Persistent Homology of Global Functional Connectivity for Depression Classification

Amy Hu, Guy Wilson

***Abstract*—*Objective*: Probing the connectivity of brain regions has yielded enormous benefits in characterizing and modeling disease states as well as understanding the dynamics of functional networks underlying a variety of cognitive tasks. In this paper, we discuss a tool known as persistent homology, providing background and a discussion on numerical stability as well as an implementation of the technique for capturing topological features of resting state functional connectivity (RSFC) at multiple scales, thereby avoiding arbitrary thresholding. *Methods*: We then apply the technique to study differences in global RSFC between healthy controls and individuals with clinical depression and create features for classification through multidimensional scaling of pairwise distances, deploying and evaluating both persistence diagrams and landscapes as topological summaries. *Results*: The results reveal lowered global RSFC in depressed participants and show that that persistent homology can provide suitable features for classification through multidimensional scaling of pairwise distances. We also show that persistent homology outperforms a variety of other graph metrics in classification yet overall performance is highest when combined with other features, suggesting that information capture by persistent homology is largely complementary to common graph-theoretic tools. *Conclusions*: In sum, these results corroborate findings of resting state network differences in clinically depressed individuals and indicate that persistent homology, through multidimensional scaling, can be effectively deployed in classification. *Significance*: The analysis highlights the potential of persistent homology in studying depression and suggest the importance of interaction effects for improving future topological machine learning models in biomedical classification tasks.**

***Keywords*—*Brain Connectivity, Graph Theory, Persistent Homology, Machine Learning*.**

I. Introduction

Coordinated brain networks underlie a variety of behaviors and can be dynamically controlled in order to flexibly adapt to changes in setting or task [1] [2] [3]. By studying statistical correlations, one can identify patterns of functional connectivity that serves as a proxy for probing communication between different brain regions without resorting to invasive methods. Understanding brain functional networks can yield insight into not only normal behavior but also psychiatric disorders, where their organization and activity often differ strongly from that of healthy individual [4] [5] [6].

One such case are differences in resting state functional connectivity (RSFC) between patients diagnosed with major depressive disorder (MDD) and healthy individuals, having been noted in numerous studies [7] [5] [8], with depressive states being correlated with abnormal network properties during resting states versus controls. In particular, MDD is associated with changes in the default mode network (DMN), a group of intercommunicating regions associated with self-rumination, mind-wandering, and thinking about the past or future [5], and other local networks as well as more global irregularities [7] [8]. In this study, we applied a new technique from algebraic topology, called persistent homology, to study MDD as well to evaluate its performance against other graph metrics commonly leveraged in human disease classification tasks.

A. Hu and G. Wilson are with the University of California, Berkeley (correspondence email: gwilson18@berkeley.edu).

II. Persistent Homology

In recent years, computational topology has emerged at the forefront of data science with applications in sensor networks [9], neuroscience [6], cell biology [10], and numerous other areas. Topological data analysis (TDA) views data points as approximations of underlying manifolds (spaces with locally Euclidean structure e.g. a sphere) whose topology is of interest, either for dimensionality reduction, insight into nonlinear relationships of different variables, or discriminative purposes. In neuroscience, TDA has been shown to be an effective tool for probing differences in brain connectivity [6] [2] and, more generally, studying network structure [11]. Topological properties of manifolds are preserved under smooth deformations and hence yield robust information for real world data containing noise. One such tool in this collection is *persistent homology*, a method for analyzing the lifespan of k-dimensional holes in data called homology classes.

A. Simplicial Complexes

In order to study the topology of a dataset, it must be converted into an object with an appropriate structure for analysis, specifically a simplicial complex. A k-simplex is the set of points in $\mathbb{R}^{k+1}$ such that

$$\{(x_1, ..., x_{k+1}) | x_1 + x_2 + ... + x_{k+1} = 1, x_i \geq 0\}$$

.

