Wasserstein Distance-Preserving Vector Space of Persistent Homology

Tananun Songdechakraiwut[1,2]([✉]), Bryan M. Krause[2], Matthew I. Banks[2],
Kirill V. Nourski[3], and Barry D. Van Veen[2]

[1] Duke University, Durham, USA
`t.song@duke.edu`
[2] University of Wisconsin–Madison, Madison, USA
[3] University of Iowa, Iowa City, USA

Abstract. Analysis of large and dense networks based on topology is
very difficult due to the computational challenges of extracting meaning-
ful topological features from networks. In this paper, we present a com-
putationally tractable approach to topological data analysis of large and
dense networks. The approach utilizes principled theory from persistent
homology and optimal transport to define a novel vector space represen-
tation for topological features. The feature vectors are based on persis-
tence diagrams of connected components and cycles and are computed
very efficiently. The associated vector space preserves the Wasserstein
distance between persistence diagrams and fully leverages the Wasser-
stein stability properties. This vector space representation enables the
application of a rich collection of vector-based models from statistics
and machine learning to topological analyses. The effectiveness of the
proposed representation is demonstrated using support vector machines
to classify measured functional brain networks. Code for the topological
vector space is available at https://github.com/topolearn.

Keywords: Persistent homology · Wasserstein distance · Graph
embedding · Functional brain networks · Consciousness

1 Introduction

Networks are ubiquitous representations for describing complex, highly intercon-
nected systems that capture intricate patterns of relationships between nodes.
[3]. Finding meaningful, computationally tractable characterizations of network
structure is very difficult, especially for large and dense networks with node
degrees ranging over multiple orders of magnitude [5].

Persistent homology [10] is an emerging tool for understanding, characteriz-
ing and quantifying the topology of complex networks [19,21]. Connected com-
ponents and cycles are the most dominant and fundamental topological features
of real networks. For example, many networks naturally organize into modules
or connected components [5]. Similarly, cycle structure is ubiquitous and is often
interpreted in terms of information propagation, redundancy and feedback loops

[14]. Topological features are represented in persistent homology using descriptors called *persistence diagrams* [10]. Effective use of such topological descriptors in machine learning requires a notion of proximity. Wasserstein distance is often used to quantify the distance between persistence diagrams, motivated by its central stability properties [18]. However, the integration of persistence diagrams and Wasserstein distance with standard learning methods from statistics and machine learning has been a challenging open problem due to the differences between Wasserstein distance and standard Euclidean-based metrics [8].

Approaches that embed persistence diagrams into vector spaces [1] or Hilbert spaces [7,13] have recently been proposed to address this challenge. None of the embedding methods proposed thus far preserve Wasserstein distance in the original space of persistence diagrams [6]. Thus, these approaches do not inherit the stability properties of Wasserstein distance.

Recently, it was shown that persistence diagrams are inherently 1-dimensional if the topological features of networks are limited to connected components and cycles, and that the Wasserstein distance between these diagrams has a closed form expression [19]. Consequently, the work in [20] provides a computationally tractable, topological clustering approach for complex networks. However, significant limitations of the result in [19,20] are that it is unclear how this approach can be incorporated with standard Euclidean-based learning methods from statistics and machine learning, and that the approach is limited to evaluating networks with the *identical* number of nodes. There are many opportunities for applications of topological analysis of networks of *different* size, such as studies of the human brain when different subjects are sampled at different resolutions.

In this work, we present a novel *topological vector space* (TopVS) that embeds 1-dimensional persistence diagrams representing connected components and cycles for networks of *different* sizes. Thus, TopVS enables topological machine learning with networks of different sizes and greatly expands the applicability of previous work. Importantly, TopVS preserves the Wasserstein distance in the original space of persistence diagrams. Preservation of the Wasserstein distance ensures the theoretical stability properties of persistence diagrams carry over to the proposed embedding. In addition to the robustness benefits, TopVS also enables the application of a wide variety of Euclidean metric-based learning methods to topological data analysis. Particularly, the utility of TopVS is demonstrated in topology-based classification problems using support vector machines. TopVS is illustrated by classifying measured functional brain networks based on data obtained from subjects with *different* numbers of electrodes. The results show that TopVS performs very well compared to other competing approaches.

