A Comparison of Brain Connectivity Graphs Across Multi-Band Acquisition Factors

Sohail Nizam
snizam@emory.edu
Emory University
Atlanta, Georgia

Abstract

The representation of data collected from brain imaging studies using graphs is a powerful technique that can be used to study many disorders and gain a better understanding in general of anatomical and functional brain connectivity. Simultaneous multislice (SMS) techniques improve temporal resolution in magnetic resonance imaging by using a multiband (MB) radiofrequency pulse to collect multiple slices at the same time, which also enables higher spatial resolution. However this technique requires extra image preprocessing steps which can result in noise amplification at varying levels across MB factors. In this study we examine the effect of the MB factor at which functional MRI scans were acquired on the structure of the brain connectivity graphs derived from those scans. fMRI scans were taken at 9 different MB factors for 32 different subjects. The resulting graphs are represented by correlation matrices. Graphs were first compared using classic summary measures such as average degree, average path length, and average clustering coefficient. Next, low-dimensional representations of the graphs were created using Anonymous Walk Embeddings. These embeddings were compared across MB factors with a variety of techniques. During the analysis, the raw weighted graphs as well as unweighted graphs created by thresholding the adjacency matrices at different values were considered.

ACM Reference Format:
Sohail Nizam. 2018. A Comparison of Brain Connectivity Graphs Across Multi-Band Acquisition Factors. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/1122445.1122456

Woodstock '18, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

1 Introduction

Our ability to construct and compare graphs using brain imaging data is of great scientific importance. Deriving and studying brain connectivity graphs can help us gain a better understanding of how the brain is structured, how it functions, and how we process information. Of particular interest is using connectivity graphs to identify and better understand how structure and function are altered by brain-related disorders. Some examples of previous work done in this area include finding differences in connectivity structure between healthy patients and those with Alzheimer's [1] and other conditions. Changes in connectivity structure have been identified as people age [2], enter different sleep cycles [3], and begin epileptic seizures [4].

One important topic that has not been studied as frequently is the effect that different methods of brain image acquisition can have on the connectivity graphs that are derived from them. When studies such as the ones cited above are conducted, the data are typically collected using only one acquisition method. Thus, knowing before-hand the how acquisition methods can change an analysis is of great importance. In this paper, we focus specifically on the effect of multiband (MB) factors on structure in graphs derived from functional magnetic resonance imaging (fMRI) scans. MB factor can be thought of essentially as a tuning parameter that dictates how the brain scan is obtained. It is a key factor in the simultaneous multislice (SMS) technique for collecting fMRI data. SMS techniques improve temporal resolution in magnetic resonance imaging by using a multiband radiofrequency pulse to collect multiple slices at the same time. This process enables higher spatial resolution and can facilitate the removal of higher-frequency artifacts and boost effective sample size. However, SMS requires an additional processing step in which the MB slices are separated, or unaliased, to recover the whole brain volume. This results in noise amplification, which increases with higher MB factors. High MB factors have been popularized by the Human Connectome Project (HCP), which seeks to map all structural and functional connections in the human brain [5]. The HCP's approach may systematically mischaracterize connections between brain regions due to noise amplification.

In this study we consider fMRI scans from 32 different subjects. Each subject had scans taken at 9 different MB factors. Each graph is represented by a 264x264 adjacency

matrix of correlation coefficients measuring the strength of association of blood-oxygen-level-dependent (BOLD) signals between regions of interest.

While problems such as graph classification are well studied, there are fewer methods for actually characterizing differences in graphs of different classes in a rigorous way. The goal of this analysis is not simply to build a system that can label a graph with an MB factor although we do consider this as a sub-task. Instead, the goal is to understand the impact of MB factor on the construction of graphs that will inevitably be used for analyses related to other outcomes.

There are two main ways in which we will approach the analysis. First, we will compare numerical summaries of graphs across MB factors. Summaries will include common graph metrics such as average degree, average path length, and average clustering coefficient. Second, we will create Anonymous Walk Embeddings of the graphs and compare them visually and numerically.

2 Related Work and Method Backgrounds

This paper is an intersection between two fields that often strive for very different goals. The field of Neuroimaging statistics typically aims to perform inference and identify significant associations between aspects of brain structure and certain outcomes such as diseases and disorders. The most advanced techniques in graph data analysis are largely concerned with graph representation learning and deep learning tasks. Before describing the approach to our specific problem, we examine work related to our problem from the two fields in an effort to make clear how we plan to leverage them both.