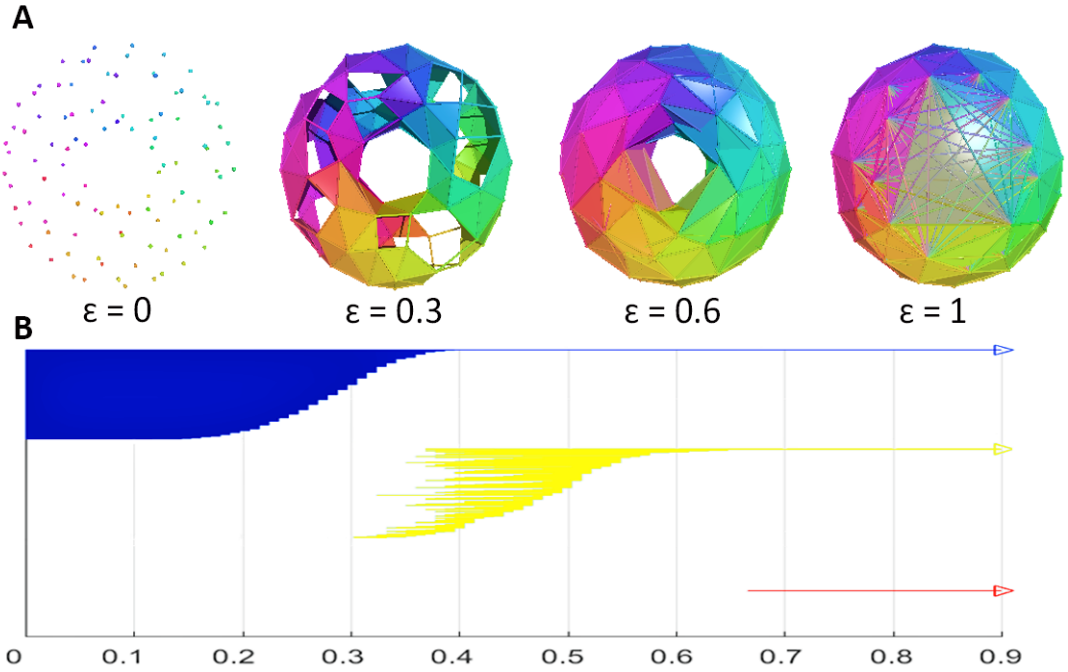

Fig. 1: A Vietoris-Rips filtration of a torus parametrized by ϵ. Each simplicial complex includes into its neighbor, resulting in a map between the homology groups that allows us to track the birth and death of k-dimensional holes (A). Barcode summaries for the 0,1, and 2-dimensional features. Longer barcodes represent more persistent topological features. Persistent homology captures the true topology of the underlying manifold; we observe one stable connected component (blue), two stable classes of 1-dimensional cycles (yellow) corresponding to longitudinal and latitudinal paths along the torus, and a void (red) on the interior (B).

For example, a 0-dimensional simplex is a point, 1-dimensional a line, 2-dimensional a triangle, 3-dimensional a tetrahedron and so on. A simplicial complex can be understood as a collection of these points, lines, triangles, and higher dimensional objects. More formally, a simplicial complex is defined as:

Given a set of simplices, $V = \{\sigma_1, \sigma_2, ..., \sigma_k\}$, and their faces, $\{\tau_1, \tau_2, ..., \tau_{k-1}\}$, we say $\tau \leq \sigma$ if the face belongs to that simplex. Then we say K is a *simplicial complex* if:
(i) $\sigma \in K$ and $\tau \leq \sigma$ implies that $\tau \in K$
(ii) $\sigma, \tau \in K$ implies that $\sigma \cap v \leq \sigma, \tau$.

There are many different methods used in uncovering the simplicial complexes in the dataset. The one used in this paper is the Vietoris-Rips complex (also called Rips complex, denoted as R). Given a set of points $\{x_1, ..., x_n\}$ in a space, X, then $\forall k \in \mathbb{N}$, if $\exists\ x_{i_1}, ...x_{i_k}$ such that $d(x_{i_a}, x_{i_b}) \leq \epsilon\ \forall i_a, i_b \in \{1, ..., k\}$, then the resultant (k-1) simplex is added to our Vietoris-Rips complex.

Some other common methods include the Cech complex, Delaunay complex, and witness complex [12]. In particular, the Cech complex is shown to capture the true topology of an underlying manifold assuming the sampled points are appropriately picked [13]. The Cech complex is formed by adding a k-simplex when there is a point of intersection with k open ϵ-balls. By contrast, the Vietoris-Rips construction adds a k-simplex wherever $k-1$ ϵ-balls have a pairwise intersection. However, the Rips complex is a good approximation of the Cech complex according to the following inclusion [13]:

$$R(\epsilon) \subseteq Cech(\sqrt{2}\epsilon) \subseteq R(\sqrt{2}\epsilon)$$

The Cech complex requires us to store all the simplices whereas the Rips complex only requires pairwise distances i.e. the entire structure can be deduced from the 1-skeleton. Thus, for the sake of computational simplicity, researchers often use the Rips complex.

B. Filtrations and Persistence

The Rips filtration generates simplicial complexes that are highly dependent on the ϵ distance chosen. In order to differentiate the complexes formed at each ϵ distance, a filtration is used. A filtration of a simplicial complex, A, is a sequence of subcomplexes $A_i \subseteq A_{i+1}$.

Furthermore, given complexes A and B with a simplicial map $f : A \to B$, there is an associated mapping between their chain complexes, f_k :

