Hyperbolic graph embedding of magnetoencephalography brain networks to study brain alterations in patients with subjective cognitive decline

by

Cole Baker

B.S. in Computer Science and Engineering, Massachusetts Institute of Technology, 2020

Submitted to the the Department of Electrical Engineering and Computer Science

in partial fulfillment of the requirements for the degree of

Master of Engineering in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2022

Author .....
the Department of Electrical Engineering and Computer Science
May 13, 2022

Certified by .....
Boris Katz
Principal Research Scientist
Thesis Supervisor

Certified by .....
Dimitrios Pantazis
Principal Research Scientist
Thesis Supervisor

Accepted by .....
Katrina LaCurts
Chair, Master of Engineering Thesis Committee

Hyperbolic graph embedding of magnetoencephalography brain networks to study brain alterations in patients with subjective cognitive decline

by

Cole Baker

Submitted to the the Department of Electrical Engineering and Computer Science
on May 13, 2022, in partial fulfillment of the
requirements for the degree of
Master of Engineering in Electrical Engineering and Computer Science

Abstract

Identifying subtle changes in brain activity in the early stages of pathology is crucial for gaining understanding of the causes and mechanisms of neurodegenerative diseases such as Alzheimer's disease (AD). Mapping high dimensional brain connectivity information to a lower dimensional latent space can allow quantitative analysis of the subtle changes in brain activity and create information-rich inputs to downstream classification tasks. Using a Hyperbolic Graph Convolutional Network (HGCN), we embed functional brain connectivity graphs derived from magnetoencephalography data to a Poincare disk instead of traditional Euclidean space. The Poincare disk is a negatively curved unit disk that encourages a continuous tree-like (and low-dimensional) embedding where paths between sibling nodes pass through a more central parent node. This model allows scale-free graphs to be embedded into 2 dimensions with low distortion while maintaining a conformal mapping of angles to Euclidean space. The Poincare model is particularly useful for neuroscientific analysis, as brain networks are generally scale-free, and the low dimensional mappings can facilitate learning despite the typically small datasets that are available in the field. The embeddings provide a parsimonious description of both similarity and hierarchy, which can be used to study the role of individual brain regions and known functional subnetworks, such as the default mode network (DMN) and ventral attention network (VAN). We used the hyperbolic embeddings to assess MEG brain network alterations in subjects with Subjective Cognitive Decline, a pre-clinical precursor to AD in which the subject cannot be objectively diagnosed through traditional neuropsychological testing. Poincare embeddings were used to classify subjects' disease state and identify functional changes in the interconnectivity of several subnetworks as well as the overall hierarchical placement of those networks.

Thesis Supervisor: Boris Katz
Title: Principal Research Scientist

Thesis Supervisor: Dimitrios Pantazis
Title: Principal Research Scientist

Acknowledgments

My great appreciation goes to Dimitrios Pantazis and Mengjia Xu for giving me this opportunity, their guidance, and the proper balance of inspiration and constructive deflation.

Thank you to Boris Katz, who has been my advisor as a undergrad and a Masters student.

Thank you to Isabel Suárez Méndez from the Universidad Complutense de Madrid for allowing me to pester about the data from across an ocean.

Lastly I want to thank my parents, Steve and Allison, for supporting my adventures and misadventures in life and learning.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Alzheimers Disease

Alzheimer's disease (AD) is neurodegenerative brain disorder and the most common form of dementia. In 2020, worldwide AD prevalance was estimated at 50 million [24], with diagnosis rates predicted to rise in the near future. Total diagnoses are predicted to rise to the worldwide total of 152 million cases by 2050 [7]. However, delaying the onset or progression of the disease by 1 year would lead to an overall drop in 9 million patients [14]. Understanding the earliest disruptions to normal neural behavior is key to treating existing pathologies as well as preventing new progressions.