2.1 Classic Techniques for Brain Graph Characterization

It is quite common for brain connectivity graphs to be compared using classic graph summary metrics such as degree distribution, average shortest path length, and average clustering coefficient. Rubinov and Sporns [6] reason that these measures can represent important features of functional connectivity in brains. For example, they state that clustering coefficient and path lengths are measures of functional segregation and functional integration respectively.

Stam et al. [1] use this approach to compare brain connectivity graphs between patients with and without Alzheimer's Disease. The graphs begin as correlation matrices between regions of interest in the brain. They start by averaging the matrices within each group resulting in two weighted graphs. Next, they choose many different threshold values that are then used to binarize the matrices thus creating two unweighted graphs. For the resulting graphs at each threshold value, they then calculate graph summary measures. In this particular paper they focus on the clustering coefficient $C = \sum_{i=1}^{n} \frac{2e_i}{k_i(k_i-1)}$, and the average shortest path length $L = \frac{1}{n}\sum_{i=1}^{n} \frac{\sum_{j=1,j\neq i}^{n} d_{ij}}{n-1}$. However, analyses could include more measures depending on what properties of the graph are of interest. One could also consider modularity M if interested in functional segregation or other metrics. The authors then note that when comparing graphs at a given threshold, the results would be influenced by differences in the mean level of correlation between those with and without Alzheimer's. Such differences are already known to exist and are not the focal point of this analysis. Thus they also repeat the analysis computing C and L as a function of the average degree K after the matrices have been thresholded. This ensures that, upon comparison, both graphs have the same number of edges and that any differences found would be differences in graph organization.

The final step in their analysis is to create plots of both C and L versus both threshold level and average degree K. Confidence intervals are produced for each measured value of C and L, linear regressions relating each metric to both threshold and degree are performed, and resulting estimates are compared.

Rubinov and Sporns [6] note that thresholding naturally causes some information loss. Thus, they suggest considering weighted counterparts of those classic graph summary measures. Options for weighted summaries include the weighted clustering coefficient $C^w = \sum_{i=1}^{n} \frac{2e_i^w}{k_i(k_i-1)}$, the weighted average path length $L^w = \frac{1}{n}\sum_{i=1}^{n} \frac{\sum_{j=1,j\neq i}^{n} d_{ij}^w}{n-1}$, and the weighted average degree $A^w = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}$. In this case, comparison of groups across threshold values is not necessary. One can simply calculate the weighted metrics for each subject's graph to compare or average the adjacency matricies within groups and calculate one metric per group.

2.2 Anonymous Walk Embeddings

In their 2018 paper, Ivanov and Burnaev [7] lay out a method for low-dimensional graph embedding via a process called Anonymous Walks. Anonymous Walks are similar to random walks in that one starts at a single node, chooses from the connecting nodes at random to step towards, and repeats this process a pre-determined number of times. The result of a single walk is the list of nodes that were stops on the walk. However, Anonymous Walks do not use global graph node labels. Nodes in Anonymous Walks are enumerated relative to the first node on the walk. For example, if we observed a random walk with labels [a, b, a, c], the corresponding Anonymous Walk would be [0, 1, 0, 2]. Another random walk that does not share any of the same nodes as the first, but displays the same pattern relative to the first node would be identical as an Anonymous Walk. For example, the random with labels [g, h, g, i] would also yield the Anonymous Walk [0, 1, 0, 2].

Embeddings for a graph are created by enumerating all of the Anonymous Walks of a given size and storing their frequencies in the graph. For larger graphs and/or large walk

sizes, a large number of Anonymous Walks can be chosen at random and the resulting frequencies act as an estimate of the true frequency distribution.

Ivanov and Burnaev demonstrate that these embeddings paired with a simple support vector machine classifier can achieve classification accuracies that are competitive with other state of the art graph classification methods.

3 Problem Formulation and Preliminaries

Based on these previous works, we now formally lay out our problem and potential solutions. The problem is one of finding potential structural differences in graphs derived from fMRI scans taken at different MB factors. In this way the study is somewhat exploratory and will likely inform future work. We follow in the footsteps of other studies that aim to characterize graphs derived from brain images by examining graph features that are known to have biological significance like degree, path length, and clustering coefficient. We veer away from tradition by constructing and examining Anonymous Walk Embeddings for the graphs.

