



Coidentification of Group-Level Hole Structures in Brain Networks via Hodge Laplacian

Hyekyoung Lee^{1(✉)}, Moo K. Chung³, Hyejin Kang², Hongyoon Choi^{1,2},
Seunggyun Ha⁴, Youngmin Huh², Eunkyung Kim¹, and Dong Soo Lee^{1,2}

¹ Seoul National University Hospital, Seoul, Republic of Korea
hklee.brain@gmail.com

² Seoul National University, Seoul, Republic of Korea

³ University of Wisconsin, Madison, USA

⁴ The Catholic University of Korea, Seoul ST. Mary's Hospital,
Seoul, Republic of Korea

Abstract. One of outstanding issues in brain network analysis is to extract common topological substructure shared by a group of individuals. Recently, methods to detect group-wise modular structure on graph Laplacians have been introduced. From the perspective of algebraic topology, the modules or clusters are the zeroth topology information of a topological space. Higher order topology information can be found in holes. In this study, we extend the concept of graph Laplacian to higher order Hodge Laplacian of weighted networks, and develop a group-level hole identification method via the Stiefel optimization. In experiments, we applied the proposed method to three synthetic data and Alzheimer's disease neuroimaging initiative (ADNI) database. Experimental results showed that the coidentification of group-level hole structures helped to find the underlying topology information of brain networks that discriminate groups well.

Keywords: Hole structure · Group analysis · Hodge Laplacian · Stiefel optimization · ADNI

1 Introduction

Persistent homology has been widely applied for brain network analysis [4,9]. Especially, the concept of filtration in persistent homology helps to solve the thresholding problem of correlation-based brain networks. During the filtration, a weighted network is decomposed into the sequence of unweighted networks at all possible thresholds, and the change of network shape is observed over thresholds. The persistent homology focuses on the birth and death of topological features, called holes, and the change of their numbers, called Betti numbers, during the filtration. Persistence diagram is a useful tool for analyzing and visualizing the change of holes and Betti numbers during the filtration. The most widely

used algorithm to compute persistent homology was developed by Zomorodian and Carlsson, called the ZC algorithm [11]. It is very efficient in computing a persistence diagram. However, since the ZC algorithm finds only one sparse representation of a hole among many possible representations, it can obscure the identification of holes in practice [6, 7].

The k -dimensional holes in a network span in the intersection of the null space of the k th and $(k + 1)$ th incidence matrices [5–7, 11]. The null space of a matrix is generally estimated by the Gaussian elimination or eigen-decomposition. The ZC algorithm is on the Smith normal form of the incidence matrices, which is the extension of the Gaussian elimination [11]. Another approach to the hole estimation is based on the eigen-decomposition of the k th Hodge Laplacian constructed by the k th and $(k + 1)$ th incidence matrices [5–7, 11]. The Gaussian elimination of sparse incidence matrices gives us the sparse representation of holes. On the other hand, the eigen-decomposition of Hodge Laplacian estimates a dense representation of holes, called a harmonic form. The harmonic form can express all possible paths around a hole with their weights representing the amount of contributions [6]. Another advantage of the harmonic form estimation method is to represent the convex optimization problem on the Stiefel manifold of Hodge Laplacian [8]. If we change the harmonic form estimation method to an optimization problem, we can easily add regularization term such as sparseness constraints as well as extend to group-level analysis [8].

A further consideration when computing holes during the filtration is that the shape of a hole is changed from its birth to death by increasing the number of edges during the filtration. The ZC algorithm chooses the youngest persistent hole at birth for the localization of a hole. However, the best solution in the hole localization is to find the consistent shape of a hole from its birth to death. In this paper, we add the weight to simplexes including edges and higher order counterparts in a network for the consistency of holes during filtration [5]. Since the connections (topology) in a network is not changed by the weight of a simplex, the Betti numbers and the birth and death of holes are not affected by the presence of weights. The harmonic form is affected both by topology and geometry. If we assign large weights to edges at birth and small weights to edges at death, it is possible to reduce the change of the shape of the harmonic form.