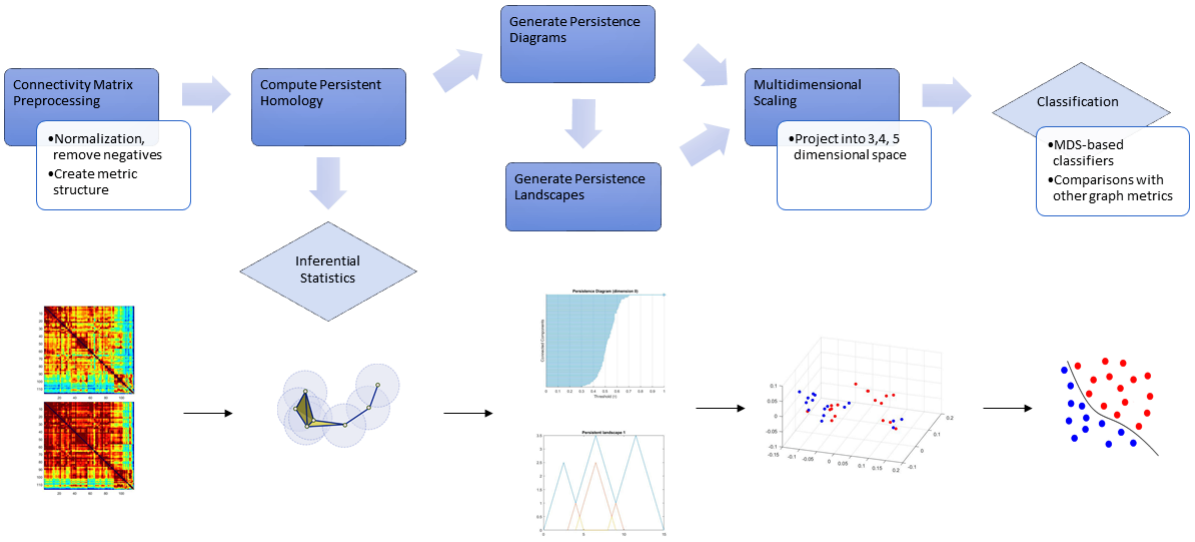

Fig. 2: Following preprocessing, Vietoris-Rips filtrations are built over the ROI connectivity data. The output can be used for statistical analysis of group-level differences and also as input for classifiers. Multidimensional scaling of pairwise bottleneck distances is used to project persistence diagrams and landscapes into a suitable metric space.

$C_k(A) \to C_k(B)$ (for those more acquainted with linear algebra, these structures can be upgraded to vector spaces and the mappings between them linear transformations). It can be shown that this chain map induces homomorphisms between the homology groups of A and B containing our higher-dimensional cycles of interest. When $A \subseteq B$, we can take f to be the natural embedding of A into B. When working with filtrations of spaces, resulting in a series of nested simplicial complexes $A_i \subseteq A_{i+1}$, we have maps:

$$A_1 \xhookrightarrow{f_1} A_2 \xhookrightarrow{f_2} A_3 ...$$
$$H_k(A_1) \xrightarrow{\varphi_{1,k}} H_k(A_2) \xrightarrow{\varphi_{2,k}} H_k(A_3) ...$$

One notion we can now try to develop is the "persistence" of an equivalence class as it moves through this sequence of homology groups. In other words, for a given $x \in H_k(A_i)$ (where k is fixed) we want to understand the following:

1) When x appears i.e. for what complex A_i is $x \in H_k(A_i)$ but $\forall i' < i,\ x \notin H_k(A_{i'})$
2) When x disappears i.e. for what complex A_i is $x \in H_k(A_i)$ but $\forall i' > i,\ x = 0 \in H_k(A_{i'})$

Technically, the x referenced above should be viewed as a collection of images (through homology maps) of an initial equivalence class. In a geometric context however, this notational leniency lends itself to a better intuition for what this information tracks: the time period during which a topological feature is present.

A commonly used visualization of the time period during which a topological feature is present is the barcode. The birth time is the start of a new bar and the death time is the end of the bar. This interval corresponds to a n-dimensional hole opening in the complex at the birth and filling in at the death. A barcode that does not have a death (i.e. interval goes to infinity) would correlate to a perpetually open hole. The longer barcodes indicate significant structures whereas the shorter barcodes are less likely to be so as they disappear much more quickly, indicating weakly characterized topological attributes in the data.

To make this data more useful in computation, persistence diagrams encode the same data in a better descriptor. Plotting the births and deaths along the horizontal and vertical axes respectively, it is easy to see when the persistent homology classes are born and die. The most complex descriptor used is the *persistence landscape*. In this approach, the births and deaths of the homology classes are transformed into continuous, piecewise functions; the persistence landscape is a sequence of these functions. These functions introduces statistical properties that were previously not available with barcodes and persistence diagrams [14].

C. Numerical Stability

While persistence diagrams are capable of encoding topological information about data, they must be equipped with a suitable metric structure in order for common machine learning methods to be deployed

[15]. In order to accomplish this, one can provide the space of persistence diagrams with the bottleneck distance as a metric, in turn imbuing Lipschitz continuity [16]. The bottleneck distance between sets X and Y, with elements in $\mathbb{R}^n$, is given by:

$$W_\infty(X,Y) = \min_{\varphi:X\to Y} \max_{x\in X} ||x - \varphi(x)||$$

In the definition, φ is a bijection between the sets and hence the bottleneck distance is found by maximizing the distance between a given point and all elements of Y, storing that value, and then taking the minimum of such a collection of distances.

III. Methods

RSFC data were obtained from a dataset [7] comprising 17 controls and 16 individuals diagnosed with major depressive disorder (MDD). For each subject, an associated connectivity matrix tracks pairwise ROI correlations, generated by parcellation of the brain according to the Automated Anatomic Labeling (AAL) template [17] into 116 regions. Correlations were obtained from eight minute scans in a 3-Tesla scanner, during which subjects were not prompted to think about anything specifically, and z-scored using Fisher's z-transform. Five MDD participants were taking anti-depressants at the time of the initial data gathering, one had a co-morbid diagnosis of PTSD, another had co-morbid diagnosis of schizophrenia, and seven had co-morbid diagnoses of anxiety and/or social phobia [7].

Persistent homology calculations were performed through the JavaPlex library [18] on MATLAB (the Mathworks, Natick, MA, USA). Persistence landscapes were computed through the Persistence Landscapes Toolbox [19]. Neural networks were constructed using the TensorFlow machine learning library [20]. Other classifiers were implemented using the MATLAB Statistics and Machine Learning Toolbox (11.1). Graph measures were computed using the Brain Connectivity Toolbox [21].