Anonymous Walk Embeddings are a natural choice for our study. Their ability to accurately represent graph information in a low-dimensional form while being relatively interpretable compared to other types of graph representation mean that one can use them for tasks beyond just classification tasks. Embeddings can be examined in detail, and one can compare the most common walks across classes. They do not make any assumptions about the importance of classic graph measures such as degree distribution or path length. Nor do they rely on pre-chosen motifs such as triangles. Furthermore, as previously mentioned, Ivanov and Burnaev [7] demonstrate an ability to do good graph classification using AWEs. Thus a secondary problem we will investigate is our ability to do graph classification with MB factors as labels.

An extra consideration to make is that of preprocessing. Most methods related to the analysis of graph data deal solely with unweighted graphs that can be represented by binary adjacency matrices. For this reason, it is typical in studies with brain imaging data to threshold the correlation matrices before conducting analyses. There is no standard for choosing threshold values, and previous literature simply recommends repeating the analysis at many different values [6]. It is also common practice to use the Graphical Lasso [8] to estimate inverse covariance (precision) matrices for use in the analysis rather than the original correlation matrices. Graphical Lasso requires the choice of a sparsity parameter which, again, has no established standard in the literature. Taking these considerations into account, an added goal of this study is to characterize how different preprocessing choices can impact our ability to examine differences in MB factor. For simplicity's sake, we will not consider the use of Graphical Lasso and the structure of precision matrices in this paper. That would, however, be of great interest in the future.

4 Technical Design

Below is an outline of the entire analysis conducted in this study.

4.1 Numerical Summaries

First, correlation matrices were averaged across subjects within each MB factor group. This resulted in 9 mean correlation matrices. The distribution of Average Weighted Degree across all nodes was examined for each MB factor. Similarly, the mean Weighted Path Length and mean Weighted Clustering Coefficient across nodes were examined for each MB factor. Differences in each metric across MB factor were noted.

4.2 Anonymous Walk Characterizations

Anonymous Walk Embeddings using walk length of 3 were created for each mean correlation matrix. The distribution of walks was noted for each MB. Mean matrices were then thresholded at several values and AWE distributions were again noted for each MB.

Next, AWEs were created for each of the 32*9 individual graphs in the dataset. Walks of length 5 and 7 were created from the graphs at each of several threshold values. These embeddings were then visualized by plotting them on the plane of the first two principal components. Each point on the plots was colored according to its MB label. Finally, our ability to do graph classification was assessed by choosing select MB factors and finding the cross-validated mean accuracy score provided by a Support Vector Machine classifier. SVM results were examined after training on embeddings of various walk sizes derived from graphs thresholded at various values.

5 Experimental Analysis

5.1 Numerical Summaries

First we examine how degree distribution differs between MB factors. Table 1 shows weighted average degree distribution summaries for each

It is clear that the degree distribution differs across MB factors with SB3 being the most dense and MB12 being the most sparse. Since it is common practice to threshold graphs derived from brain images, we also examine mean degree for each MB after a variety of threshold values. These results can be found in Figure 1. Again, there is a clear difference across thresholds, however the difference seems to shrink as the threshold passes a correlation of 0.2. It must be noted, however, that there are differences at the highest thresholds. These differences are being obscured by the scale of the plot. Table 1 and Figure 1 show us that, if nothing else, the scale of the correlations in the matrices differs by MB. The question

MB	Mean	SD	Min	Q1	Median	Q3	Max
SB3	24.43	5.63	9.86	20.78	24.81	28.51	37.99
SB2	24.24	5.13	10.75	20.75	24.83	28.08	36.2
MB2	21.84	5.32	8.33	18.12	21.49	25.96	37.11
MB3	20.91	5.23	6.8	17.87	21.14	24.55	35
MB4	21.89	5.58	6.85	18.5	22.42	25.66	38.3
MB6	20.59	5.43	4.95	17.01	21.02	24.17	36.67
MB8	18.47	5.73	4.94	14.83	18.89	22.29	36.67
MB9	16.2	4.87	3.89	13.22	16.26	19.57	29.89
MB12	13.33	4.34	3.57	10.47	13.59	16.18	26.64

Table 1. Average Weighted Degree distribution summaries for mean matrices across all 32 subjects within each MB factor.