2   Wasserstein Distance-Preserving Vector Space

2.1   One-Dimensional Persistence Diagrams

Define a network as an undirected weighted graph $G = (V, \boldsymbol{w})$ with a set of nodes V and a weighted adjacency matrix $\boldsymbol{w} = (w_{ij})$. Define a binary graph G_ϵ with the identical node set V by thresholding the edge weights so that an edge between

nodes i and j exists if $w_{ij} > \epsilon$. The binary graph is viewed as a 1-skeleton [15]. As ϵ increases, more and more edges are removed from the network G. Thus, we have a graph filtration: $G_{\epsilon_0} \supseteq G_{\epsilon_1} \supseteq \cdots \supseteq G_{\epsilon_k}$, where $\epsilon_0 \leq \epsilon_1 \leq \cdots \leq \epsilon_k$ are called filtration values. Persistent homology keeps track of the birth and death of topological features over filtration values ϵ. A topological feature that is born at a filtration b_i and persists up to a filtration d_i, is represented by a point (b_i, d_i) in a 2D plane. A set of all the points $\{(b_i, d_i)\}$ is called *persistence diagram* [10]. In the 1-skeleton, the only non-trivial topological features are connected components and cycles [22]. As ϵ increases, the number of connected components $\beta_0(G_\epsilon)$ and cycles $\beta_1(G_\epsilon)$ are monotonically increasing and decreasing, respectively [19]. Thus, the representation of the connected components and cycles can be simplified to a collection of sorted birth values $B(G) = \{b_i\}_{i=1}^{|V|-1}$ and a collection of sorted death values $D(G) = \{d_i\}_{i=1}^{1+|V|(|V|-3)/2}$, respectively [19]. $B(G)$ comprises edge weights in the *maximum spanning tree* (MST) of G. Once $B(G)$ is identified, $D(G)$ is given as the remaining edge weights that are not in the MST. Thus $B(G)$ and $D(G)$ are computed very efficiently in $O(n \log n)$ operations with n number of edges in networks.

2.2   Closed-Form Wasserstein Distance for Different-Size Networks

The Wasserstein distance between the 1-dimensional persistence diagrams can be obtained using a closed-form solution. Let G_1 and G_2 be two given networks possibly with different node sizes, i.e., their birth and death sets may *differ* in size. Their underlying empirical distributions on the persistence diagrams for connected components are defined in the form of Dirac masses [23]: $f_{G_1,B}(x) := \frac{1}{|B(G_1)|} \sum_{b \in B(G_1)} \delta(x-b)$ and $f_{G_2,B}(x) := \frac{1}{|B(G_2)|} \sum_{b \in B(G_2)} \delta(x-b)$, where $\delta(x-b)$ is a Dirac delta centered at the point b. Then the empirical distribution functions are the integration of $f_{G_1,B}$ and $f_{G_2,B}$ as $F_{G_1,B}(x) = \frac{1}{|B(G_1)|} \sum_{b \in B(G_1)} \mathbb{1}_{b \leq x}$ and $F_{G_2,B}(x) = \frac{1}{|B(G_2)|} \sum_{b \in B(G_2)} \mathbb{1}_{b \leq x}$, where $\mathbb{1}_{b \leq x}$ is an indicator function taking the value 1 if $b \leq x$, and 0 otherwise. A pseudoinverse of $F_{G_1,B}$ is defined as $F_{G_1,B}^{-1}(z) = \inf\{b \in \mathbb{R} \,|\, F_{G_1,B}(b) \geq z\}$, i.e., $F_{G_1,B}^{-1}(z)$ is the smallest b for which $F_{G_1,B}(b) \geq z$. Similarly, we define a pseudoinverse of $F_{G_2,B}$ as $F_{G_2,B}^{-1}(z) = \inf\{b \in \mathbb{R} \,|\, F_{G_2,B}(b) \geq z\}$. Then the empirical Wasserstein distance for connected components has a closed-form solution in terms of these pseudoinverses as

$$W_{p,B}(G_1, G_2) = \left(\int_0^1 |F_{G_1,B}^{-1}(z) - F_{G_2,B}^{-1}(z)|^p \, dz \right)^{1/p}. \tag{1}$$