A. Pipeline

Correlation matrices were rescaled between -1 and 1 across all subjects while retaining relative connectivity differences and, for problem simplification, we set negative values to zero as performed in a previous study [6]. Distances between ROIs were defined using the transformation:

$$d(x_i, x_j) = 1 - corr(x_i, x_j)$$

which captures topological features of the network by associating tightly correlated brain regions in the resulting space [6].

Distance matrices were converted into metric spaces whose filtrations were built using a Vietoris-Rips construction thresholded over a parameter that varied across the interval [0 1] in 300 steps, ensuring a fully connected simplicial complex construction due to matrix normalization. Persistence diagrams for zero-dimensional homology were calculated for all subjects and barcode endpoints were extracted for further testing between groups (birth times of connected components all begin at $\epsilon = 0$).

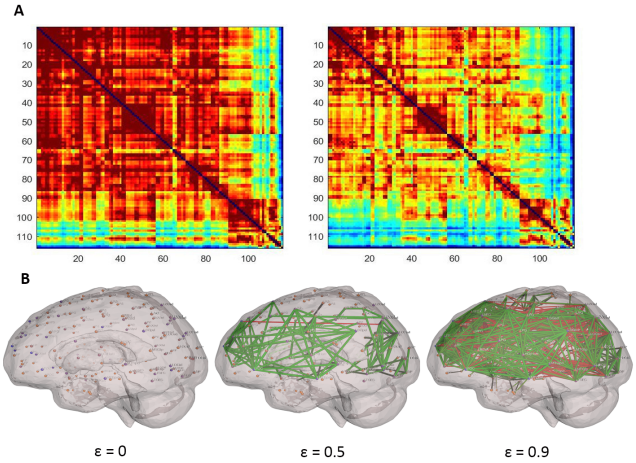

Fig. 3: Group-averaged connectivity matrices for controls (left) and MDD (right), color range from 0 to 1 (A). A filtration of a ROI set with ϵ increasing as one moves to the right. Note that each complex includes all neighbors to its left. One can track the timescale of various topological features in the data, including connected components and holes (B).

Dissimilarity between persistence diagrams was calculated through the bottleneck distance, resulting in a new distance matrix reflecting differences in the 0-persistence diagrams of subjects' global functional connectivity. Classical multidimensional scaling (MDS) was used to embed distances in $\mathbb{R}^n$, yielding a geometric interpretation of the topological differences that lent itself to several classification techniques. Classical MDS works by finding a point set such that the strain function, measuring the error between a point set approximation and true distances, is minimized. Specifically, we seek a minimization of:

$$Strain(x_1 ... x_n) = \frac{\sum_{i,j} d_{i,j} - ||x_i - x_j||}{\sum_{i,j} d_{i,j}}$$

Higher dimensional (approximate) embeddings more accurately preserve distances but at the cost of a higher feature number. When used in a setting with sparse training data, this excess of features can reduce classifier accuracy due to the "curse of dimensionality" - as we increase our set of features the number of hypotheses relating features to response variable(s) grows massively. On the other hand, our hypothesis that topological information is an important discriminant in classification suggests that we should preserve

measures of distance between persistence diagrams as faithfully as possible. In the event of a tradeoff, we found point sets in 3,4, and 5-dimensional space to identify resulting local or absolute optima for various classifiers. Principal component analysis (PCA) was applied to reduce the dimensionality of data in the event of correlations between features (95% variance explained after application).

A feedforward neural network, composed of two dense hidden layers with five ReLU activation units each and a softmax output for class probabilities, was deployed as well. The model was trained using Adam as an optimizer and 0.001 initial learning rate. Average cross-entropy was used to measure loss. We skipped partitioning a validation set due to the small number of data points, instead letting the network stop after 32 epochs of training. While this procedure risks allowing the network to overfit to the training data, we hypothesized that the relative risk of severely reducing the amount of training justified such a step. This potential trade-off was confirmed by later testing (mean accuracy difference 15% for 10 runs each). PCA was skipped for neural network training but features were rescaled using z-score normalization.

Both decision trees used 100 splits; one utilized Gini's diversity index for split criterion whereas the other used maximum deviance reduction. The nearest neighbors classifiers used k = 3. SVMs used a linear kernel with auto kernel scaling and soft-margin penalty c set to 1.

For all classifiers, we used a leave-one-out cross-validation technique and took average accuracy across all folds due to sparse data. In a leave-one-out approach for n data points, a classifier is trained on $n-1$ observations and tested on the excluded observation. The approach is repeated on all n points hence, in this setting, mean accuracy is calculated simply by dividing the number of correct classifications by the size of our dataset (33 subjects).

B. Graph Measures

We computed eight graph measures for evaluating persistent homology as a feature for classification: transitivity, characteristic path length, global efficiency, small worldness, assortativity, modularity, global betweenness, and global flow coefficient.

Crucially, we avoided binarizing our matrices using a threshold, as such a procedure can result in information loss and thereby lower discriminative ability [7] [22]. Additionally, the choice of threshold introduces a degree of freedom into data analysis that can produce variability in results [23] [24]. For global flow calculations, a binarized matrix is required and we therefore used a threshold of 25% due to unreliable test-retest reliability for lower values [8].

