions f and g are simply given as $f(x_i) = y_i$ and $g(y_i) = x_i$ and Eq. (1) is discretized as

$$d_{GH}(X,Y) = \frac{1}{2} \max_{\forall i,j} |d_X(x_i, x_j) - d_Y(y_i, y_j)|.$$

3 Experimental Results

3.1 Data Description.

The data consists of 24 ADHD (19 boys, mean age: 8.2 ± 1.6 years), 26 ASD (24 boys, mean age: 6.0 ± 1.8 years) and 11 PedCon (7 boys, mean age: 9.7 ± 2.5 years). PET images were preprocessed using Statistical Parametric Mapping (SPM) package. After spatial normalization to the standard template space, mean FDG uptake within 96 ROIs were extracted. The values of FDG uptake were globally normalized to the individuals total gray matter mean count.
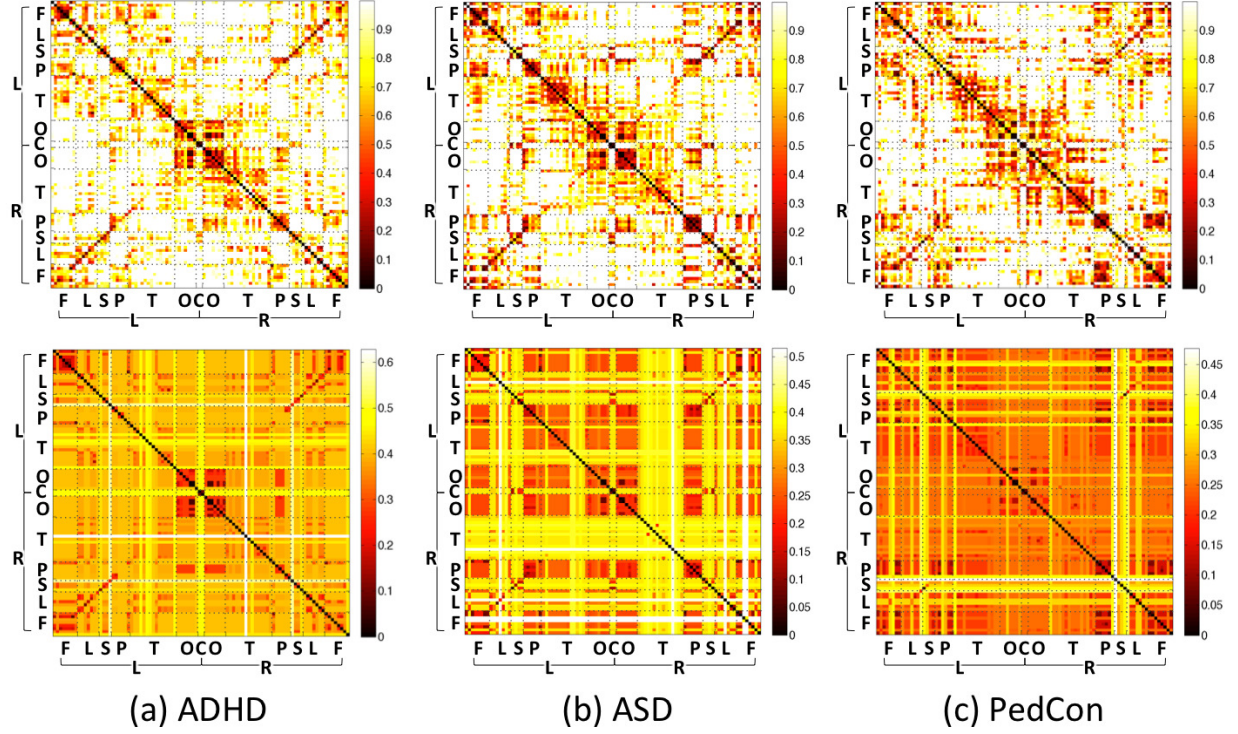
Figure 3: The correlation-based distance c_X (top) and single linkage matrix d_X (bottom) for (a) ADHD, (b) ASD and (c) PedCon. In each connectivity matrix, the upper-left and the lower-right 48 ROIs are from left and right hemispheres, respectively. The order of ROIs of the left and the right hemispheres are horizontally and vertically symmetric, thus, the diagonal terms from the top-right to the bottom-left represents bilateral symmetry of brain.