MB	Avg. Path Length	Avg. Clustering Coef.
SB3	0.003	0.0922
SB2	0.0031	0.0937
MB2	0.0025	0.0833
MB3	0.0024	0.0823
MB4	0.0027	0.0868
MB6	0.0022	0.0831
MB8	0.002	0.0788
MB9	0.0016	0.0734
MB12	0.0015	0.0661

Table 2. Average path length and clustering coefficient for mean matrices across all 32 subjects within each MB factor.

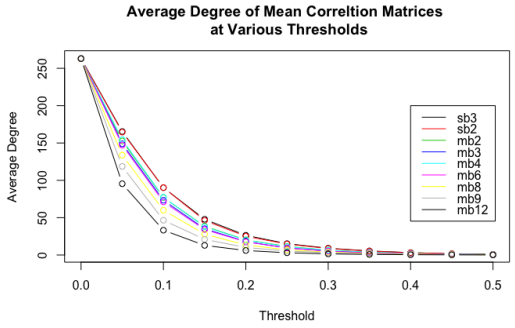

Figure 1

now becomes whether or not the pattern of correlations changes.

We can examine this further by looking at average path length and average clustering coefficient. Rubinov and Sporns [6] note that average path lengths and average clustering coefficients can be considered measures of functional segregation and integration in the brain respectively. Functional segregation is the brain's ability to conduct specialized processing within groups of brain regions that are interconnected. Functional integration is the brain's ability to combine information from different brain regions. Table 2 shows average path lengths and clustering coefficients for the mean correlation matrices across all 32 subjects within each MB factor. It is clear that as MB factor increases, both path length and clustering coefficient decrease. This suggests that representations of brains from scans at higher MB factors suggest higher functional integration and lower functional segregation.

5.2 Anonymous Walk Embeddings

Anonymous Walk Embeddings were created both for the 9 mean correlation matrices and for the 32*9 individual correlation matrices. We first discuss insights derived from the mean correlation matrix embeddings.

AWEs with walk length of 3 were created for each of the 9 mean matrices before thresholding and after thresholding at values .3 and .5. Figures 2-7 Show the distributions of these walks just for SB3 and MB12. Note first in Figures 2 and 3 that, before thresholding, the walk distributions seem nearly identical in that the most common walk by far (comprising over 95% of all walks in each case) is walk number 4 which is represented by [0, 1, 2, 3]. This is the walk with a new node at every step. At first glance this may be surprising, however, we must note that, while there are many small correlations in each matrix, there are very few correlations of exactly 0. This means that each graph is close to fully connected. At any given node, there are close to 263 options to take a step towards with non-zero probability. So, especially for such short walks, it makes sense that the most likely scenario is a new node at each step. Thus it seems that in raw form, the correlation matrices are too noisy to analyze, at least in this way.

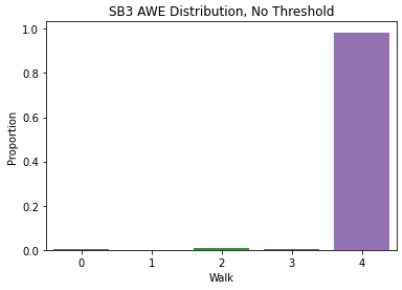

Figure 2

We investigated further by thresholding the matrices before computing AWEs. With thresholds at .3, Figures 4 and 5 still show [0, 1, 2, 3] as the most common walk, but now that walk comprises about 70% of the SB3 distribution and just 35% of the MB12 distribution. Furthermore, the distribution of the remaining walks seems to differ between SB3 and MB12.

With thresholds at .5 in Figures 6 and 7, we see the proportion of [0, 1, 2, 3] walks decrease again and the difference