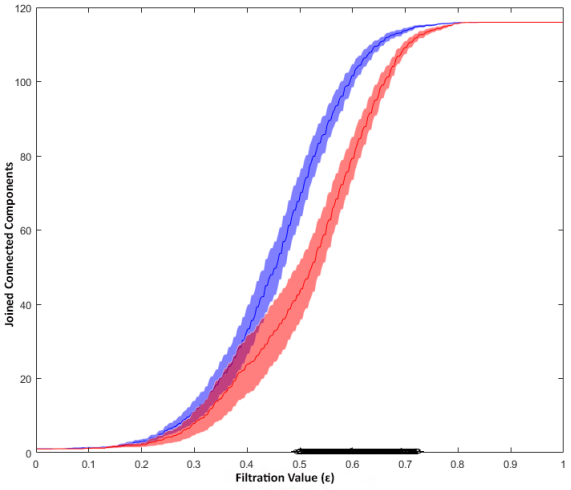

Fig. 4: The distributions show the average number of joined components (darker line) as a function of the filtration value for both groups as well as standard deviations (shading). At ϵ = 0 all components are separate and at ϵ = 1 the network is a complete graph. The dark line on the x-axis shows the location of a significant cluster.

IV. Results

A. Barcode Differences

A summary of network connectivity was generated by measuring the number of connected components joined as a function of the filtration value (Fig. 4). The resulting distributions were compared between groups using a non-parametric permutation test (Monte Carlo method, 5000 permutations, α = 0.05) with cluster correction, revealing a significant span of filtration values (p = 0.019. Fig. 2). The test does not assume normality and corrects for multiple comparisons by identifying clusters of significant points and thresholding based upon a secondary permutation test of cluster sizes (summed t-statistic). Cluster correction is advantageous in this setting as the number of connected components for similar ϵ values is not independent; Bonferroni correction would too harshly penalize the resulting statistics.

Inspection of barcode differences revealed shorter average lengths in controls relative to subjects with MDD ($\Delta_\mu = -0.0574$), implying a relatively higher degree of connectivity. This observation is corroborated by Fig. 2, which shows that nodes are joined more quickly in controls relative to the MDD group.

B. Diagram-Based Classifiers

Six classifiers were trained with three different settings for MDS output dimensionality. Classifier performance was evaluated by mean accuracy due to relatively balanced class sizes.

For comparison with a random classifier, a permutation test (5000 permutations) was used in which

class labels were randomized within folds and cross-validated. Permutation tests are critical in testing cross-validated performance as procedures such as binomial tests and t-tests rely on independence assumptions that do not hold [25] [26]. The observed classifier performance is then compared against the distribution of generated accuracies that are greater than or equal to 50%, the p-value being approximated by the number of observed values greater than the classifier performance divided by the number of observations in the permutation distribution.

Classifier Performance			
Classifier Type	MDS = 3	MDS = 4	MDS = 5
Decision Tree (I)	84.8%	84.8%	81.8%
Decision Tree (II)	84.8%	87.9%	87.9%
Logistic Regression	63.6%	63.6%	63.6%
Linear SVM	66.7%	66.7%	63.6%
1-KNN	48.5%	57.6%	60.6%
Neural network	66.7%	63.6%	63.6%

TABLE I: Classifier accuracies: note decision tree I uses Gini's diversity index for split criterion whereas decision tree II uses maximum deviance reduction.

The decision tree with maximum deviance reduction performed best with a statistically significant maximum accuracy of 87.9% ($p < 3.99 \times 10^{-4}$, sensitivity = 87.5%, specificity = 88.2% for MDD as positive class labels) compared to a random classifier. Both trees achieved statistical significance in classification for all MDS settings whereas a maximum classification accuracy of 66.7% ($p = 0.0419$) among all other classifiers failed to achieve significance following Bonferroni correction.

C. Landscape-Based Classifiers

We then investigated the use of persistence landscapes in classification. Since PLs are numerically stable under the bottleneck distance [14], we calculated pairwise distances to compare persistence diagram and landscape-based classification performance across MDS settings and algorithms.

A permutation test showed that PL-based classifiers did not perform significantly better or worse across classifiers and MDS settings (1000-permutation test, $p = 0.871$) although significant speed gains were observed ($\Delta_\mu = 4.590$ sec) when comparing pairwise bottleneck distance runtimes (15 randomly sampled pairwise distances each in a 1000-permutation test, $p = 0.0001$). Speed calculations were performed on a computer with i5-6200 2.3 GHz processor and 8 GB memory.

D. Comparison with Other Measures

In order to characterize the utility of persistent homology relative to other graph-theoretic measures, we implemented two approaches as described in [8] that consider linear SVM performance. The choice of SVM reflects the relative ease of interpreting the importance of various features to the overall model, reflected by absolute weight values. In the first approach, a SVM using all twelve features is trained using LOOCV. Feature weight ranks were computed for each fold due to the ease of cross-fold interpretation as opposed to weights, which reflect relative importance within a fold. Rankings were then averaged across folds (table 2). In the second approach, we generated 4095 SVMs by exhaustively considering all combinations of features from a set of twelve total. The average accuracy and standard deviation of all SVMs including a given feature is then computed. In both cases, PCA is skipped as it would generate a lower-dimensional basis using linear combinations of features, thus obscuring contributions of individual features.

The general SVM using all metrics performed significantly above random classification based upon a one-sided permutation test (mean accuracy = 72.7%, p < 0.011). Analysis of average feature weights showed that the first axis of the PD-based MDS output had the highest average ranking (and hence importance in classification) followed by *small worldness* and then *transitivity*. Small worldness quantifies the balance between local clustering and characteristic path length [27]. Transitivity measures the ratio of triangles (2-simplices if we view our graph as a low-dimensional simplicial complex) to connected triplets, reflecting the degree of clustering in the graph [8].

