

# STATISTICAL INFERENCE ON THE NUMBER OF CYCLES IN BRAIN NETWORKS

Moo K. Chung<sup>1</sup>, Shih-Gu Huang<sup>1</sup>, Andrey Gritsenko<sup>1</sup>, Li Shen<sup>2</sup>, Hyekyoung Lee<sup>3</sup>

<sup>1</sup>University of Wisconsin, Madison, USA

<sup>2</sup>University of Pennsylvania, Philadelphia, USA

<sup>3</sup>Seoul National University, Seoul, Korea

## ABSTRACT

A cycle in a graph is a subset of a connected component with redundant additional connections. If there are many cycles in a connected component, the connected component is more densely connected. While the number of connected components represents the integration of the brain network, the number of cycles represents how strong the integration is. However, enumerating cycles in the network is not easy and often requires brute force enumerations. In this study, we present a new scalable algorithm for enumerating the number of cycles in the network. We show that the number of cycles is monotonically decreasing with respect to the filtration values during graph filtration. We further develop a new statistical inference framework for determining the significance of the number of cycles. The methods are applied in determining if the number of cycles is a statistically significant heritable network feature in the functional human brain network.

## 1. INTRODUCTION

The modular structure or connected components are one of the fundamental topological features of brain network. Brain networks with higher number of connected components have many disjoint clusters and the transfer of information will likely to be impeded. Modular structures are often studied through the Q-modularity in graph theory [1, 2] and the 0-th Betti number in persistent homology [3, 4, 5]. The 0-th Betti number, the number of connected components, is considered as the first order topological feature and often *not* sensitive enough to discriminate more subtle topological differences.

Persistent homology provides a coherent framework for obtaining higher order topological features beyond modular structures [6, 7]. A graph can be treated as the 1-skeleton of a simplicial complex, where the 0-dimensional hole is the connected component, and the 1-dimensional hole is a cycle. The number of  $k$ -dimensional holes is called the  $k$ -th Betti number and denoted as  $\beta_k$  [3, 8, 9, 10]. In applying persistent homology to a brain network, it is necessary to threshold

edge weights somewhere and make the network into a binary graph. Unfortunately, the choice of threshold affects the topological structure of the network. By performing graph filtration that builds the binary graphs at every possible threshold, we bypass the problem of arbitrary thresholding [5].

Motivated by persistent homology, we will study higher order topological changes of brain networks using cycles. The cycle structure in networks is important for information propagation, redundancy and feedback loops [11]. If a cycle exists in the network, the information can be delivered using two different redundant paths and interpreted as strongly connected. Alternately, it can be viewed as diffusing the spread of information and creating information bottlenecks [12].

While cycles in a network have been widely studied in graph theory, especially in path analysis, they are rarely used in brain network analysis [13]. Existing graph analysis packages such as Brain Connectivity (<http://sites.google.com/site/bctnet>) do not provide any tool related to cycles. Cycles are often computed using the brute-force depth-first search algorithm [12]. Few recent brain network studies on cycles are all based on persistent homology computation, which can be computationally involving [3, 8].

The main contributions of the paper are as follows. 1) We present a new scalable algorithm for computing the number of cycles in the network quickly. 2) We establish the monotonic property of the number of cycles over graph filtration for the first time. The monotonicity is then used in constructing a test statistic for topologically differentiating two networks. 3) The method is applied to the large-scale resting-state twin fMRI study in determining the heritability of the number of cycles.

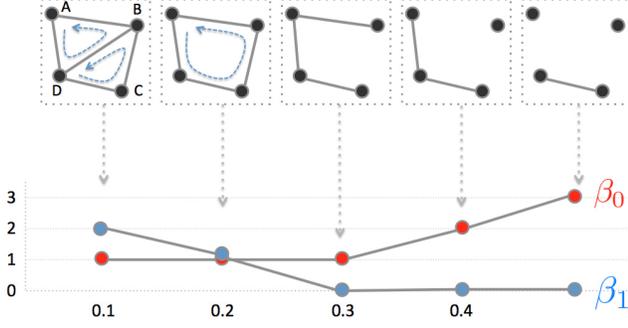
## 2. METHODS

### 2.1. Monotonicity of number of cycles

Given network  $\mathcal{X} = (V, w)$  with node set  $V$  and edge weight  $w = (w_{ij})$  between nodes  $i$  and  $j$ , define binary network  $\mathcal{X}_\epsilon = (V, w_\epsilon)$ , where any edge weight of  $w$  less than or equal to  $\epsilon$  is made into zero while edge weight larger than  $\epsilon$  is made into one. Then we have graph filtration [5, 14]

$$\mathcal{X}_{\epsilon_0} \supset \mathcal{X}_{\epsilon_1} \supset \cdots \supset \mathcal{X}_{\epsilon_q}, \quad (1)$$