The contribution of this paper is (1) to introduce a weighted version of the k th Hodge Laplacian for computing harmonic forms of a brain network during filtration, (2) to propose a harmonic form estimation method based on the Stiefel optimization, and (3) to extend a group analysis of harmonic form estimation by adding the constraint of pairwise similarity between harmonic forms in a group. In experiments, we applied to three synthetic data and the FDG PET dataset in Alzheimer’s disease neuroimaging initiative (ADNI) database. The experimental results showed that coidentification of group-level hole structure helped to find the underlying topology information that improves the clustering accuracy between the groups.

2 Methods

2.1 Datasets and Preprocessing

FDG PET images in ADNI data set was used for application¹. It consists of 4 groups, 181 normal controls (NC), 91 mild cognitive impairment non-converters (MCI_n), 77 MCI converters (MCI_c), and 135 Alzheimer’s disease (AD). The FDG PET images were preprocessed by statistical parametric mapping (SPM8)². The whole brain image was divided into 94 regions of interest (ROIs) based on automated anatomical labeling (AAL2) excluding cerebellum. The distance between two ROIs (nodes) was obtained by diffusion distance on positive correlation between the measurements [2]. We transformed the distance to the similarity (weight) between two nodes by Gaussian kernel with bandwidth 0.005 [1]. The edge weight between nodes i and j are denoted by w_{ij} .

2.2 Hodge Laplacian of Weighted Simplicial Network

Weighted Simplicial Network. Given a non-empty node set V , a k -simplex is an element with nodes, $v_1, \dots, v_{k+1} \in V$, denoted by $\sigma_k = [v_1, \dots, v_{k+1}]$. An abstract simplicial complex K is a subset of the power set of V , i.e., $K \subseteq 2^V$ such that (1) $\emptyset \in K$, and (2) if $\sigma \in K$ and $\tau \in \sigma$, $\tau \in K$ [11]. The dimension of K , denoted by $\dim K$, is the maximum dimension of $\sigma \in K$. The collection of σ_k ’s in K is denoted by K_k ($-1 \leq k \leq \dim K$). The number of simplices in K_k is denoted by $|K_k|$. The $(k - 1)$ -face of σ_k is obtained by $\sigma_{k,\check{j}} = [v_1, \dots, v_{j-1}, v_{j+1}, \dots, v_{k+1}] \in K_{k-1}$. We call a simplicial complex a (simplicial) network for convenience [6].

In a weighted simplicial network, each simplex has its own weight function $w : 2^V \rightarrow (0, \infty)$. Suppose that a weighted simplicial network has the ordered $\sigma_k^i \in K_k$, $i = 1, \dots, |K_k|$ for $0 \leq k \leq \dim K$. The weight matrix of σ_k ’s are a diagonal matrix such that $\mathbf{W}_k = \text{diag} \{w(\sigma_k^1), \dots, w(\sigma_k^{|K_k|})\} \in \mathbb{R}^{|K_k| \times |K_k|}$. In this study, we define the weight of simplexes as follows [10]:

$$\begin{cases} w(\sigma_0 = [v_i]) = 1(1 \leq i \leq p), \\ w(\sigma_1 = [v_i, v_j]) = w_{ij}(1 \leq i < j \leq p), \\ w(\sigma_k) = \min(\sigma_{k,\check{1}}, \dots, \sigma_{k,\check{k+1}}) \text{ for } k > 1. \end{cases}$$

Hodge Laplacian. Given a finite simplicial complex K , a chain complex C_k is defined by $\mathbb{Z}^{|K_k|}$ with $C_{-1} = \mathbb{Z}$. \mathbb{Z}^k is a k -dimensional integer space [11]. The boundary operator ∂_k and coboundary operator ∂_k^\top are functions such that $\partial_k : C_k \rightarrow C_{k-1}$ and $\partial_k^\top : C_{k-1} \rightarrow C_k$, respectively. For $k = 0$, $\partial_0 : C_0 \rightarrow 0$.

¹ <http://adni.loni.usc.edu>.
² <http://www.fil.ion.ucl.ac.uk/spm>.

Given $\sigma_k = [v_1, \dots, v_{k+1}] \in C_k$ and its weight $w(\sigma_k)$, the boundary of σ_k is algebraically defined as [5, 10]

$$\partial_k \sigma_k = \sum_{j=1}^{k+1} \frac{w(\sigma_k)}{w(\sigma_{k,\check{j}})} (-1)^{j-1} \sigma_{k,\check{j}}. \tag{1}$$

If the sign of σ_{k-1} in $\partial_k \sigma_k$ is positive/negative, it is called positively/negatively oriented with respect to σ_k . It is denoted by $\sigma_{k-1} \in_{+(-)} \sigma_k$.