1.1.1 Known biomarkers

Several biomarkers form a profile of AD that can provide clues for the structural causes and early detection of AD as well as evidence of its damage. These include atrophy of the cortex, reduction in cerebral spinal fluid (CSF) decreased fluorodeoxyglucose 18F (FDG) uptake on Positron emission tomography (PET) and elevated CSF tau and APOE genetic status [40]. Amyloid beta accumulation have been studied particularly as a biomarker that exists in the precursor stages to AD. Animal models suggest that the excitation/inhibition balance is disrupted during progression of AD, leading to the alteration of large-scale brain networks. In particular, increased levels of Amyloid-

Beta proteins damage inhibitory terminals of GABAergic neurons, leading to neural hyperactivity.[9] This is followed by a hypoactivity that is characterstic of the later stages of AD. This monotonic increase, followed by monotic decrease is known as the "X-Model" of AD conversion.[39]

1.1.2 MCI

Mild Cognitive Impairment is a key prodromal stage of AD that is characterized by disruptions on neurophysiological tests such as clinical dementia rating, mini-mental state examination (MMSE) score and Montrial Congnitive Assessment (MoCa) [22]. It does not meet the criteria of dementia, but MCI patients present with many of the biomarkers for fully developed AD. MCI, can be stable or progressive with approximately 10-15 percent of patients progressing to dementia per year, while others remain stable or improve over time [15]. According to the X-model, progressive MCI demonstrates the characteristic hyperactivity which is closely followed by neural disconnect and the onset of dementia.[39]

1.1.3 Subjective Cognitive Decline

Subjective Cognitive Decline is an early potential precursor to MCI and AD. It is defined by self-experienced persistent decline in cognitive capacity in comparison with a previously normal status that is unexplainable by an acute event, psychiatric disease or neurological disease. [25] Although a variety of diseases may cause SCD, there is evidence that a majority of SCD patients already contain AD biomarkers. There is a significant conversion to MCI, which some argue to be as high as 8% annually [41]. Many clinical tests cannot detect the self-perceived cognitive deficiency at the SCD stage. Detection and understanding of the neural signatures of SCD may lead to preventative treatments or retroactive cures by understanding the root causes of the origins of the AD pathological cascade.

1.2 Magnetoencephalography

Magnetoencophalograph (MEG) is a functional brain scan that directly measures the magnetic fields produced by intraneuronal circuits, a more direct measurement of neuronal activity than methods relying on metabolic responses (fMRI,FDG-PET [26][2][23]. Direct measurement allows for an excellent temporal resolution in MEG imaging, as it need not rely on the rest of the body to react in order to measure a proxy of neural activity. This temporal resolution is a major advantage over current biomarkers, as it allows MEG to capture subtle brain alterations associated with different brain disorders [38][52][31]. MEGs are taken via a casing surrounding the head with no direct contact and as such is completely non-invasive. Electropysological measurements have shown to successfully track several AD biomarkers, such as increased Amyloid-β levels and tau deposits and corresponding synaptic disruptions. [10][44][4]. Given this, MEG imaging is a promising avenue for developing accessible biomarkers for tracking, diagnosing and understanding stages on the pathological cascade of AD.

1.2.1 Brain Wave Types

Direct measurement of neural activity allows the waveforms to be split into bands of activity. Certain frequency regions have been classically used in neuroscience, and are described in Table 1.1 [1]. Disruptions in each of these bands have been studied as biomarkers on the AD cascade. Hyper/Hyposynchronicity in the alpha/beta bands have studied as indicators of MCI onset as discussed in [31][3]. MCI and AD are associated with decreases in the power of gamma rhythms, especially in conjunction with genetic APOE4 mutations [35][19], although these studies consider the overall intensity of the rhythms and not the synchronicity across regions. Alternatively, increases in the power of the theta band has been linked to MCI and atrophy in the hippocampus [34] [20]. Clearly, evidence from each of these bands is complex and non-linear, and as such deserve extended analysis.

Frequency Bands	
Band Type	Band Frequency
Theta (θ)	4-8 Hz
Alpha (α)	8-12 Hz
Beta (β)	12-35 Hz
Gamma (γ)	35+ Hz

Table 1.1: Frequency Bands

1.2.2 Regions of Interest

MEG imaging measures the magnetic activity at discrete grids in the brain. A key step in the analysis of this step is the parcelation of those grids into larger, interpretable Regions of Interest (ROIs). This analysis uses the AAL90 Template, which describes 90 ROIs, 45 pairs of symmetric regions [36]. Further discussion of parcellation is found in Section 3.2.3.