Correspondence should be sent to Moo Chung (email: [mkchung@wisc.edu](mailto:mkchung@wisc.edu)). This study was funded by NIH Grants EB022856, EB022574 and Basic Science Research Program through the National Research Foundation (NRF) of Korea (NRF-2016 R1D1A1B 03935463).



**Fig. 1.** Schematic of how Betti numbers change over graph filtration. When each edge is removed from the smallest to the largest edge weights, the number of connected components ( $\beta_0$ ) increases while the number of cycles ( $\beta_1$ ) decreases. At filtration value 0.1, ADB, DCB and ADCB are all cycles but ADCB is the sum of ABD and DCB. So only 2 algebraically independent cycles are counted as  $\beta_1$  in persistent homology [3, 8]. At filtration value 0.2, ADCB is a cycle.

where  $\epsilon_0 < \epsilon_1 < \dots < \epsilon_q$  are  $q$  unique edge weights (Figure 1). Note that  $\mathcal{X}_{-\infty}$  is a complete graph while  $\mathcal{X}_{\infty}$  is the node set  $V$ . By increasing the threshold value, we are thresholding at higher connectivity so more edges are removed.

**Theorem 1** *In a graph, Betti numbers  $\beta_0$  and  $\beta_1$  are monotone over filtration (1).*

*Proof.* Under the graph filtration, the edges are deleted one at a time. Since an edge has only two end points, the deletion of the edge disconnects the graph into at most two. Thus, the number of connected components ( $\beta_0$ ) always increases and the increase is at most by one. The Euler characteristic  $\chi$  of the graph is given by  $\chi = \beta_0 - \beta_1 = p - q$ , where  $p$  and  $q$  are the number of nodes and edges respectively [15]. Thus,

$$\beta_1 = \beta_0 - p + q.$$

Note  $p$  is fixed over the filtration but  $q$  is decreasing by one while  $\beta_0$  increases at most by one. Hence,  $\beta_1$  always decreases and the decrease is at most by one.  $\square$

Figure 1 illustrates how Betti numbers monotonically change for a toy example. Once we compute  $\beta_0$  numbers,  $\beta_1$  number is simply given by  $\beta_0 - p + q$ .

## 2.2. Inference on Betti numbers $\beta_0$ and $\beta_1$

The graph filtration can be quantified using monotone Betti numbers:

$$\begin{aligned} \beta_0(\mathcal{X}_{\epsilon_0}) &\leq \beta_0(\mathcal{X}_{\epsilon_1}) \leq \dots \leq \beta_0(\mathcal{X}_{\epsilon_q}), \\ \beta_1(\mathcal{X}_{\epsilon_0}) &\geq \beta_1(\mathcal{X}_{\epsilon_1}) \geq \dots \geq \beta_1(\mathcal{X}_{\epsilon_q}). \end{aligned}$$

Given two networks  $\mathcal{X}^1 = (V, w^1)$  and  $\mathcal{X}^2 = (V, w^2)$ , Kolmogorov-Smirnov (KS) distance between  $\mathcal{X}^1$  and  $\mathcal{X}^2$  is

defined as

$$D_q(\mathcal{X}^1, \mathcal{X}^2) = \sup_{0 \leq k \leq q} |\beta_i(\mathcal{X}_{\epsilon_k}^1) - \beta_i(\mathcal{X}_{\epsilon_k}^2)|,$$

The distance  $D_q$  is motivated by Kolmogorov-Smirnov (KS) test for determining the equivalence of two cumulative distribution functions [5]. The distribution on  $D_q$  can be computed.