Given $K_k, K_{k-1} \subset K$, the k th incidence matrix $\mathbf{M}_k \in \mathbb{Z}^{|K_{k-1}| \times |K_k|}$ is defined by $[\mathbf{M}_k]_{ij} = 1$, if $\sigma_{k-1}^i \in_+ \sigma_k^j$; -1 , if $\sigma_{k-1}^i \in_- \sigma_k^j$; 0 otherwise [5]. Then, the boundary operator ∂_k in (1) is written by $\partial_k = \mathbf{W}_k^{-1} \mathbf{M}_k \mathbf{W}_k$, and The coboundary operator ∂_k^\top is written by $\mathbf{W}_k \mathbf{M}_k^\top \mathbf{W}_{k-1}^{-1}$ [3]. The k th Hodge Laplacian $\mathbf{L}_k : C_k \rightarrow C_k$ of a weighted simplicial network is defined by [3, 5]

$$\begin{aligned} \mathbf{L}_k &= \partial_k^\top \partial_k + \partial_{k+1} \partial_{k+1}^\top \\ &= \mathbf{W}_k \mathbf{M}_k^\top \mathbf{W}_{k-1}^{-2} \mathbf{M}_k \mathbf{W}_k + \mathbf{W}_k^{-1} \mathbf{M}_{k+1} \mathbf{W}_{k+1}^2 \mathbf{M}_{k+1}^\top \mathbf{W}_k^{-1}. \end{aligned} \tag{2}$$

Harmonic Form. The kernel of \mathbf{L}_k are denoted by $\ker \mathbf{L}_k$. The dimension of $\ker \mathbf{L}_k$, i.e., the number of zero eigenvalues of \mathbf{L}_k is the k th Betti number, denoted by β_k . Eigenvectors with zero eigenvalues in $\ker \mathbf{L}_k$ are called a harmonic k -form, denoted by $\mathbf{H}_k = [\mathbf{h}_{1,k}, \dots, \mathbf{h}_{\beta_k,k}] \in \mathbb{R}^{|K_k| \times \beta_k}$. The eigenvalue of $\mathbf{h}_{i,k}$ can be written by $\lambda_{i,k} = \mathbf{h}_{i,k}^\top \mathbf{L}_k \mathbf{h}_{i,k} = 0$ for $i = 1, \dots, \beta_k$.

2.3 Stiefel Optimization for Group-Level Harmonic Forms

Given a network K , the problem of estimating harmonic k -forms of \mathbf{L}_k can be written by an optimization problem on a Stiefel manifold

$$\min_{\mathbf{H}_k \in \mathcal{S}(|K_k|, r)} \mathbf{H}_k^\top \mathbf{L}_k \mathbf{H}_k + \beta \|\mathbf{H}_k\|_1, \tag{3}$$

where $\mathcal{S}(|K_k|, r)$ is a Stiefel manifold which is the set of all r -tuples of orthonormal vectors in $\mathbb{R}^{|K_k|}$, $\|\cdot\|_1$ is the l_1 -norm of \cdot , and β is the control parameter for sparseness [8].

Pairwise Similarity Constraint for Group Analysis. Suppose that there are N simplicial networks in a group. Their k th Hodge Laplacians and harmonic forms are denoted by $\mathbf{L}_k^{(1)}, \dots, \mathbf{L}_k^{(N)}$ and $\mathbf{H}_k^{(1)}, \dots, \mathbf{H}_k^{(N)}$, respectively. To estimate group-level harmonic forms, we extend (3) to

$$\begin{aligned} \min_{\mathbf{H}_n \in \mathcal{S}(|K_k|, r)} \sum_{n=1}^N \left((\mathbf{H}_k^{(n)})^\top \mathbf{L}_k^{(n)} \mathbf{H}_k^{(n)} + \beta \|\mathbf{H}_k^{(n)}\|_1 \right) \\ + \frac{\mu}{2} \sum \sum_{m,n, 1 \leq m \neq n \leq N} \|\mathbf{H}_k^{(m)} - \mathbf{H}_k^{(n)}\|_2^2, \end{aligned} \tag{4}$$

where μ is a parameter to control pairwise similarity between harmonic forms. When $\mu = 0$, the obtained $\mathbf{H}_k^{(n)}$ s are an individual-level harmonic form. The larger the μ , the stronger the group-level constraint. The derivative of the optimization problem in (4) is $2\mathbf{L}_k^{(n)}\mathbf{H}_k^{(n)} + \beta \text{sign}(\mathbf{H}_k^{(n)}) + \mu \sum_{m=1, m \neq n}^N (-\mathbf{H}_k^{(m)})$. We used the trust-region algorithm on a Stiefel manifold for the proposed optimization problem (4).