1.2.3 Connectivity Matrix

MEG imaging provides the raw waveform of activity over an extended period of time at those discrete grids. Many metrics exist for synthesizing these individual waveforms into single measures of connectivity. These include Pearson correlation, mutual information, phase locking value (PLV), and correction imaginary phase locking value (ciPLV)[37][8][50]. While our analysis uses PLV, the application of any of these metrics allows different functional neuroimaging methods (fMRI, PET and MEG) to be translated into the general form of a connectivity matrix. This takes the form of a symmetric square matrix C, where the value at $C[i, j] = C[j, i]$ indicates the connectivity value of regions i, j.

Chapter 2

Previous Work

2.1 Brain as a Graph

Given the ROIs of a brain and a connection between them, the brain can naturally be embedding as a graph with nodes and edges representing ROIs and connections respectively. The most natural method is to pick a threshold for connectivity value at which the nodes are considered to be linked. Although we will use the PLV matrix from MEG scans, this method generalizes to any functional brain image that can provide a connectivity matrix.

2.2 Graph Embedding

While graphs are a useful method for conceptualizing structured data, it is not trivial to perform normal mathematical operations on a graph. For that reason, it is helpful to map nodes to a latent geometric space that maintains the structures of the graph. In particular, these methods try to faithfully maintain the relative closeness of nodes in the embedding space to closeness in the graph. Graph embeddings have found success learning meaningful representations of various complex networks such as airport networks, internet routing and protein folding. Traditional methods include Isomap and node2vec, which have been applied to brain networks in fMRI. Xu et al embeds brain graphs into a probabilistic geometry that is used to predict progression or sta-

bility of MCI for both fMRI and MEGs [51][50]. In general, embedding methods are self-supervised and thus do not need large training sets. Many methods are transductive, that is they operate only on one graph and to not transfer knowledge from one datapoint to another.

Metrics of the fidelity of a graphs embedding include Mean Average Precision (mAP) and distortion. mAP is a relative measure that compares how often non-nieghbor nodes are mapped closer than neighbor nodes

2.3 Graph Machine Learning

The past few years have seen the emergence of Graph Neural Networks (GNNs) that operate explicitly in the graph domain by passing weights only from node to connected node. Due to their interpretability and superior performance, these methods have become the default method for graph analysis in diverse modalities. More recently, GNN methods have appeared in fMRI literature. [28] uses GNNs with ROI pooling for disease detection and interpretation Wang and Kim use traditional GNN methods for various tasks on the Human Connectome Project (HCP).

2.3.1 Message Passing Framework

GNNs take an input feature vector for every node, which is aggregated between neighboring nodes in successive layers. At every layer, node i calculates it's hidden state

$$h_i^l = W^l x_i^{l-1} + \boldsymbol{b}^l \tag{2.1}$$

where x_i^k is the representation of node i at step l, $W^l \in \mathbb{R}^{d_{l-1} \times d_l}$ where d_l is the dimension on the representation at layer l and the matrix W contains the learned parameters of the model. This hidden state is then passed as a message to all of its direct neighbors.

The state of a node evolves according to

$$x_i^{l+1} = \sigma(h_i^l + AGG_{j \in N(i)}(h_j^l) \tag{2.2}$$

where AGG is a generic aggregation function, and σ is a nonlinear function. Common aggregation functions include averages, summations and learned attention aggregation networks[11] . The choice of input feature x_i^0 is discussed further in 3.3.3 .

2.4 Hyperbolic Spaces

2.4.1 Hyperbolic Geometry

Hyperbolic geometry is a non-Euclidean geometry that studies spaces of negative curvature. In network science, hyperbolic geometry has gained attention for its ability to model hierarchical data. Intuitively, hyperbolic space can be thought of as a 2D disk with a fixed radius. Distances between two points "cost" exponentially more as they reach the edge of the disk, such that distance to reach the edge itself is infinite. Shortest paths in this space are not straight lines, but arcs that bend closer to the origin to take advantage of the lower cost. The midpoint of that arc is analogous to a common parent node in the tree. Just as the number of child nodes in a branching tree grows exponentially with distance from the root, the continuous space for embeddings grows with distance from the origin. Networks with tree-like structures can be embedded in this space with fewer dimensions and with less distortion than in the Euclidean space. Figures 2-1(a,b) give a sense for the intuition behind using hyperbolic geometry to embed trees.