**Theorem 2** [5, 14]

$$\lim_{q \rightarrow \infty} P\left(\frac{D_q}{\sqrt{2q}} \geq d\right) = 2 \sum_{i=1}^{\infty} (-1)^{i-1} e^{-2i^2 d^2}.$$

When  $q \geq 100$ , 10 terms are more than sufficient for convergence. KS-distance method does not assume any statistical distribution on graph features other than that they have to be monotonic. The MATLAB code for computing Betti numbers,  $D_q$  and the corresponding  $p$ -values are available at <http://www.stat.wisc.edu/~mchung/twins>.

## 3. APPLICATION

### 3.1. Resting-state fMRI

We used the resting-state fMRI of genetically confirmed  $m = 131$  monozygotic (MZ) twin pairs and  $n = 77$  same-sex dizygotic (DZ) twin pairs from the Human Connectome Project [16]. fMRI data has undergone spatial and temporal preprocessing including motion and physiological noise removal. Using the resting-state fMRI, we employed the Automated Anatomical Labeling (AAL) template to parcellate the brain volume into 116 regions [17]. The fMRI were then averaged within each brain region for each subject. The averaged fMRI signal in each parcellation is temporally smoothed as follows.

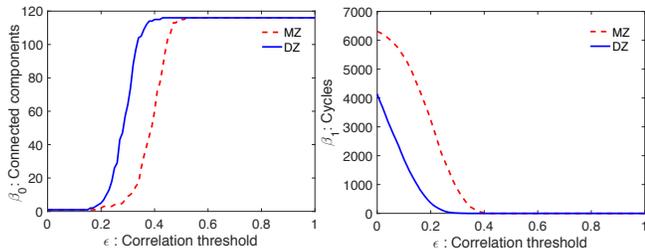
Given fMRI time series at the  $i$ -th parcellation  $\zeta_i(t)$  at time  $t$ , we scaled it to fit to unit interval  $[0, 1]$ . Then subtracted its mean over time  $\int_0^1 \zeta_i(t) dt$ . The resulting scaled and translated time series was represented as

$$\zeta_i(t) = \sum_{l=0}^k d_{li} \psi_l(t), \quad t \in [0, 1],$$

where  $\psi_0(t) = 1$ ,  $\psi_l(t) = \sqrt{2} \cos(l\pi t)$  are cosine basis functions and  $d_{li}$  are coefficients estimated in the least squares fashion [18]. For our study,  $k = 119$  was used such that fMRI were compressed into 10% of the original data size. The resulting Fourier coefficient vector  $\mathbf{d}_i = (d_{0i}, d_{1i}, \dots, d_{ki})$  was used to represent the fMRI in each parcellation as a vector feature in the frequency domain.

### 3.2. Twin correlations

The subject level connectivity matrix  $C = (c_{ij})$  is computed by correlating 120 features in the frequency domain. Between



**Fig. 2.** Betti-plots showing Betti numbers over correlation  $\epsilon$  as filtration. MZ-twins (dashed line) shows more higher correlation connections and cycles compared to DZ-twins (solid line). The maximum gaps between the plots are KS-distances.

$i$ - and  $j$ -th parcellations, the connectivity is measured by correlating  $\mathbf{d}_i$  and  $\mathbf{d}_j$  over 120 features, i.e.,

$$c_{ij} = \text{corr}(\mathbf{d}_i, \mathbf{d}_j).$$

Let  $(c_{ij}^{1k}, c_{ij}^{2k})$  be the connectivity of the  $k$ -th twin pair at the edge level. Let  $\mathbf{c}_{ij}^1 = (c_{ij}^{11}, \dots, c_{ij}^{1m})$  and  $\mathbf{c}_{ij}^2 = (c_{ij}^{21}, \dots, c_{ij}^{2m})$  be the connectivity of all the MZ-twins. Then the MZ twin correlation is computed as

$$c_{ij}^{MZ} = \text{corr}(\mathbf{c}_{ij}^1, \mathbf{c}_{ij}^2),$$

which is the *correlation of correlations*. Similarly we can compute the DZ twin correlation  $c_{ij}^{DZ}$ . However, since there is no preference in the order of twins, twins are randomly permuted in a pairwise fashion in computing the twin correlation and their average is taken as the estimate for twin correlation. Due to high correlation between pairs, only 35 and 52 permutations were required for MZ- and DZ-twins to guarantee the convergence within 4 decimal places in terms of the mean of absolute error of the entries. The resulting group level twin correlations matrices are  $C_{MZ} = (c_{ij}^{MZ})$  and  $C_{DZ} = (c_{ij}^{DZ})$ . The heritability index (HI) is then estimated using the standard ACE-model, which determines the amount of variation due to genetic influence in a population as [19]

$$\text{HI} = 2(C_{MZ} - C_{DZ}).$$

The statistical significance of HI is determined by computing the KS-distance between  $C_{MZ}$  and  $C_{DZ}$  [5].