Threshold-Dependency Constraint for Filtration. Suppose that the threshold of weights (similarities between nodes) is given by $\epsilon_1 \geq \epsilon_2 \geq \dots \geq \epsilon_T$. Here, we mainly consider the 2nd Hodge Laplacians and their harmonic 2-forms ($k = 2$). Therefore, we omit k in \mathbf{L}_k , $\mathbf{h}_{i,k}$ and $\lambda_{i,k}$ in (1-4).

The Hodge Laplacian at ϵ_t in the n th network is now written by $\mathbf{L}_{\epsilon_t}^{(n)}$. The group-level harmonic form in (4) is denoted by $\mathbf{H}_{\epsilon_t}^{(n)} = [\mathbf{h}_{1t}^{(n)} \dots \mathbf{h}_{rt}^{(n)}] \in \mathbb{R}^{q \times r}$, where q is the number of edges in $K^{(n)}$ and r is β_2 at ϵ_t . The eigenvalue of $\mathbf{H}_{\epsilon_t}^{(n)}$ is denoted by $\lambda_{\epsilon_t}^{(n)} = [\lambda_{it}^{(n)} = (\mathbf{h}_{it}^{(n)})^\top \mathbf{L}_{\epsilon_t}^{(n)} \mathbf{h}_{it}^{(n)}] \in \mathbb{R}^r$. Then, we have the hole sequence of the n th network at $\epsilon_1, \epsilon_2, \dots, \epsilon_T$ defined by

$$\mathcal{H}_n : (\mathbf{H}_{\epsilon_1}^{(n)}, \lambda_{\epsilon_1}^{(n)}) \rightarrow \dots \rightarrow (\mathbf{H}_{\epsilon_T}^{(n)}, \lambda_{\epsilon_T}^{(n)}). \tag{5}$$

To impose the dependency between $(\mathbf{H}_{\epsilon_{t-1}}^{(n)}, \lambda_{\epsilon_{t-1}}^{(n)})$ and $(\mathbf{H}_{\epsilon_t}^{(n)}, \lambda_{\epsilon_t}^{(n)})$, we replace $\mathbf{L}_{\epsilon_t}^{(n)}$ in (4) with $\tilde{\mathbf{L}}_{\epsilon_t}^{(n)} = (1 - \alpha)\mathbf{L}_{\epsilon_{t-1}}^{(n)} + \alpha\mathbf{L}_{\epsilon_t}^{(n)}$ ($0.5 \ll \alpha \leq 1$). Then, a persistent harmonic form lasted from the threshold ϵ_{t-1} will be selected first at ϵ_t .

2.4 Similarity Between Hole Sequences

As eigenvalue $\lambda_{it}^{(n)}$ approaches zero, the corresponding eigenvector $\mathbf{h}_{it}^{(n)}$ becomes a harmonic form. Given two hole sequences, \mathcal{H}_m and \mathcal{H}_n , we define a similarity between \mathcal{H}_m and \mathcal{H}_n as follows:

$$\Gamma(\mathcal{H}_m, \mathcal{H}_n) = \sum_{i=1}^r \sum_{t=1}^T w(\lambda_{it}^{(m)}, \lambda_{it}^{(n)}) \left| (\mathbf{h}_{it}^{(m)})^\top \mathbf{h}_{it}^{(n)} \right|,$$

where $w(\lambda_{it}^{(m)}, \lambda_{it}^{(n)}) = \frac{\exp(-\lambda_{it}^{(m)}/s) \exp(-\lambda_{it}^{(n)}/s)}{\sum_t \exp(-\lambda_{it}^{(m)}/s) \exp(-\lambda_{it}^{(n)}/s)}$. Since $\mathbf{h}_{it}^{(m)}$ and $\mathbf{h}_{it}^{(n)}$ are an eigenvector, $0 \leq \left| (\mathbf{h}_{it}^{true})^\top \mathbf{h}_{it}^{(n)} \right| \leq 1$. When $\mathcal{H}_m = \mathcal{H}_n$, $\Gamma(\mathcal{H}_m, \mathcal{H}_n) = 1$. It compares two hole sequences using the shape of holes in the eigenvectors $\mathbf{H}_{\epsilon_t}^{(n)}$ and the existence of the holes in the eigenvalues $\lambda_{\epsilon_t}^{(n)}$.