Feature Rankings			
Feature	Mean Rank	SD	Overall
PD (first coordinate)	1.000	0	1
PD (second coordinate)	9.000	1.225	9
PD (third coordinate)	11.061	0.496	11
PD (fourth coordinate)	5.273	1.098	5
Transitivity	3.515	1.149	3
Characteristic Path Length	7.667	1.242	8
Global Efficiency	4.030	1.759	4
Small Worldness	2.333	1.021	2
Assortativity	6.273	1.567	6
Modularity	11.606	0.864	12
Global Betweenness	9.455	1.148	10
Global Flow Coefficient	6.788	1.244	7

TABLE II: Ranking of features across all folds according to relative weights of a linear SVM.

Of the SVMs trained with only the first axis MDS feature, transitivity, or small worldness, only the one using transitivity showed significant performance following Bonferroni corection (one-sided permutation test, 5000 permutations, p = 0.030, 0.016, 0.301 respectively).

In the second approach, we find that the first MDS coordinate yields fifth place for overall accuracy

across SVMs. We additionally examined SVMs associated with the roughly top 10% of scores (26 total scores, top 3 = 11.5% corresponding to 1556 SVMs) and counted feature occurrences. This method places the first and fourth MDS coordinate well above an expected 129.67 SVM count assuming each feature is evenly represented in the pool. Overall, the first MDS coordinate ties with transitivity for third highest SVM count.

Feature Accuracies			
Feature	Mean Accuracy	SD	10% Count
PD (first coordinate)	70.4%	0.0513	151
PD (second coordinate)	69.0%	0.0567	94
PD (third coordinate)	68.8%	0.0599	90
PD (fourth coordinate)	70.2%	0.0597	148
Transitivity	70.9%	0.0473	151
Characteristic Path Length	69.7%	0.0570	140
Global Efficiency	70.0%	0.0556	110
Small Worldness	71.8%	0.0518	188
Assortativity	70.9%	0.0467	146
Modularity	69.3%	0.0608	109
Global Betweenness	69.7%	0.0478	64
Global Flow Coefficient	70.5%	0.0533	165

TABLE III: Mean accuracies for all SVMs including a given feature along with standard deviations. Column three shows feature representation in a pool of 1556 SVMs corresponding to approximately the top 10% of accuracy scores.

In order to separate discriminative ability from group-level population differences and test reported findings in earlier studies, we performed two-sided 10,000-permutation tests on the eight introduced graph metrics. Transitivity ($p = 0.046$), assortativity ($p = 0.0072$), and global betweenness ($p = 0.0259$) differences were significant although no metrics remained so following Bonferroni correction.

V. Discussion

Analysis of network connectivity changes as a function of the filtration parameter revealed a significant group level difference (Fig. 4), indicating that MDD participants on average exhibited lower global functional connectivity compared to controls. The result replicates the findings in [7] and corroborates other studies examining global functional connectivity abnormalities [28] [29], suggesting that global connectivity measures are promising biomarkers for depression.

Previous studies have indicated the utility of small worldness in classifying a number of psychiatric disorders [8] [22]. In particular, one study found that small worldness had the highest average ranking compared to other common graph measures in depression classification [8]. Importantly, small worldness was not sufficient for significant classification accuracy, a finding backed by another rs-fMRI study of MDD that reported similar small worldness properties across MDD and control groups [4] (albeit, classification performance and population differences do not always translate directly). Our results corroborate these findings: in a linear SVM trained on all features, small worldness had the second highest average weight ranking across all folds, second only to a MDS-transformed persistence diagram feature, and the highest associated accuracy. Despite these performance benchmarks, group differences were small ($p > 0.5$) and a SVM trained solely on small worldness failed to achieve significant performance in classification ($p > 0.3$), implying that interactions between small worldness and various other features drive its relative importance. We also note that the previous study used tractography-based graph metrics whereas our data originate from fMRI BOLD signals, suggesting that small worldness as a classification feature for MDD may be relatively robust to anatomical or functional measures of connectivity, albeit a detailed analysis is warranted.

Following Bonferroni correction, none of the observed graph metric differences remained statistically significant. However, we suspect that some graph metrics may have true population differences and that the current analysis is relatively underpowered. In particular, reduced characteristic path length is indicative of lowered overall connectivity, a finding supported by the observed 0-persistent homology differences.

Our results also show the utility of persistent homology as a classification feature with decision trees using MDS-scaled PD and PL distances performed significantly better than random classifiers. Additionally, the first MDS-scaled coordinate had the highest associated weight rank across all folds in a linear SVM trained on all features. This weight ranking disparity means that the resultant hyperplane used to separate classes is more orthogonal to the MDS-scaled first coordinate axis relative to all other features, implying that the distribution along this axis has higher discriminative importance. While none of the MDS coordinates exhibited high mean accuracy rank, the prominence of the first coordinate in the SVMs associated with the top 10% of scores may imply that, while persistent homology alone does not yield high discriminative ability in SVMs, an interaction effect may be at work as high performing SVMs had an average of 6.25 features each (median = 6).