Similarly, the Wasserstein distance for cycles $W_{p,D}(G_1, G_2)$ is defined in terms of empirical distributions for death sets $D(G_1)$ and $D(G_2)$.

The empirical Wasserstein distances $W_{p,B}$ and $W_{p,D}$ are approximated by computing the Lebesgue integration in (1) numerically as follows. Let $\widehat{B}(G_1) = \{F_{G_1,B}^{-1}(1/m), F_{G_1,B}^{-1}(2/m), ..., F_{G_1,B}^{-1}(m/m)\}$ and $\widehat{D}(G_1) = \{F_{G_1,D}^{-1}(1/n), ..., F_{G_1,D}^{-1}(n/n)\}$ be pseudoinverses for network G_1 sampled with partitions of equal intervals. Let $\widehat{B}(G_2)$ and $\widehat{D}(G_2)$ be sampled pseudoinverses

for network G_2 with the same partitions of m and n, respectively. Then the approximated Wasserstein distances are given by

$$\widehat{W}_{p,B}(G_1, G_2) = \Big(\frac{1}{m^p} \sum_{k=1}^{m} \big| F^{-1}_{G_1,B}(k/m) - F^{-1}_{G_2,B}(k/m) \big|^p \Big)^{1/p}, \qquad (2)$$

$$\widehat{W}_{p,D}(G_1, G_2) = \Big(\frac{1}{n^p} \sum_{k=1}^{n} \big| F^{-1}_{G_1,D}(k/n) - F^{-1}_{G_2,D}(k/n) \big|^p \Big)^{1/p}. \qquad (3)$$

For a special case when networks G_1 and G_2 have the same number of nodes, i.e., $|B(G_1)| = |B(G_2)|$ and $|D(G_1)| = |D(G_2)|$, then exact computation of the Wasserstein distance is achieved using those birth and death sets, and setting m to the cardinality of the birth sets and n to that of the death sets.

2.3 Vector Representation of Persistence Diagrams

A collection of 1-dimensional persistence diagrams together with the Wasserstein distance is a metric space. 1-dimensional persistence diagrams can be embedded into a vector space that preserves the Wasserstein metric on the original space of persistence diagrams as follows. Let $G_1, G_2, ..., G_N$ be N observed networks possibly with different node sizes. Let $F^{-1}_{G_i,B}$ be a pseudoinverse of network G_i. The vector representation of a persistence diagram for connected components in network G_i is defined as a vector of the pseudoinverse sampled at $1/m, 2/m, ..., m/m$, i.e., $\boldsymbol{v}_{B,i} := \big(F^{-1}_{G_i,B}(1/m), F^{-1}_{G_i,B}(2/m), ..., F^{-1}_{G_i,B}(m/m) \big)^{\top}$. A collection of these vectors $M_B = \{\boldsymbol{v}_{B,i}\}_{i=1}^{N}$ with the p-norm $|| \cdot ||_p$ induces the p-norm metric $d_{p,B}$ given by $d_{p,B}(\boldsymbol{v}_{B,i}, \boldsymbol{v}_{B,j}) = ||\boldsymbol{v}_{B,i} - \boldsymbol{v}_{B,j}||_p = m\widehat{W}_{p,B}$. Thus, for $p = 1$ the proposed vector space describes Manhattan distance, $p = 2$ Euclidean distance, and $p \to \infty$ the maximum metric, which in turn correspond to the earth mover's distance (W_1), 2-Wasserstein distance (W_2), and the bottleneck distance (W_∞), respectively, in the original space of persistence diagrams. Similarly, we can define a vector space of persistence diagrams for cycles $M_D = \{\boldsymbol{v}_{D,i}\}_{i=1}^{N}$ with the p-norm metric $d_{p,D}$. The normed vector space $(M_B, d_{p,B})$ describes topological space of connected components in networks, while $(M_D, d_{p,D})$ describes topological space of cycles in networks.