2.4.2 Hyperbolic Math and Machine Learning

Applying standard deep learning algorithms in hyperbolic space presents many difficulties, as many standard operations are not obviously defined. Here, we give a brief primer on equations necessary for intuition, and offer literature for a deeper dive

[16][42].

The fundamental operation in the hyperbolic space is the distance function:

$$d(u,v) = \arccos\left(1 + 2\frac{||u-v||^2}{(1-||u||^2)(1-||v||^2)}\right) \tag{2.3}$$

Intuitively, the hyperbolic distance increases whenever the distance between u and v increases, or either node tends towards the radius of the disk.

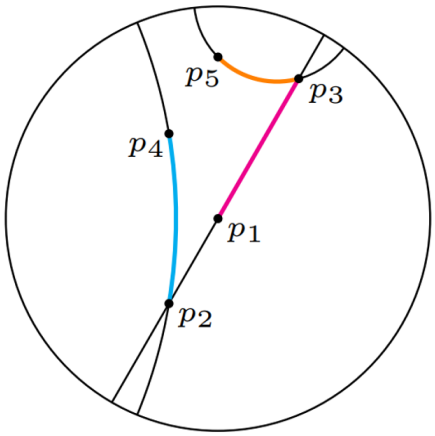
In the case where a particular operation is not-defined or intractable in hyperbolic space, the logarithmic and exponential maps are used. The exponential map takes x_h, a point in hyperbolic space, v_x and the tangent vector that is hyperbolic to the hyperbolic space at x, and transforms $x_h -> x_e$ the euclidean equivalent. After normal euclidean operations are applied in this space, the logarithmic mapping transforms $x'_e -> x'_t$. This trick is computationally useful, but will exacerbate errors due to noise or numerical precision [11].

2.4.3 Hyperbolic Embedding

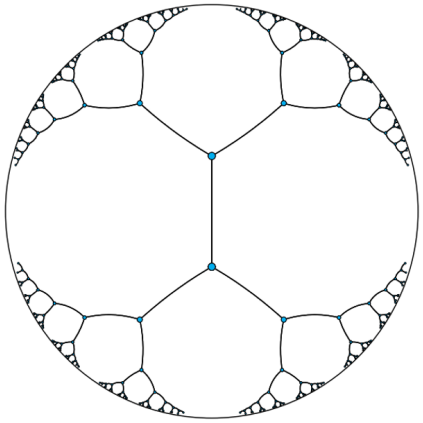
As such, deep learning has been enabled in the hyperbolic space and forms an exciting frontier for brain embeddings. There have been a few attempts to leverage hyperbolic deep learning for connectivity embeddings. [49] uses Mercator, a hyperbolic implementation of Isomap and studies population differences in patients with Autism, while [17] implements the Poincare Embeddings from [32] to recreate known functional subnetworks such as the Default Mode Network and the Executive System

2.4.4 Hyperbolic Graph Neural Network

Hyperbolic geometry is a natural fit for GNNs, as many of the complex networks that are analyzed as graphs have implicit hierarchies built in. Several concurrent works have adapted GNNs to work in a space where some common operations, such as taking a mean of a set, do not have known closed formed solutions. Liu et al. [29] presents HGNN which projects hyperbolic embeddings to the Euclidean tangent space and aggregates messages with a Euclidean mean. Alternatively, Chami et al [11] presents



(a) Geodesics on the Poincare Disk



(b) Embedding a perfect tree on 2D Poincare Disk

Figure 2-1: Hyperbolic Visualization (a) Due to the negative curvature, the distance between points grows exponentially with respect to Euclidean distance as points approach the boundary. This leads to shortest paths that are not straight lines, but segments of circles orthogonal to the boundary (b) A tree embedding with all connected nodes equally spaced (all segments are of equal hyperbolic length). (Visualations taken from [32])

HGCN that learns the curvature of the space instead of assuming a fixed curvature, which allows for more precise projection to Euclidean space, and learns an attention vector that weights messages for aggregation. These two papers cover fundamentally different tasks. HGCN takes one or a few large graphs, predicts missing links and classifies nodes. For example, the model takes a graph of the world's airports and predicts the population of the region to which the airport belongs. Alternatively, HGNN classifies the graphs themselves (i.e., predicting properties of a molecule).