3 Results

3.1 Experiments on Synthetic Data

The synthetic data used in this experiment were (1) house data with 5 nodes and one hole; (2) snowman data with 11 nodes and two holes; and (3) three

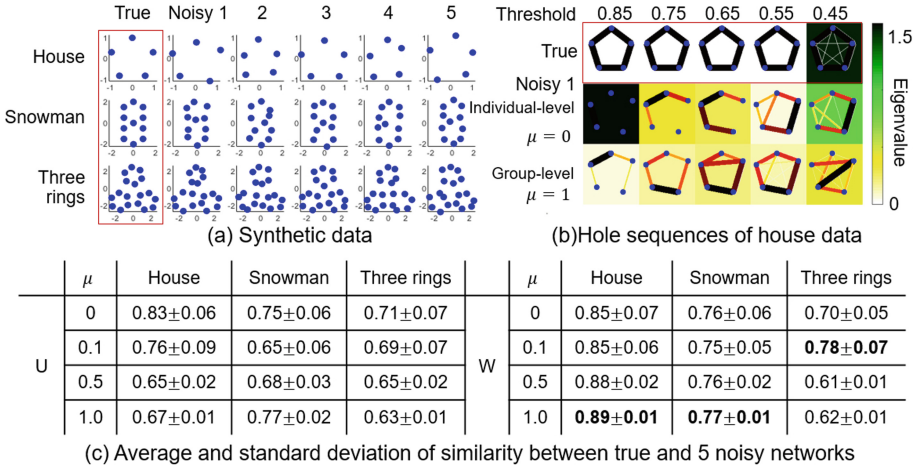


Fig. 1. (a) Synthetic data, house, snowman, and three rings. Each data consists of true network and 5 noisy networks. (b) The first row shows the hole sequence of the true network, and the second and third rows respectively show the individual- and group-level hole sequences of the noisy network 1 in the house data. The similarity between the true and individual-level hole sequences is 0.74, and the similarity between the true and group-level hole sequences is 0.88. (c) Average and standard deviation of the similarities of hole sequences between true and 5 noisy networks. U and W represent when not using and using the weight of simplexes during the harmonic form estimation. (Color figure online)

rings data with 18 nodes and three holes. We generated 5 noisy networks by adding Gaussian noise with mean 0 and standard deviation 0.13 to the position of nodes in the true networks in Fig. 1(a). The measurement of a node is the position of a node in (a), and the edge weight were estimated by Gaussian kernel of the measurements of two nodes. We assumed that the individual-level hole sequences of noisy networks ($\mu = 0$) were different from the true hole sequence, while the group-level hole sequences of 5 noisy networks ($\mu > 0$) were close to true topology. We estimated the group-level hole sequences of 5 noisy networks for $\mu = 0, 0.1, 0.5, 1$, and estimated the similarity between the true and 5 individual- and group-level hole sequences. Figure 1(b) showed the true hole sequence, individual- and group-level hole sequences of house data from top to bottom. The background color of each hole was determined by the eigenvalue of the hole as shown in the right colorbar. In each hole, a thick and dark edge had large weight which is proportional to its contribution to the hole. The average of 5 similarities were shown in Fig. 1(c). The similarity to the true network was maximized when we used the weight of simplexes (W) for $\mu > 0$. It meant that the group-level hole estimation could find the true topology of a network.