### 3.3. Results

We used 101 filtration values between 0 and 1 at 0.01 increment. This gives a reasonably accurate estimate of the maximum gap in the  $\beta_i$ -plots between the twins (Figure 2). For  $\beta_0$ -plots, the maximum gap is 84, which gives the  $p$ -value smaller than  $10^{-26}$ . For  $\beta_1$ -plots, the maximum gap is 3627, which gives the  $p$ -value smaller than  $10^{-32}$ . At the same correlation value, MZ-twins are more similar in their connectivity patterns, and thus have more high correlation edges than

DZ-twins. Also MZ-twins have more cycles than DZ-twins. Such topological differences are contributed to heritability.

Figure 3 displays the HI that gives 100% heritability. The most heritable connections include the left frontal gyrus, left and right middle frontal gyri, left superior frontal gyrus, left parahippocampal gyrus, left and right thalami, left and right caudate nuclei among other regions. The left and right caudate nuclei are identified as most heritable hub nodes.

## 4. DISCUSSION

The consistent presence of cycles across different networks is likely to be the network signal. We presented a new property of the monotonicity of the number of cycles over graph filtration. The number of cycles was used to characterize the high order topological changes of the brain network. The existing inference procedure was adapted for the number of cycles.

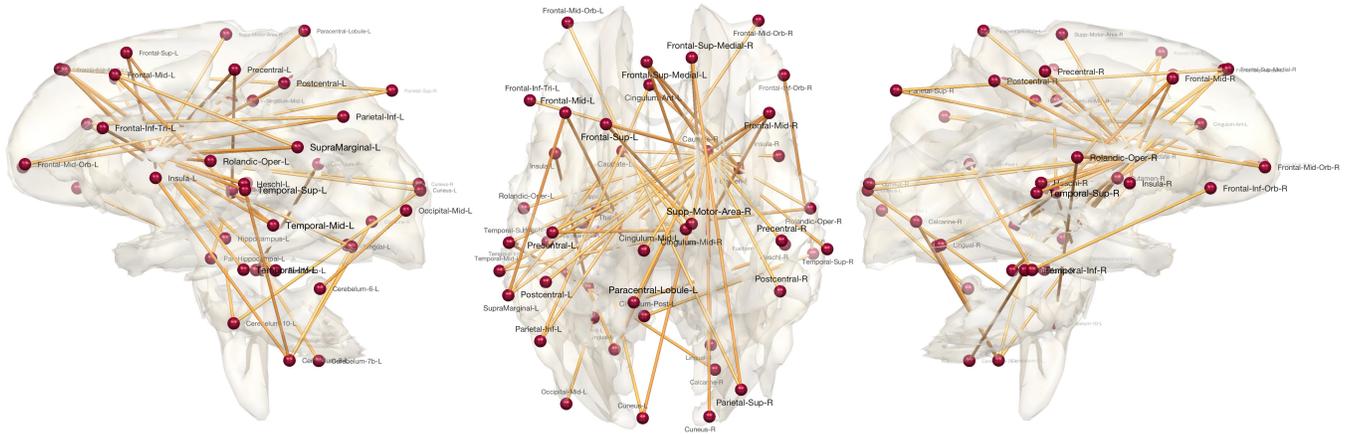
The proposed framework compute the number of cycles in graph filtration without identifying cycles. Theoretically it is possible to identify each cycle that is removed by carefully tracing at what filtration values, cycles disappear. This is left as a future study.

Although we did not provide a comparison against baseline approaches, KS-distance was compared against other topological distances and matrix norms before [14].

**Acknowledgements** We thank Martin Lindquist of Johns Hopkins University, Hernando Ombao of King Abdullah University of Science and Technology, Gregory Kirk of University of Wisconsin-Madison and Alex DiChristofano of Washington University at St. Louise for supports and discussions.