Chapter 3

Methods

The following section will present the pipeline for the MEG embeddings and corresponding design choices. Section 3.1 describes the patient study, Section 3.2 describes how the connectivity matrix is formed from raw MEG, Section 3.3 describes choices in creating the graph from the connectivity matrix, Section 3.4 describes the core HGCN embedding algorithm and modifications made too it.

3.1 Participants

The following study consists of an initial cognitive screening and MEG, an experimental period and a follow up screening and MEG. For further detail, see[45]
Recruitment for this study was from Centre for Prevention of Cognitive Impairment (Madrid Salud), the FAculty of Psychology of the Complutense University of Madrid (UCM), and the hospital Clinico San Carlos (HCSC) in Madrid, Spain between January 2014 and December 2015. All participants were between 65-80 years old, right-handed and Spanish natives.

During the initial screening, patients were assessed on various cognitive tasks including – and placed into the SCD or Healthy Control (HC) groups based on diagnosis. The diagnosis of SCD was agreed on by multidisciplinary consensus in accordance with the following criteria i. Self-reported cognitive concern ii. No possible SCD con-

founders such as medication or psycho-affective disorders). iii. Patient $>$ 60 years at the onset of SCD iv. SCD occuring within last 5 years v. Patient does not have MCI. MCI patients were additionally not eligble for the health control group. The eligible group was further narrowed by removing 34 patients who did not atttent the second MEG session and 22 patients whose scans presented technical reconstruction issues. The resulting sample consists of 41 HC and 49 SCD patients. These patients were randomly split into a trained group (n=46, 24 SCD) that underwent a cognitive training program and placebo group (n=44, 25 of which SCD). These groups were adjusted for age. The experimental group (Trained) underwent the training according to the UMAM Method[33].

3.2 MEG

3.2.1 Data Acquisition and pre-processing

MEGs were acquired at the Laboratory of Cognitive and Computational Neuroscience (UCM-UPM) in the Centre of Biomedical Technology (CTB)(Madrid, Spain) using a 306 channel (102 magnetometers, 204 planar gradiometers). Four minutes of resting-state electrophysiological activity were recording while patients rested awake with their eyes closed. MEG data were acquired using a sampling rate of 1000 Hz and an online anti-alias band-pass filter between 0.1 and 330 Hz. Recordings were processed offline using a tempo-spatial filtering algorithm [47] to eliminate magnetic noise originating outside the head. This algorithm was repeated to compensate for head movement during the recording. Data were segmented into 4-s intervals of artifact-free activity (epochs). The number of clean epochs did not differ across groups,

3.2.2 Source Reconstruction

Clean epochs were band-pass filtered into classic MEG bands Table 1.1 using a 1800th order FIR filter using a hamming window. To avoid edge effects, the epochs were padded with 2 seconds of real data on each end, which were removed upon filtering.

The source model consisted of 1220 sources that were places in a homogeneous grid of 1 cm using the Montreal Neurological Institute (MNI) template which was converted to subject space by affine transformation. MEG sources were anatomically parcelled into 90 regions of interest (ROI) as defined in the Automated Anatomical Labeling atlas [36]. The lead field was calculated with a single-shell head model [37] with a unique boundary defined by the inner skull generated from the individual T-1 weighted MRI using Fieldtrip. Source reconstruction was performed for each subject using a Linearly Constrained Minimum Variance (LCMV) beamformers [48]. Beamformer Filters were obtained using the computed lead field, the epoch averaged covariance matrix, and a 1pct regularization matrix.