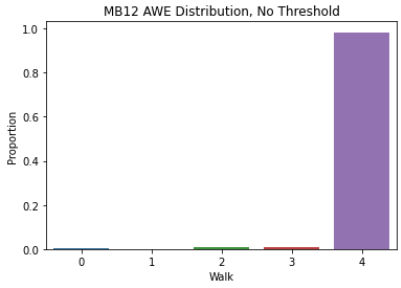

Figure 3

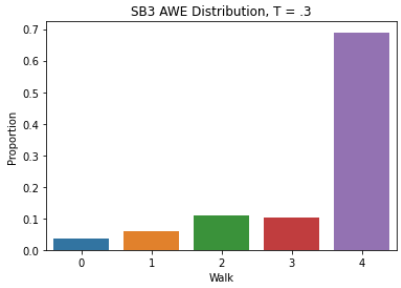

Figure 4

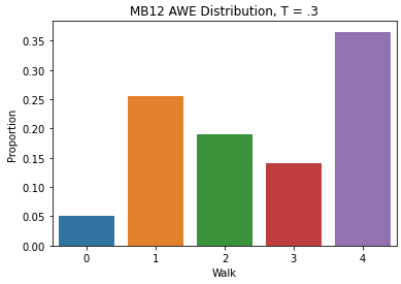

Figure 5

in distributions overall becomes more pronounced. Walks 2 and 3 ([0, 1, 2, 1] and [0, 1, 0, 2] respectively) are far more prevalent for MB12 than SB3 while walk 4 ([0, 1, 2, 3]) remains more prevalent than walks 2 and 3 for MB12. In each case, walk 1, [0, 1, 0, 1], is the most prevalent.

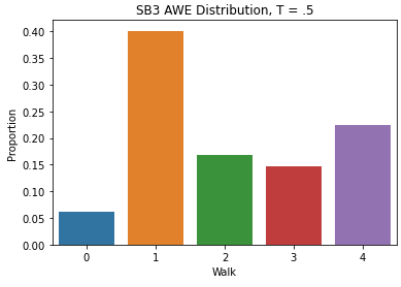

Figure 6

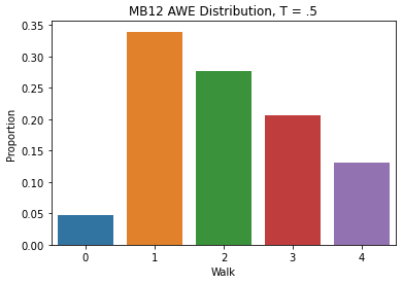

Figure 7

Next we create AWEs using walk lengths of 5 for each individual matrix without thresholding and after thresholding at several values. There are 52 unique anonymous walks of length 5, and as we've seen, walk distributions can differ while still having the same most and least frequent walks. So rather than simply examining most common walks here, we try to visualize the embeddings. Embeddings for all 32*9 unthresholded graphs using walk length 5 were stored in a single matrix (one graph embedding per row), and principal components analysis was performed. The embeddings were plotted on the plane of the first two principal components as shown in Figure 8. Points in the plane were colored according to MB factor.

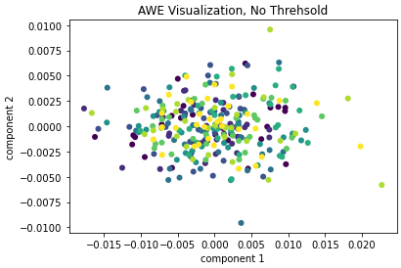

Figure 8

Without thresholding, there is no clear separation between the graphs of different MB factors. It must be noted that only 57% of the variation is captured by these first two PCs, however, it is still notable that there is absolutely no class separation. Contrast this image with Figure 9 which shows embeddings generated after thresholding at the value .5.

In this case, the first two principal components captured 97% of the variation in the data. Now there seems to be a clear color gradient showing that graphs with larger MB factors tend to be farther away from graphs with smaller MB factors. This process was repeated once more. The threshold was kept at .5, but walks of length 7 were used.

Again, the clear color gradient indicates separation between the classes.

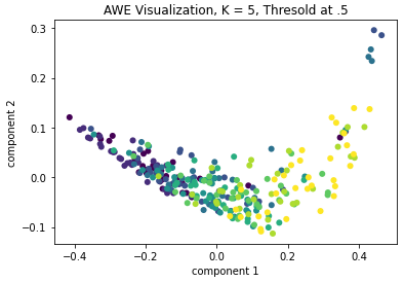

Figure 9

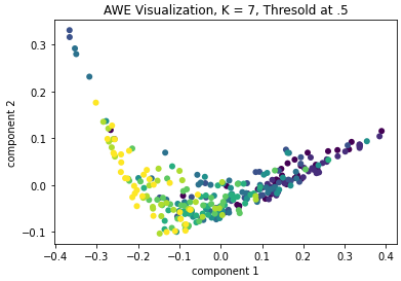

Figure 10

5.3 Graph Classification

For the final element of of our investigation, based on the seperability shown in the PCA plots at certain thresholds, we check our ability to do graph classification with a simple Support Vector Machine. Rather than treat this as a multiclass problem, we choose several pairs of MB factors and do binary classification. Table 3 shows the results of classifiction under several setups.