The topology of a network viewed as a 1-skeleton is *completely* characterized by connected components and cycles. Thus, we can fully describe the network topology using both M_B and M_D as follows. Let $M_B \times M_D = \{(\boldsymbol{v}_{B,i}, \boldsymbol{v}_{D,i}) \,|\, \boldsymbol{v}_{B,i} \in M_B, \boldsymbol{v}_{D,i} \in M_D\}$ be the Cartesian product between M_B and M_D so the vectors in $M_B \times M_D$ are the concatenations of $\boldsymbol{v}_{B,i}$ and $\boldsymbol{v}_{D,i}$. For this product space to represent meaningful topology of network G_i, the vectors $\boldsymbol{v}_{B,i}$ and $\boldsymbol{v}_{D,i}$ must be a network decomposition, as discussed in Sect. 2.1. Thus $\boldsymbol{v}_{B,i}$ and $\boldsymbol{v}_{D,i}$ are constructed by sampling their psudoinverses with $m = \mathcal{V} - 1$ and $n = 1 + \frac{\mathcal{V}(\mathcal{V}-3)}{2}$, respectively, where $\mathcal{V}$ is a free parameter indicating a reference network size. The metrics $d_{p,B}$ and $d_{p,D}$ can be put together to form a

p-product metric $d_{p,\times}$ on $M_B \times M_D$ as [9]

$$d_{p,\times}\big((\boldsymbol{v}_{B,i}, \boldsymbol{v}_{D,i}), (\boldsymbol{v}_{B,j}, \boldsymbol{v}_{D,j})\big) = \big([d_{p,B}(\boldsymbol{v}_{B,i}, \boldsymbol{v}_{B,j})]^p + [d_{p,D}(\boldsymbol{v}_{D,i}, \boldsymbol{v}_{D,j})]^p\big)^{1/p}$$
$$= \big([m\widehat{W}_{p,B}]^p + [n\widehat{W}_{p,D}]^p\big)^{1/p}, \tag{4}$$

where $(\boldsymbol{v}_{B,i}, \boldsymbol{v}_{D,i}), (\boldsymbol{v}_{B,j}, \boldsymbol{v}_{D,j}) \in M_B \times M_D$, $m = \mathcal{V} - 1$ and $n = 1 + \frac{\mathcal{V}(\mathcal{V}-3)}{2}$. Thus, $d_{p,\times}$ is a weighted combination of p-Wasserstein distances, and is simply the p-norm metric between vectors constructed by concatenating $\boldsymbol{v}_{B,i}$ and $\boldsymbol{v}_{D,i}$. The normed vector space $(M_B \times M_D, d_{p,\times})$ is termed *topological vector space* (TopVS). Note the form of $d_{p,\times}$ given in (4) results in an unnormalized mass after multiplying m and n by their reciprocals given in (2) and (3). This unnormalized variant of Wasserstein distance is widely used in both theory [8,18] and application [7,19,21] of persistent homology. A direct consequence of the equality given in (4) is that the mean of persistence diagrams under the approximated Wasserstein distance is equivalent to the sample mean vector in TopVS. In addition, the proposed vector representation is highly interpretable because persistence diagrams can be easily reconstructed from vectors by separating sorted births and deaths.

For a special case in which networks $G_1, G_2, ..., G_N$ have the same number of nodes, the vectors $\boldsymbol{v}_{B,i}$ and $\boldsymbol{v}_{D,i}$ are simply the original birth set $B(G_i)$ and death set $D(G_i)$, respectively, and the p-norm metric $d_{p,\times}$ is expressed in terms of exact Wasserstein distances as $d_{p,\times} = ([mW_{p,B}]^p + [nW_{p,D}]^p)^{1/p}$.