3.2.3 Connectivity Analysis

Functional Connectivity (FC) between each source was estimated using the phase locking value (PLV) which estimates source to source connectivity according the phase synchronization [37]. This metric is based on the assumption that the degree of non-uniformity of phase differences between two time-series is a proxy of their coupling. PLV has shown high reliability across MEG recordings [18]. Connectivity of ROI to ROI was calculated by averaging the pairwise PLV values between all sources in any two ROIs:

$$PLV_{A,B} = \frac{1}{N_A N B} \sum_{N_A} \sum_{N_B} \left| \frac{1}{T} \sum_{t} e^{-j(\phi_{A_k}(t) - \phi_{B_l}(t))} \right| \tag{3.1}$$

where $\phi_{A_k}(t)$ and $\phi_{B_l}(t)$ are the instantaneous phases of the signal A_k and B_l at the instant t, T is the number of temporal points per epoch, j is the imaginary unit, N_A is the number of sources in area A, and A_k is the kth source inside that area. We calculate this value for every source, for every frequency band, and are left with six 90x90 connectivity matrices (one for each frequency band).

3.2.4 RSN Assignment

Resting State Networks (RSNs) are subnetworks of ROIs that act together for different mental functions. Changes in the patterns of these RSNs are known to be indicators

of neural disturbances, such as the pDMN hyperactivation in patients with SCD [45].

Atlases that describe ROIs of these functional networks often differ greatly from structural atlases like the AAL because they are based on patterns of communication instead of physical structures in the brain. Through an accumulation of studies into different RSNs [43][6][21][5], we form a collective atlas which we will refer to at $AtlRSN$. $AtlRSN$ describes 8 RSNs. RSNs contain anywhere from 4 to 12 Functional ROIs, which we are defined by a 3d coordinates. In order to assign ROIs in AAL to RSNs described in $AtlRSN$, we adopt a simple technique of mapping regions that have significant spatial overlap using the center coordinates of each $AtlRSN$ ROI and the grid coordinate of each MEG source. Each source is included in a functional ROI if it falls within a certain radius r of the center of that ROI. Each AAL ROI is given an inclusion score for each RSN indicating the percentage of its sources that fall within r. Mathematically speaking:

$$P_{R,A} = \frac{1}{N_A} \sum_{N_A} \sum_{N_R} \begin{cases} 1, & \text{if } dist(C_{A_k}, C_{R_f}) < r \\ 0, & \text{otherwise} \end{cases} \tag{3.2}$$

Where A,N_A,and A_k are as before, R is a certain functional RSN in $AtlRSN$, R_f is a the fth functional ROI belonging to R, N_R is the number of functional ROIs in R, $P_{R,A}$ is the probability that region A belongs to R, $dist$ is the distance between two points, C_i is the 3d subject space coordinate of point i, and r is the threshold radius. Note that an average is only taken over the number N_A, not N_R, because the score calculates the percentage of *all* nodes in A that are near *any* ROI in R.
Lastly, an inclusion threshold t is selected such that A is assigned to R if $P_{R,A} > t$. ROIs are allowed to be assigned to multiple RSNs. r was selected at 1.5 cm to account for the 1 cm resolution of MNI coordinates. This threshold gives enough space for $\sqrt{2cm}$ which allows for all coordinate cubes that share at least one edge. Given r, t was chosen to be .3, as it limited all RSNs to reasonable size and regions known to belong to a network were correctly placed. Figure 3-1 shows a summary of RSNs, while Appendix A shows full inclusion lists for each RSN.

Resting State Network Summary			
RSN Name	Abrv.	Ex. ROI	ROI Abrv.
Posterior Default Mode Network	pDMN	Precuneus	Precu
Anterior Default Mode Network	aDMN	Cingulate gyrus, Anterior part	ACC
Dorsal Attention Network	DAN	Precentral gyrus	PreCG
Fronto-parietal Network	FPN	Angular gyrus	Ang
Visual Network	VN	Middle Occipital Lobe	MOccL
Ventral Attention Network	VAN	Inferior Frontal gyrus, Triangular	IFGt
Salience Network	SN	Amygdala	Amyg
Sensorimotor Network	SMN	Supplementary Motor area	SMA

Figure 3-1: Resting State Networks and an example ROI.