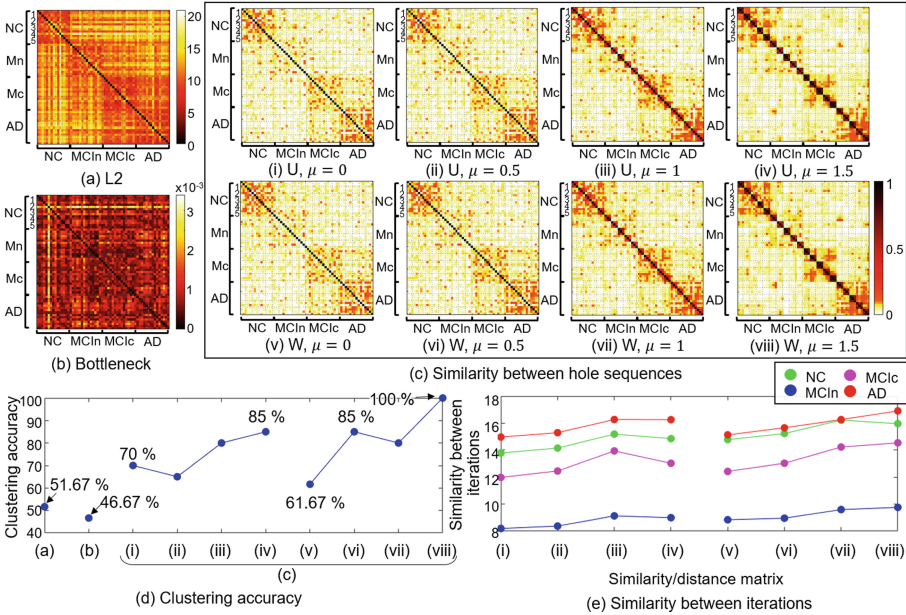


Fig. 2. (a) L2-norm, (b) Bottleneck distance of holes. (c) Similarity matrix between hole sequences for U (top) and W (middle), and $\mu = 0, 0.5, 1.0, 1.5$ from left to right. The similarity matrix of individual-level hole sequences were in (i, v) in (c). When μ increased, the 3 group-level hole sequences in a group resembled each other. Thus, dark 3×3 block matrices with large similarity were found in the diagonal term of the matrices in (iv, viii) in (c). (d) Clustering accuracy. (e) Similarity between hole sequences over CV iterations. The most of maximum accuracy and similarity were found at (viii) W, $\mu = 1.5$ in (c).

3.2 Experiments on ADNI Data

We randomly divided the data of a group into 3 parts, and constructed 3 brain networks for each group, NC, MCln, MClc, and AD. We estimated 3 individual- and group-level hole sequences from 3 networks by varying $\mu = 0, 0.5, 1, 1.5$. We repeated this 3-fold cross validation (CV) procedure 5 times, and compared the individual- and group-level hole structures of 5 CV iterations. The number of networks we used was 3 folds/group \times 4 groups \times 5 CV iterations = 60.

Figure 2(c) showed the similarity between 60 hole sequences when μ increased from left to right. If the group-level holes identified the ground truth hole structure of a group, the group-level holes were reproducible over CV iterations. Moreover, the closer the group-level hole structures were to the ground truth, the better the group-level holes discriminated the groups. We clustered 60 hole sequences into 4 groups by Ward’s hierarchical clustering method based on the similarity matrix in (c). The clustering accuracy was shown in Fig. 2(d) with the accuracy based on L2-norm and bottleneck distance of holes in (a) and (b). The clustering accuracy when using the weight of simplexes in (v–viii) in (c) was