3   Application to Functional Brain Networks

Dataset. We evaluate our method using functional brain networks from the anesthesia study reported by [2]. The brain networks are based on alpha band (8–12 Hz) weighted phase lag index applied to 10-second segments of resting state intracranial electroencephalography recordings. These recordings were made from eleven neurosurgical patients during administration of increasing doses of the general anesthetic propofol just prior to surgery. Each segment is labeled as one of the three arousal states: pre-drug *wake* (W), *sedated* but responsive to command (S), or *unresponsive* (U). The number of brain networks belonging to each subject varies from 71 to 119, resulting in the total of 977 networks from all the subjects. The network size varies from 89 to 199 nodes across subjects.

Classification Performance Evaluation. We are interested in whether candidate methods 1) can differentiate arousal states within individual subjects, and 2) generalize their learned knowledge to unknown subjects afterwards. As a result, we consider two different nested cross validation (CV) tasks as follows.

1. For the first task, we classify a collection of brain networks belonging to each subject separately. Specifically, we apply a nested CV comprising an outer loop of stratified 2-fold CV and an inner loop of stratified 3-fold CV. Since we may get a different split of data folds each time, we perform the nested CV

for 100 trials and report an average accuracy score and standard deviation for each subject. We also average these individual accuracy scores across subjects (11×100 scores) to obtain an overall accuracy.

2. For the second task, we use a different nested CV comprising both outer and inner loops with a leave-one-subject-out scheme. That is, a classifier is trained using all but one test subject. The inner loop is used to determine optimal hyperparameters, while the outer loop is used to assess generalization capacity of the candidate methods to unknown subjects in the population.

Method Comparison. Brain networks are used to compare the classification performance of the proposed TopVS relative to that of five state-of-the-art kernel methods and two well-established graph neural network methods. Three of these kernel methods are based on conventional 2-dimensional persistence diagrams for connected components and cycles: the *persistence image* (PI) vectorization [1], the *sliced Wasserstein kernel* (SWK) [7] and the *persistence weighted Gaussian kernel* (PWGK) [13]. The other two kernel methods are based on graph kernels: the *propagation kernel* (Prop) [16] and the *GraphHopper kernel* (GHK) [11]. The PI method embeds persistence diagrams into a vector space in which classification is performed using linear support vector machines (SVMs). The non-linear SWK, PWGK, Prop and GHK methods are combined with SVMs to perform classification. While nearly any classifier may be used with TopVS, here we illustrate results using the SVM with the linear kernel, which maximizes Wasserstein distance-based margin. When the TopVS method is applied to different-size networks, we upsample birth and death sets of smaller networks to match that of the largest network in size. Hyperparameters are tuned using grid search. SVMs have a regularization parameter $\mathcal{C} = \{0.01, 1, 100\}$. Thus, a grid search trains TopVS and PI methods with each $C \in \mathcal{C}$. The SWK and WGK methods have a bandwidth parameter $\Sigma = \{0.1, 1, 10\}$, and thus grid search trains both methods with each pair $(C, \sigma) \in \mathcal{C} \times \Sigma$. The Prop method has a maximum number of propagation iterations $T_{max} = \{1, 5, 10\}$, and thus is trained with each pair $(C, t_{max}) \in \mathcal{C} \times T_{max}$. GHK method uses the RBF kernel with a parameter $\Gamma = \{0.1, 1, 10\}$ between node attributes, and thus is trained with each pair $(C, \gamma) \in \mathcal{C} \times \Gamma$.

In addition, we also evaluate two well-established graph neural network methods including *graph convolutional networks* (GCN) [12] and *graph isomorphism network* (GIN) [25]. GCN and GIN are based on configurations and choices of hyperparameter values used in [25] as follows. Five graph neural network layers are applied, and the Adam optimizer with initial learning rate and weight decay of 0.01 are employed. We tune the following hyperparameters: the number of hidden units in $\{16, 32\}$, the batch size in $\{32, 128\}$ and the dropout ratio in $\{0, 0.5\}$. The number of epochs is set to 100 to train both methods.