MB Factors	Walk Length	Threshold	Accuracy
SB3 vs MB12	5	None	.4859
SB3 vs MB12	5	.5	.9679
SB3 vs MB12	7	.5	.9679
MB2 vs MB9	5	None	.5026
MB2 vs MB9	5	.5	.7526
MB2 vs MB9	7	.5	.7679
MB4 vs MB8	5	None	.4026
MB4 vs MB8	5	.5	.6859
MB4 vs MB8	5	.5	.6397

Table 3. 5-fold cross validated mean accuracy for binary classification under a variety of embedding setups.

We can see that our ability to predict MB factor increases greatly after appropriate thresholding but that MB factors that are closer together are harder to distinguish. Notably, increasing walk length does not, by default, improve accuracy.

6 Conclusion

This paper has been an initial investigation into our ability to discern differences between graphs derived from fMRIs taken using different MB factors. We examined simple graph summary measures as is typical in neuroimaging statistics studies. We also took a new approach by creating and examining Anonymous Walk Embeddings. The results provide interesting insight into how one can use AWEs in a general graph comparison problem as well as our specific problem.

We found that graphs constructed at different MB factors differ in terms of degree distribution, average path length, and average clustering coefficient. This is a relatively shallow way of comparing graphs, but indicates to us that it is worth investigating further differences in graph structure.

Using Anonymous Walk Embeddings we discovered that we can uncover differences in graph structure but that the thresholding level matters greatly. Counter-intuitively, the raw correlation matrices do not seem to be the best representations of the information at least as far as Anonymous Walk Embeddings go. The noise created by very small but nonzero correlations in the adjacency matrices results in uninformative walk distributions that are dominated by the [0, 1, 2, ..., k] walk (in the case of a k-step walk). Thresholding can act as a de-noising device, removing most of these uninformative walks and preserving the graph structure. Once the proportion of [0, 1, 2, ..., k] walks is decreased, it is possible that a distinctive walk distribution will emerge. We observed that after appropriate thresholding, clear separation in MB factors can be observed on the plane of the first two principal components of the graph embeddings.

Finally, we discovered that using a simple support vector machine classifier, we may have goo ability to predict MB factor. Again, this depends heavily on thresholding value. It also becomes more difficult as MB factors in the two groups become closer to each other. This makes sense intuitively and based on the PCA plots shown.

In the future, it would be valuable to derive a statistical test based on anonymous walk embeddings. This would involve creating some test statistic for which we are able to derive a distribution. It would also be valuable to more deeply investigate Anonymous Walk Embeddings on a larger set of more diverse data to try to identify biologically meaningful associations between specific walk patterns and brain structure. It is possible that certain walks or walk patterns are associated with certain brain properties. Finally, since we identified the great importance of thresholding our correlation matrices, it would be valuable to investigate the impact of using Graphical Lasso precision matrix estimation as an extra pre-processing step on our results. Graphical Lasso allows us to control the sparsity level in the resultant graph, and it may de-noise the data in a more structured way than simply choosing different thresholds.

References

[1] Cornelis J Stam, BF Jones, G Nolte, M Breakspear, and Ph Scheltens. Small-world networks and functional connectivity in alzheimer's disease. *Cerebral cortex*, 17(1):92–99, 2007.

[2] Sophie Achard and Ed Bullmore. Efficiency and cost of economical brain functional networks. *PLoS Comput Biol*, 3(2):e17, 2007.

[3] Stavros I Dimitriadis, Nikolaos A Laskaris, Yolanda Del Rio-Portilla, and George Ch Koudounis. Characterizing dynamic functional connectivity across sleep stages from eeg. *Brain topography*, 22(2):119–133, 2009.

[4] Mark A Kramer, Eric D Kolaczyk, and Heidi E Kirsch. Emergent network topology at seizure onset in humans. *Epilepsy research*, 79(2-3):173–186, 2008.

[5] Matthew F Glasser, Stephen M Smith, Daniel S Marcus, Jesper LR Andersson, Edward J Auerbach, Timothy EJ Behrens, Timothy S Coalson, Michael P Harms, Mark Jenkinson, Steen Moeller, et al. The human connectome project's neuroimaging approach. *Nature neuroscience*, 19 (9):1175–1187, 2016.

[6] Mikail Rubinov and Olaf Sporns. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, 52(3):1059–1069, 2010.

[7] Sergey Ivanov and Evgeny Burnaev. Anonymous walk embeddings. *arXiv preprint arXiv:1805.11921*, 2018.

[8] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008.