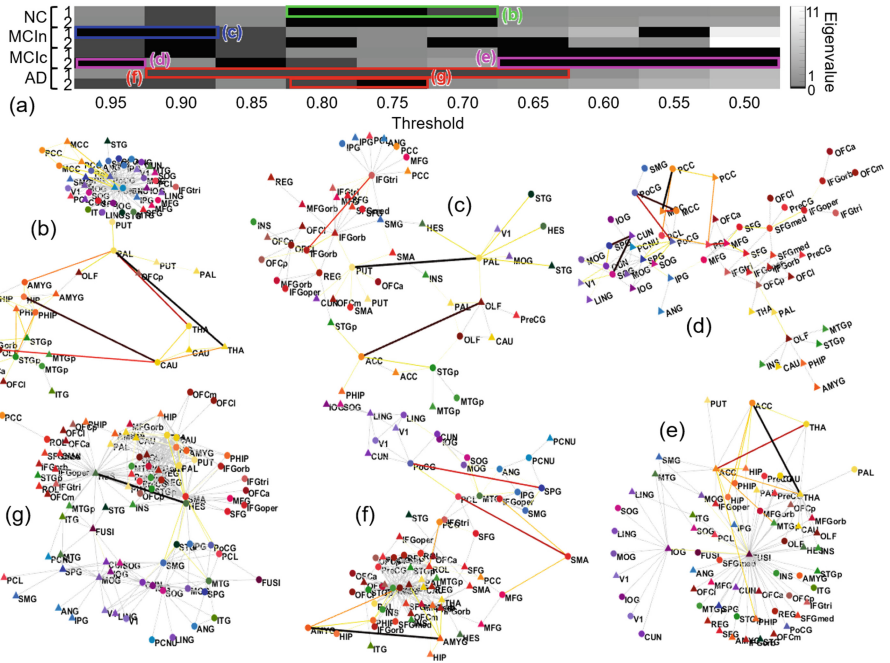


Fig. 3. (a) Sequence of the two smallest eigenvalues of Hodge Laplacian over thresholds. When the eigenvalue was close to zero (dark region), the corresponding eigenvector became a group-level harmonic form. We selected 6 group-level holes from (a). Holes in (b) NC, (c) MCIn, (d, e) MCIC, and (f, g) AD.

better than the clustering accuracy without weights in (i–iv) in Fig. 2(c). The larger the μ , i.e., the stronger the group-level constraint, the better the clustering accuracy. We also estimated the sum of similarity between hole sequences obtained from different CV iterations in Fig. 2(e). The similarity of hole structures increased as μ increased in Fig. 2(e). It meant that the group-level hole structure was reproducible. The similarity was the largest in AD, followed by NC, MCIC, and MCIn. It might be related with the data homogeneity. The sequence of the two smallest eigenvalues over thresholds in each group were shown in Fig. 3(a), and six selected holes were plotted in (b–g). A hole was found in sub-cortical region in NC in (b), while there were holes in cortico-cortical connections in AD in (f, g).

Acknowledgements. Data used in preparation of this article were obtained from the ADNI database <http://adni.loni.usc.edu>. This work is supported by NRF Grants funded by the Korean Government (No. 2013R1A1A2064593, No. 2016R1D1A1B03935463, No. 2015M3C7A1028926, No. 2017M3C7A1048079, No. 2016R1D1A1A02937497, No. 2017R1A5A1015626, and No. 2011-0030815), and NIH grant EB022856.

References

1. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer, New York (2006)
2. Coifman, R.R., Lafon, S., Lee, A.B., Maggioni, M., Warner, F., Zucker, S.: Geometric diffusions as a tool for harmonic analysis and structure definition of data: diffusion maps. In: Proceedings of the National Academy of Sciences (2005)
3. Dawson, R.J.M.: Homology of weighted simplicial complexes. *Cahiers de Topologie et Géométrie Différentielle Catégoriques* **31**(3), 229–243 (1990)
4. Giusti, C., Ghrist, R., Bassett, D.S.: Two’s company, three (or more) is a simplex. *J. Comput. Neurosci.* **41**(1), 1–14 (2016)
5. Horak, D., Jost, J.: Spectra of combinatorial Laplace operators on simplicial complexes. *Adv. Math.* **244**, 303–336 (2013)
6. Kim, Y.J., Kook, W.: Harmonic cycles for graphs. *Linear Multilinear Algebra* **67**, 1–11 (2018)
7. Lee, H., Chung, M.K., Kang, H., Lee, D.S.: Hole detection in metabolic connectivity of alzheimer’s disease using k -laplacian. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) MICCAI 2014. LNCS, vol. 8675, pp. 297–304. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10443-0_38
8. Lu, C., Yan, S., Lin, Z.: Convex sparse spectral clustering: Single-view to multi-view. *IEEE Trans. Image Process.* **25**(6), 2833–2843 (2016)
9. Wu, P., et al.: Optimal topological cycles and their application in cardiac trabeculae restoration. In: Niethammer, M., et al. (eds.) IPMI 2017. LNCS, vol. 10265, pp. 80–92. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59050-9_7
10. Zomorodian, A.: Fast construction of the Vietoris-Rips complex. *Comput. Graph.* **34**, 263–271 (2010)
11. Zomorodian, A., Carlsson, G.: Computing persistent homology. *Discrete Comput. Geom.* **33**, 249–274 (2005)