Results. Results for the first task are summarized in Fig. 1, in which classification accuracy for individual subjects is shown. There is variability in individual subject performance because a different subject's network has a different number of electrodes, different electrode locations and different effective signal to
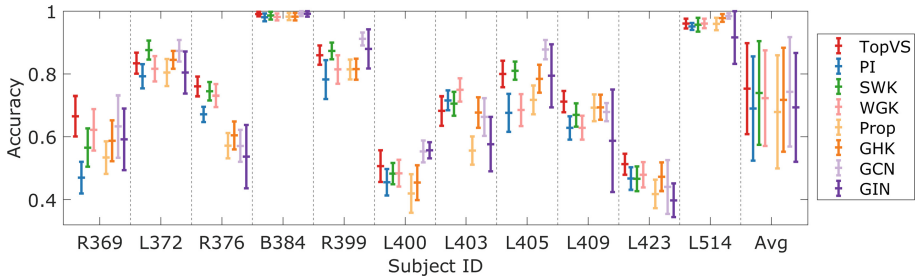
Fig. 1. Accuracy classifying brain networks within individual subjects. The last column displays the average accuracy obtained across all subjects. The center markers and bars depict the means and standard deviations obtained over 100 different trials.

noise ratio. So we expect these subjects to exhibit a diverse set of topological features across spatial resolutions. In most subjects all methods perform relatively well. The consistently poorer performance of PI, Prop and GIN is evident in the lower overall performance. On the other hand, our TopVS method is consistently among the best performing classifiers, resulting in the higher overall performance. For classification accuracy across subjects from the second task, we have 0.65 ± 0.21 for TopVS, 0.58 ± 0.22 for PI, 0.57 ± 0.20 for SWK, 0.60 ± 0.21 for WGK, 0.36 ± 0.12 for Prop, 0.43 ± 0.14 for GHK, 0.53 ± 0.20 for GCN and 0.48 ± 0.19 for GIN. TopVS is also among the best methods for classifying across subjects, while the performance of all the graph neural networks and graph kernels is significantly weaker. These results suggest that the use of computationally demanding and complex classification methods, such as GCN and GIN, does not result in significant increase in generalizability when classifying brain networks.

In addition, we compute confusion matrices to gain insights into the across-subject predictions for the second task, as displayed in Fig. 2. The persistent homology based methods, including TopVS, PI, SWK and WGK, are generally effective for separating unresponsive (U) from the other two states, and the majority of classification errors are associated with the differentiation between wake (W) and sedated (S) states. Prior work [2] demonstrated that wake and sedated brains are expected to have a great deal of similarity in comparison to the less similar unresponsive brains. However, the work in [2] performed the analysis on each individual subject separately while the results presented here are based on the analysis across subjects. Thus, not only the results here are consistent with the previous work [2] but also suggest that such biological expectation carries over to brains across subjects and that topology based methods can potentially derive biomarkers of changes in arousal states in the population, which underlie transitions into and out of consciousness, informing our understanding of the neural correlates of consciousness in clinical settings. TopVS shows clear advantages over all other topological baseline methods for differentiating wake and sedated states, suggesting that the proposed vector representation is an effective choice for representing subtle topological structure in brain networks.
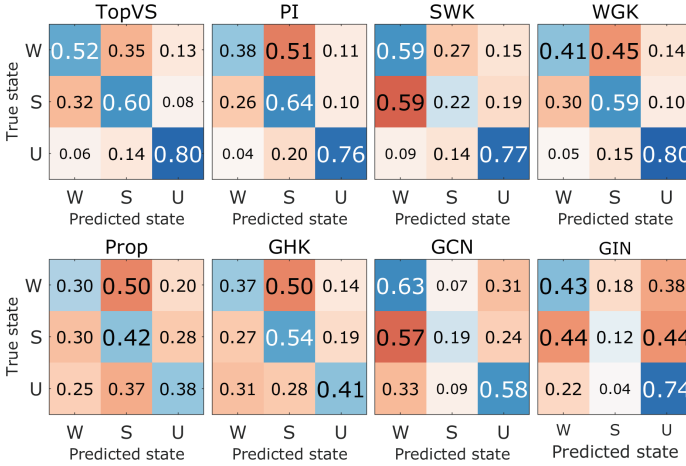
Fig. 2. Confusion matrices illustrating method performance for classifying across subjects. The numbers represent the fraction of brain networks in the test subjects being predicted as one of the three states: wake (W), sedated (S), and unresponsive (U). The confusion matrices are normalized with the entries in each row summing to 1.