A. Limitations

While we argue that persistent homology, through methods such as MDS, can be used to create promising biomarkers, the secondary analysis follows an initial study of whether or not 0-persistent homology differs between classes. This concurrency may risk introducing bias into the experimental design; a feature is studied for potential significance and the subsequent development of a model can appear as a form of ”double dipping”. That said, such bias is offset by the original hypothesis that persistent homology could

be leveraged through embedding methods i.e. group-level statistics did not inform consequent application to classification. Nevertheless, the choice of MDS dimension is a source of bias that should be accounted for in future studies, either through analytic methods, testing on a validation set, or through persistence images [15], numerically stable, vectorized representations of persistence diagrams that bypass the need for procedures such as MDS or relative distance computations.

A better examination of persistent homology's utility in rs-fMRI depression classification is hindered by small study sizes common to fMRI analyses. The lack of training data meant we were unable to engage in any hyperparameter tuning with a validation set, a process integral to optimizing machine learning methods. We hence caution against any conclusions of relative classifier performance across types.

The current population pool included multiple participants with co-morbidities. As schizophrenia, depression, and PTSD are all correlated with resting state network dysfunction [22] [5] [30], the obtained classifier accuracy likely reflects differences between healthy individuals and those suffering from depression as well as a variety of other psychiatric diseases. In order to fully characterize persistent homology's capacity as a discriminative feature for depression, further studies examining multi-class classification are necessary.

Additionally, we limited testing of persistence landscapes to MDS-scaled comparisons with persistence diagrams in order to avoid increases in the FWER. A more cogent analysis of persistence landscape performance is warranted due to the variety of methods enabled by their relatively well-behaved mathematical properties as well as evidence in favor of persistence landscapes in classifier performance [15] [31].

VI. Conclusion

Earlier studies have used persistent homology to study network-level differences in disease, including autism [6], attention deficit hyperactivity disorder [6], and a mouse model of depression [32]. To our knowledge, this is the first study to apply persistent homology to the study of depression in human beings as well as an evaluation of its performance relative to common graph measures through the use of multidimensional scaling. The results indicate that 0-persistent homology is an effective tool at capturing global rs-fMRI connectivity differences between healthy and depressed individuals. Through the use of persistence diagrams and landscapes, persistent homology could provide promising biomarkers for supervised classification in tandem with common graph measures.

Future directions might include the use of transformations such as persistence images as well as incorporating higher dimensional homology information. Additionally, larger study pools may enable more powerful machine learning techniques to be leveraged, inlcuding deep learning methods that require substantial hyperparameter tuning and training set sizes.

Acknowledgments

The authors thank Berman *et al.* [7] for creating and making publicly available the dataset used in this study. The authors also thank Jeff Hicks for feedback and advice regarding TDA theory.

References

[1] L. Pessoa, "Understanding brain networks and brain organization," *Physics of life reviews*, vol. 11, no. 3, pp. 400–435, 2014.

[2] J. Hahm, H. Lee, H. Park, E. Kang, Y. K. Kim, C. K. Chung, H. Kang, and D. S. Lee, "Gating of memory encoding of time-delayed cross-frequency meg networks revealed by graph filtration based on persistent homology," *Scientific reports*, vol. 7, 2017.

[3] 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, no. 101, 2014. [Online]. Available: http://rsif.royalsocietypublishing.org/content/11/101/20140873

[4] M. Ye, T. Yang, P. Qing, X. Lei, J. Qiu, and G. Liu, "Changes of functional brain networks in major depressive disorder: A graph theoretical analysis of resting-state fmri," *PLOS ONE*, vol. 10, no. 9, pp. 1–16, 09 2015. [Online]. Available: https://doi.org/10.1371/journal.pone.0133775

[5] K. RH, A.-H. JR, W. TD, and P. DA, "Large-scale network dysfunction in major depressive disorder: A meta-analysis of resting-state functional connectivity," *JAMA Psychiatry*, vol. 72, no. 6, pp. 603–611, 2015. [Online]. Available: +http://dx.doi.org/10.1001/jamapsychiatry.2015.0071

[6] H. Lee, M. K. Chung, H. Kang, B. N. Kim, and D. S. Lee, "Discriminative persistent homology of brain networks," in *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, March 2011, pp. 841–844.

[7] M. G. Berman, B. Misic, M. Buschkuehl, E. Kross, P. J. Deldin, S. Peltier, N. W. Churchill, S. M. Jaeggi, V. Vakorin, A. R. McIntosh, and J. Jonides, "Does resting-state connectivity reflect depressive rumination? a tale of two analyses," *NeuroImage*, vol. 103, pp. 267 – 279, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1053811914007691

[8] M. D. Sacchet, G. Prasad, L. C. Foland-Ross, P. M. Thompson, and I. H. Gotlib, "Support vector machine classification of major depressive disorder using diffusion-weighted neuroimaging and graph theory," *Frontiers in Psychiatry*, vol. 6, p. 21, 2015. [Online]. Available: http://journal.frontiersin.org/article/10.3389/fpsyt.2015.00021

[9] V. De Silva and R. Ghrist, "Coverage in sensor networks via persistent homology," *Algebraic & Geometric Topology*, vol. 7, no. 1, pp. 339–358, 2007.

[10] M. Nicolau, A. J. Levine, and G. Carlsson, "Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival," *Proceedings of the National Academy of Sciences*, vol. 108, no. 17, pp. 7265–7270, 2011.

[11] H. Liang and H. Wang, "Structure-function network mapping and its assessment via persistent homology," *PLOS Computational Biology*, vol. 13, no. 1, pp. 1–19, 01 2017. [Online]. Available: https://doi.org/10.1371/journal.pcbi.1005325

[12] N. Otter, M. A. Porter, U. Tillmann, P. Grindrod, and H. A. Harrington, "A roadmap for the computation of persistent homology," 2015.