## 5. REFERENCES

- [1] D. Meunier, R. Lambiotte, A. Fornito, K.D. Ersche, and E.T. Bullmore, "Hierarchical modularity in human brain functional networks," *Frontiers in neuroinformatics*, vol. 3:37, 2009.
- [2] M.E.J. Newman, A.L. Barabasi, and D.J. Watts, *The structure and dynamics of networks*, Princeton University Press, 2006.
- [3] H. Lee, Kang H. Chung, M.K., and D.S. Lee, "Hole detection in metabolic connectivity of Alzheimer's disease using k-Laplacian," in *International Conference on Medical Image Computing and Computer-Assisted Intervention, Lecture Notes in Computer Science*, 2014, pp. 297–304.
- [4] G. Carlsson and F. Memoli, "Persistent clustering and a theorem of J. Kleinberg," *arXiv preprint arXiv:0808.2241*, 2008.



**Fig. 3.** The connections with 100% heritability ( $HI \geq 1$ ) are only shown. The left and right caudate nuclei are hub nodes.

- [5] 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," *Information Processing in Medical Imaging (IPMI), Lecture Notes in Computer Science (LNCS)*, vol. 10265, pp. 299–310, 2017.
- [6] H. Edelsbrunner and J. Harer, "Persistent homology - a survey," *Contemporary Mathematics*, vol. 453, pp. 257–282, 2008.
- [7] A.J. Zomorodian and G. Carlsson, "Computing persistent homology," *Discrete and Computational Geometry*, vol. 33, pp. 249–274, 2005.
- [8] H. Lee, M.K. Chung, H. Kang, H. Choi, Y.K. Kim, and D.S. Lee, "Abnormal hole detection in brain connectivity by kernel density of persistence diagram and Hodge Laplacian," in *IEEE International Symposium on Biomedical Imaging (ISBI)*, 2018, pp. 20–23.
- [9] G. Petri, P. Expert, F. Turkheimer, R. Carhart-Harris, D. Nutt, P.J. Hellyer, and F. Vaccarino, "Homological scaffolds of brain functional networks," *Journal of The Royal Society Interface*, vol. 11, pp. 20140873, 2014.
- [10] A.E. Sizemore, C. Giusti, A. Kahn, J.M. Vettel, R.F. Betzel, and D.S. Bassett, "Cliques and cavities in the human connectome," *Journal of computational neuroscience*, vol. 44, pp. 115–145, 2018.
- [11] P.G. Lind, M.C. Gonzalez, and H.J. Herrmann, "Cycles and clustering in bipartite networks," *Physical review E*, vol. 72, pp. 056127, 2005.
- [12] R. Tarjan, "Depth-first search and linear graph algorithms," *SIAM Journal on Computing*, vol. 1, pp. 146–160, 1972.
- [13] O. Sporns, G. Tononi, and GM Edelman, "Theoretical neuroanatomy: relating anatomical and functional connectivity in graphs and cortical connection matrices," *Cerebral Cortex*, vol. 10, pp. 127, 2000.
- [14] M.K. Chung, H. Lee, V. Solo, R.J. Davidson, and S.D. Pollak, "Topological distances between brain networks," *International Workshop on Connectomics in Neuroimaging, Lecture Notes in Computer Science*, pp. 161–170, 2017.
- [15] R.J. Adler, O. Bobrowski, M.S. Borman, E. Subag, and S. Weinberger, "Persistent homology for random fields and complexes," in *Borrowing strength: theory powering applications—a Festschrift for Lawrence D. Brown*, pp. 124–143. Institute of Mathematical Statistics, 2010.
- [16] S.M. Smith, C.F. Beckmann, J. Andersson, E.J. Auerbach, J. Bijsterbosch, and et. al., "Resting-state fMRI in the Human Connectome Project," *NeuroImage*, vol. 80, pp. 144 – 168, 2013.
- [17] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot, "Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the MNI MRI single-subject brain," *NeuroImage*, vol. 15, pp. 273–289, 2002.
- [18] M.K. Chung, N. Adluru, J.E. Lee, M. Lazar, J.E. Lainhart, and A.L. Alexander, "Cosine series representation of 3d curves and its application to white matter fiber bundles in diffusion tensor imaging," *Statistics and Its Interface*, vol. 3, pp. 69–80, 2010.
- [19] D. Falconer and T Mackay, *Introduction to Quantitative Genetics, 4th ed.*, Longman, 1995.