Runtime Experiment. The kernel candidate methods are evaluated for a runtime experiment based on Intel Core i7 CPU with 16 GB of RAM. Figure 3 displays the runtime vs input size plot. The result clearly shows that all three persistent homology based kernels (PI, SWK and WGK) are limited to dense networks with a few hundred nodes, representing the current scaling limit of persistent homology embedding methods. On the other hand, TopVS is able to compute a kernel between 2000-node networks each with approx. two million edges in about one second. The computational practicality of TopVS extends its applicability to the large-scale analyses of brain networks that cannot be analyzed using prior methods based on conventional 2-dimensional persistence diagrams. Note that the time complexity of Prop is linear while TopVS has the slightly higher complexity as linearithmic. While Prop is the most efficient among all the methods, it has the lowest average accuracy when classifying the brain network data.

Potential Impact and Limitation. An open problem in neuroscience is identifying an algorithm that reliably extracts a patient's level of consciousness from passively recorded brain signals (i.e., biomarkers) and is robust to inter-patient variability, including where the signals are recorded in the brain. Conveniently, the anesthesia dataset is labeled according to consciousness state, and electrode placement (node location) was dictated solely by clinical considerations and thus varied across patients. Importantly, the relatively robust performance across patients suggests there are reliable topological signatures of consciousness captured by TopVS. The distinction between Wake and Sedated states involves relatively nuanced differences in connectivity, yet TopVS exploits the subtle differences in topology that differentiate these states better than the com-
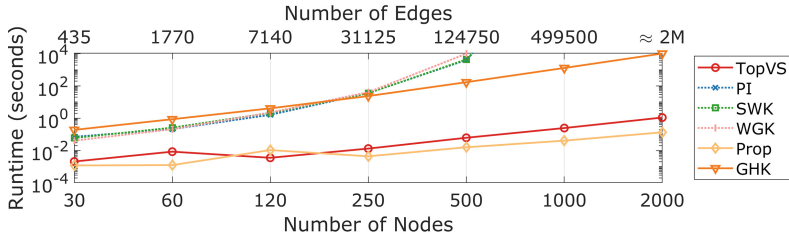
Fig. 3. Runtime experiment. We measured the runtime as the average amount of time each algorithm takes to compute its kernel between two complete graphs starting from edge weights as a given input. The runtime is plotted with respect to network size in terms of both the number of nodes and edges.

peting methods. Our results suggest that the neural correlates of consciousness can be captured in measurements of brain network topology, a longstanding problem of great significance. Additionally, TopVS is a principled framework that connects persistent homology theory with practical applications. Our versatile vector representation can be used with various vector-based statistical and machine learning models, expanding the potential for analyzing extensive and intricate networks beyond the scope of this paper. While TopVS is limited to representing connected components and cycles, assessment of higher-order topological features beyond cycles is of limited value due to their relative rarity and interpretive challenges, and consequent minimal discriminitive power [4,17,24].