[13] E. Munch, "Applications of persistent homology to time varying systems," 2013. [Online]. Available: https://dukespace.lib.duke.edu/dspace/bitstream/handle/10161/7180/Munch_duke_0066D_11841.pdf%3bsequence=1

[14] P. Bubenik, "Statistical topological data analysis using persistence landscapes," *The Journal of Machine Learning Research*, vol. 16, no. 1, pp. 77–102, 2015.

[15] S. Chepushtanova, T. Emerson, E. M. Hanson, M. Kirby, F. C. Motta, R. Neville, C. Peterson, P. D. Shipman, and L. Ziegelmeier, "Persistence images: An alternative persistent homology representation," *CoRR*, vol. abs/1507.06217, 2015. [Online]. Available: http://arxiv.org/abs/1507.06217

[16] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer, "Stability of persistence diagrams," *Discrete & Computational Geometry*, vol. 37, no. 1, pp. 103–120, Jan 2007. [Online]. Available: https://doi.org/10.1007/s00454-006-1276-5

[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, no. 1, pp. 273–289, 2002.

[18] A. Tausz, M. Vejdemo-Johansson, and H. Adams, "JavaPlex: A research software package for persistent (co)homology," in *Proceedings of ICMS 2014*, ser. Lecture Notes in Computer Science 8592, H. Hong and C. Yap, Eds., 2014, pp. 129–136, software available at http://appliedtopology.github.io/javaplex/.

[19] P. Bubenik and P. Dłotko, "A persistence landscapes toolbox for topological statistics," *Journal of Symbolic Computation*, vol. 78, pp. 91–114, 2017.

[20] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, software available from tensorflow.org. [Online]. Available: http://tensorflow.org/

[21] M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: uses and interpretations," *Neuroimage*, vol. 52, no. 3, pp. 1059–1069, 2010.

[22] A. Anderson and M. S. Cohen, "Decreased small-world functional network connectivity and clustering across resting state networks in schizophrenia: an fmri classification tutorial," *Frontiers in human neuroscience*, vol. 7, 2013.

[23] E. L. Dennis, N. Jahanshad, A. W. Toga, K. L. McMahon, G. I. De Zubicaray, N. G. Martin, M. J. Wright, and P. M. Thompson, "Test-retest reliability of graph theory measures of structural brain connectivity," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2012, pp. 305–312.

[24] G. Gong, Y. He, L. Concha, C. Lebel, D. W. Gross, A. C. Evans, and C. Beaulieu, "Mapping anatomical connectivity patterns of human cerebral cortex using in vivo diffusion tensor imaging tractography," *Cerebral cortex*, vol. 19, no. 3, pp. 524–536, 2008.

[25] Q. Noirhomme, D. Lesenfants, F. Gomez, A. Soddu, J. Schrouff, G. Garraux, A. Luxen, C. Phillips, and S. Laureys, "Biased binomial assessment of cross-validated estimation of classification accuracies illustrated in diagnosis predictions," *NeuroImage: Clinical*, vol. 4, pp. 687 – 694, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2213158214000485

[26] M. Ojala and G. C. Garriga, "Permutation tests for studying classifier performance," *Journal of Machine Learning Research*, vol. 11, no. Jun, pp. 1833–1863, 2010.

[27] M. D. Humphries and K. Gurney, "Network small-worldness: A quantitative method for determining canonical network equivalence," *PLOS ONE*, vol. 3, no. 4, pp. 1–10, 04 2008. [Online]. Available: https://doi.org/10.1371/journal.pone.0002051

[28] M. Sacchet, T. Ho, C. Connolly, O. Tymofiyeva, K. Z Lewinn, L. Han, E. Blom, S. F Tapert, J. E Max, G. Frank, M. Paulus, A. Simmons, I. Gotlib, and T. T Yang, "Large-scale hypoconnectivity between resting-state functional networks in unmedicated adolescent major depressive disorder," vol. 41, 05 2016.

[29] I. M. Veer, C. Beckmann, M.-J. Van Tol, L. Ferrarini, J. Milles, D. Veltman, A. Aleman, M. Van Buchem, N. Van Der Wee, and S. Rombouts, "Whole brain resting-state analysis reveals decreased functional connectivity in major depression," *Frontiers in Systems Neuroscience*, vol. 4, p. 41, 2010. [Online]. Available: http://journal.frontiersin.org/article/10.3389/fnsys.2010.00041

[30] C. Sylvester, M. Corbetta, M. Raichle, T. Rodebaugh, B. Schlaggar, Y. Sheline, C. Zorumski, and E. Lenze, "Functional network dysfunction in anxiety and anxiety disorders," *Trends in neurosciences*, vol. 35, no. 9, pp. 527–535, 2012.

[31] J. Liu, S. Jeng, and Y. Yang, "Applying topological persistence in convolutional neural network for music audio signals," *CoRR*, vol. abs/1608.07373, 2016. [Online]. Available: http://arxiv.org/abs/1608.07373

[32] A. Khalid, B. S. Kim, M. K. Chung, J. C. Ye, and D. Jeon, "Tracing the evolution of multi-scale functional networks in a mouse model of depression using persistent brain network homology," *NeuroImage*, vol. 101, pp. 351–363, 2014.