3.2.5 Noisy Augmentation

The data is augmented by randomly injecting gaussian noise into the ROI connectivity matrix of each band at low levels ($\sigma = \sigma_b * .01$) where σ_b is the standard deviation of the MEG band b for the entire dataset. This is applied 3 times per scan to create $n = 180 * 3 = 540$ scans per band.

3.3 Creating Brain-Graph

3.3.1 Band Selection

Given the scope of our analysis, we limit our data to two MEG bands. Previous statistical analysis of this dataset in [45] studies effects of SCD on the alpha band connectivity, which clearly warrant further analysis. In order to maximize information, we would like to analyze the band that shares the least information with the alpha band. Figure 3-2 shows the average correlation value between bands for the PLV values of the corresponding ROIs. We note that the alpha band is significantly less correlated with other bands. Additionally, the lowest correlation of any two bands is Alpha with Gamma. With this in mind and given other research showing gamma band pathological influence, we will study Alpha and Gamma bands.

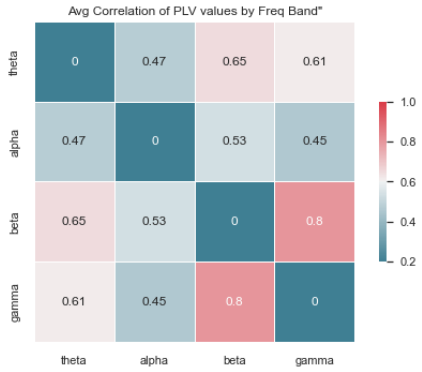


Figure 3-2: Average correlation of bands tends to increase based on their spectral proximity

3.3.2 Threshold Selection

The connectivity matrix is transformed into a binary adjacency matrix by applying a threshold. Values above this threshold will become a 1 and signify an edge while the rest will be become 0s. The PLV distributions of the bands differ, so we select a consistent percentile that will translate to the individual thresholds in each band. We select the eightieth percentile, which corresponds to .329 and .247 PLV for the alpha and gamma bands respectively. This is the highest threshold value before the graphs become disconnected. A highly disconnected graph causes problems for a HGCN as nodes become isolated from the any incoming messages. This problem is slightly remedied by adding a virtual node that is with connections to every node as discussed in [29]

3.3.3 Input Features

Selection of input features is key to a meaningful information aggregation in all graph-based neural networks. The selected information will be passed back and forth but this aggregation cannot make meaning out of nothing. For example,an atom can be represented as a one-hot feature for describing its atom type and an integer of its atomic weight [29], while a PPI can be represented as a 16-dimensional feature of RNA expression levels of the corresponding proteins [46] and an article in a citation network can be represented by a bag-of-words vector indicating word presence in

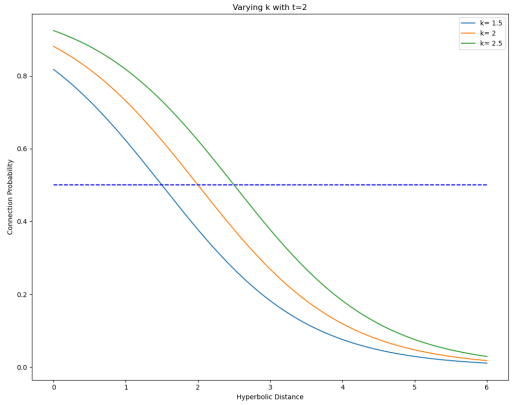
the title or text body. For graphs without obvious node features, structural node features, such as node degree, can help GNNs capture information on neighborhood patterns.[13]. Alternatively, each node can be given a unique one-hot identity vector to capture information unique to that node, although this be unsuitable for large graphs as the number of parameters will explode with the number of nodes.
Previous GNN research in brain connectomics has used a mix of the previous methods. [12] The connectivity vector for each node can be used as the input feature [28] , although this runs into the scaling consideration as the identity vector. A final idea is to leverage the symmetry of the AAL atlas by creating a binary "Side-of-Brain" variable and an identity vector that is shared by the corresponding left and right ROIs, shrinking the input space by slightly less than half.