References

1. Adams, H., et al.: Persistence images: a stable vector representation of persistent homology. JMLR **18** (2017)
2. Banks, M.I., et al.: Cortical functional connectivity indexes arousal state during sleep and anesthesia. NeuroImage **211**, 116627 (2020)
3. Barrat, A., Barthelemy, M., Pastor-Satorras, R., Vespignani, A.: The architecture of complex weighted networks. Proc. Natl. Acad. Sci. **101**(11), 3747–3752 (2004)
4. Biagetti, M., Cole, A., Shiu, G.: The persistence of large scale structures. Part I. Primordial non-Gaussianity. J. Cosmol. Astropart. Phys. **2021**(04), 061 (2021)
5. Bullmore, E., Sporns, O.: Complex brain networks: graph theoretical analysis of structural and functional systems. Nat. Rev. Neurosci. **10**(3), 186–198 (2009)
6. Carrière, M., Bauer, U.: On the metric distortion of embedding persistence diagrams into separable Hilbert spaces. In: Symposium on Computational Geometry (2019)
7. Carriere, M., Cuturi, M., Oudot, S.: Sliced Wasserstein kernel for persistence diagrams. In: International Conference on Machine Learning (2017)
8. Cohen-Steiner, D., Edelsbrunner, H., Harer, J., Mileyko, Y.: Lipschitz functions have LP-stable persistence. Found. Comput. Math. **10**, 127–139 (2010)
9. Deza, M.M., Deza, E.: Encyclopedia of Distances. Springer, Heidelberg (2009)
10. Edelsbrunner, H., Harer, J.L.: Computational Topology: An Introduction. American Mathematical Society (2022)

11. Feragen, A., Kasenburg, N., Petersen, J., de Bruijne, M., Borgwardt, K.M.: Scalable kernels for graphs with continuous attributes. In: NIPS (2013)
12. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. In: International Conference on Learning Representations (2017)
13. Kusano, G., Hiraoka, Y., Fukumizu, K.: Persistence weighted Gaussian kernel for topological data analysis. In: International Conference on Machine Learning (2016)
14. Kwon, Y.K., Cho, K.H.: Analysis of feedback loops and robustness in network evolution based on Boolean models. BMC Bioinform. **8**(1), 1–9 (2007)
15. Munkres, J.R.: Elements of Algebraic Topology. CRC Press, Boca Raton (2018)
16. Neumann, M., Garnett, R., Bauckhage, C., Kersting, K.: Propagation kernels: efficient graph kernels from propagated information. Mach. Learn. **102**(2), 209–245 (2016)
17. Sizemore, A.E., Giusti, C., Kahn, A., Vettel, J.M., Betzel, R.F., Bassett, D.S.: Cliques and cavities in the human connectome. J. Comput. Neurosci. **44**, 115–145 (2018)
18. Skraba, P., Turner, K.: Wasserstein stability for persistence diagrams. arXiv preprint arXiv:2006.16824 (2020)
19. Songdechakraiwut, T., Chung, M.K.: Topological learning for brain networks. Ann. Appl. Stat. **17**(1), 403–433 (2023)
20. Songdechakraiwut, T., Krause, B.M., Banks, M.I., Nourski, K.V., Veen, B.D.V.: Fast topological clustering with Wasserstein distance. In: International Conference on Learning Representations (ICLR) (2022)
21. Songdechakraiwut, T., Shen, L., Chung, M.: Topological learning and its application to multimodal brain network integration. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12902, pp. 166–176. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87196-3_16
22. Sunada, T.: Homology groups of graphs. In: Topological Crystallography. Surveys and Tutorials in the Applied Mathematical Sciences, vol. 6, pp. 37–51. Springer, Tokyo (2013). https://doi.org/10.1007/978-4-431-54177-6_4
23. Turner, K., Mileyko, Y., Mukherjee, S., Harer, J.: Fréchet means for distributions of persistence diagrams. Discrete Comput. Geom. **52**(1), 44–70 (2014)
24. Xia, K., Wei, G.W.: Persistent homology analysis of protein structure, flexibility, and folding. Int. J. Numer. Methods Biomed. Eng. **30**(8), 814–844 (2014)
25. Xu, K., Hu, W., Leskovec, J., Jegelka, S.: How powerful are graph neural networks? In: International Conference on Learning Representations (2019)