3.2 Comparison of the connectivity matrix.

The distance matrices obtained from correlation c_X and single linkage matrices d_X are shown in Fig. 3. The group difference is more evident in the single linkage matrices. The maximum single linkage distances of ADHD, ASD and PedCon are 0.62, 0.51, 0.48. The most regions in ADHD are weakly connected except a few strongly connected regions within the occipital (O) and left frontal (F) regions and between the right and the left frontal regions [16]. On the other hand, PedCon network is well-connected in the whole brain regions. In ASD, the connection is segmented according to lobes and temporal (T) asymmetry is obviously visible [17].
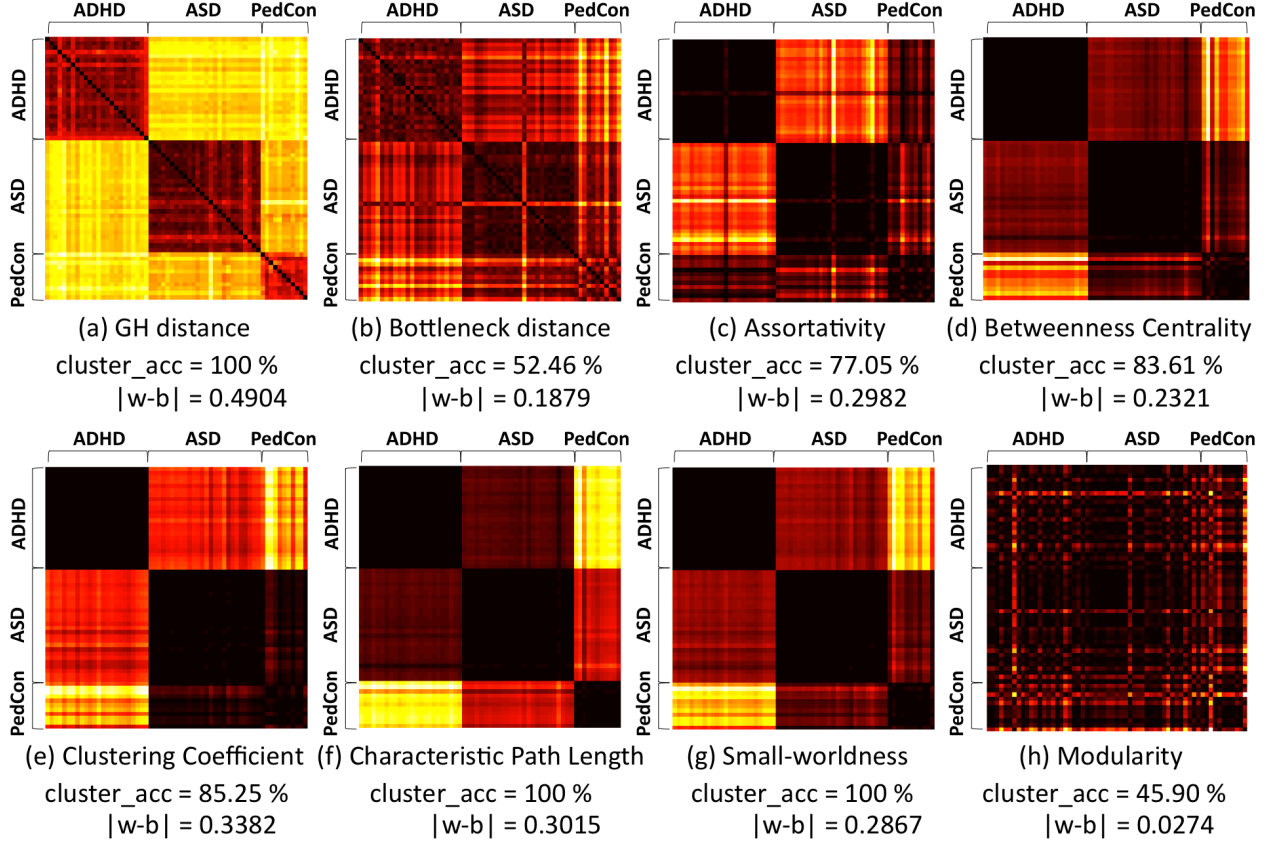
Figure 4: Network comparison using various network measures: (a) GH distance, (b) bottleneck distance, (c) assortativity, (d) centrality, (e) clustering coefficient, (f) characteristic path length, (g) small-worldness and (h) modularity.

3.3 Performance against other network measures.

We performed the leave-one-out cross validation on the network differences using 8 different measures including the GH distance, bottleneck distance, assortativity, centrality, clustering coefficient, characteristic path length, small-worldness and modularity (Fig. 4) [9, 1, 6].

After constructing the distance matrices, we divided the networks into 3 clusters using the hierarchical clustering and evaluated the clustering accuracy by comparing the assigned labels with the true labels. The clustering accuracies of GH distance, characteristic path length and small-worldness are all 100 %. However, the distance between the groups is much larger than the distance within the groups in the GH metric, i.e. $|w - b| = 0.49$ in (Fig.

4(a)), indicating the superior performance of the GH-metric.

4 Conclusions

We presented a novel framework for computing the distance on networks. Using the graph filtration, we defined the shape of the network as a sequence of nested subgraphs. The graph filtration is then transformed into an algebraic form called the single linkage matrix. The single linkage matrices were demonstrated to differentiate the group differences in the ADHD, ASD and PedCon populations. The distance between different single linkage matrices is quantified using the Gromov-Hausdorff metric. The Gromov-Hausdorff metric was validated against other global network measures from graph theory and persistent homology: bottleneck distance, assortativity, centrality, clustering coefficient, characteristic path length, small-worldness and modularity. The GH metric outperforms all of them in terms of the clustering accuracy and the difference between the within- and the between-group distance.

Acknowledgment

References

[1] Rubinov, M., Sporns, O.: Complex network measures of brain connectivity: Uses and interpretations. NeuroImage **52** (2010) 1059–1069

[2] Laurienti, P., Hugenschmidt, C., Hayasaka, S.: Modularity maps reveal community structure in the resting human brain. Nature Preceedings (2009)

[3] Bassett, D.: Small-world brain networks. The Neuroscientist **12** (2006) 512–523

[4] Sporns, O., Tononi, G., Kotter, R.: The human connectome: a structural description of the human brain. PLoS Computational Biology **1** (2005)

[5] Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., Hwang, D.: Complex networks: Structure and dynamics. Physics Reports **424** (2006) 175–308

[6] Bullmore, E., Sporns, O.: Complex brain networks: graph theoretical analysis of structural and functional systems. Nature Reviews Neuroscience **10** (2009) 186–198

[7] Li, L., Alderson, D., Doyle, J.C., Willinger, W.: Towards a theory of scale-free graphs: Definition, properties, and implications. Internet Mathematics **2** (2005) 431–523

[8] Adler, R., Bobrowski, O., Borman, M., Subag, E., Weinberger, S.: Persistent homology for random fields and complexes. ArXiv e-prints (2010)

[9] Edelsbrunner, H., Harer, J.: Persistent homology - a survey. Contemporary Mathematics **453** (2008) 257–282

[10] Mémoli, F.: Gromov-hausdorff distances in euclidean spaces. In: Workshop on Non-Rigid Shape Analysis and Deformable Image Alignment (CVPR workshop, NORDIA'08). (2008)

[11] Bronstein, A.M., Bronstein, M.M., Kimmel, R.: Efficient computation of isometry-invariant distances between surfaces. SIAM Journal on Scientific Computing **28** (2006) 1812–1836

[12] Kalton, N.J., Ostrovskii, M.: Distances between banach spaces. Forum Mathematicum **11** (1999) 17–48

[13] Carlsson, G., Mémoli, F.: Characterization, stability and convergence of hierarchical clustering methods. Journal of Machine Learning Research **11** (2010) 1425–1470

[14] Mémoli, F., Sapiro, G.: Comparing point clouds. In: SGP '04: Proceedings of the 2004 Eurographics/ACM SIGGRAPH symposium on Geometry processing. (2004) 32–40

[15] Mémoli, F., Sapiro, G.: A theoretical and computational framework for isometry invariant recognition of point cloud data. Foundations of Computational Mathematics **5** (2005) 313–347

[16] Kraina, A., Castellanos, F.: Brain development and adhd. Clinical Psychology Review **26** (2006) 433–444

[17] Minshew, N., Williams, D.: The new neurobiology of autism: Cortex, connectivity, and neuronal organization. Arch Neurol. (2007) 945–950