3.4 HGCN

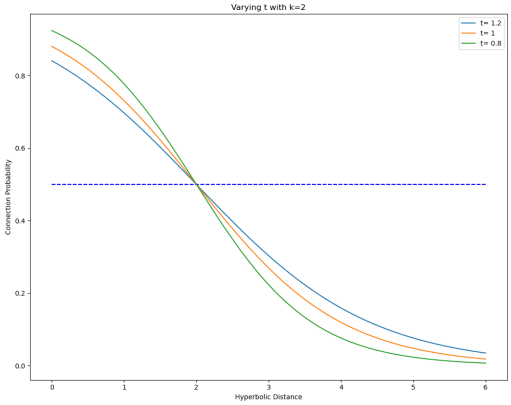
We use the HGCN architecture found in [11] as the core of our embedding model. However, we make several modifications. As we discuss, some modifications are made to make the model more suitable to our problem, while others to update the model given recent advancements in hyperbolic machine learning.

3.4.1 Prediction loss via FermiDirac Decoder

For linking prediction and optimization, the HGCN uses the Fermi-Dirac decoder as described in [32][27]. The Fermi-Dirac decoder is a generalization that computes, given some metric distance $d(i, j)$, the probability of an edge existing between nodes i,j according to $P(i \in N_j|x_i^L, x_j^L) = \frac{1}{e^{(d(i,j)-r)/t}+1}$. Here, r represents the inflection point at which the probability of an edge transitions from above .5 to below and t represents the steepness the probability rise and falls beyond that point, as demonstrated in Figure 3-3. We expand on the FermiDirac decoder in section Figure 3.4.2 to create a loss function weighted by the connection strength instead of our derived binary edges. Additionally, we explore learnable fermi-dirac parameters in Figure 3.4.5



(a) The k parameter horizontally shifts the probability distribution



(b) The t parameter shifts steepness of the probability distribution.

Figure 3-3: Fermi-Dirac Probability Distributions with varying parameters

3.4.2 Weighted Loss Function

Although our model uses binary edges, we would like to make as much use of the initial PLV values as possible. One possibility is to use the PLV features as input vectors to the HGCN. Additionally, we can infuse the PLV values into our loss function. The Fermi-Dirac decoder is described as predicting connection probability, but the predicted quantity need not be an actual probability. Following this idea, we stretch the PLV values to the [0,1] range and treat this value as the ground truth probability of a connection. The model then learns to predict this value, so that nodes connected by the highest PLV value will not be treated the same as nodes with a PLV value just above the threshold.

3.4.3 Iterative Training

The original HGCN framework is designed for link prediction in large scale graphs, such as airport networks and disease networks with nodes from between 500 to 5000 nodes. In general, they are single graphs that are learned with the validation being held out edges on the same graph. The HGCN is inductive, as shown by an experiment on a disease progression network in which the model was trained on one distinct

subgraph and validated on the another.

Our task is fundamentally different as it seeks to synthesizing learning from many similar graphs. Thus we implemented an iterative training schedule with batch learning that is standard in current deep learning.

3.4.4 Frechet Mean

As discussed in Section 2.3.1, the aggregation of messages is a fundamental step in any Graph Convolutional Neural Network. The most common method of aggregation is to take the mean of all incoming messages. In the hyperbolic space, the execution of a mean operation is not a trivial. The hyperbolic analogue of the mean is the Frechet Mean, which did not have a close form, differentiable solution until developments after the publication of the HGCN [30]. With this in mind, we replace the simplify the HGCN by removing the attention-based aggregation layer with a Frechet Mean aggregation. This eliminates unnecessary parameters as well as avoiding an additional logarithmic map to the tangent space.

3.4.5 Learnable Fermi Dirac Parameters

Previous work on hyperbolic embeddings has treated the k and t parameters as hyperparameters, where both papers take them to be 2 and 1 respectively. Changes to these parameters change the optimal embedding space, so a deeper inspection seems to be warranted. As mention in [11], changes to the curvature of the geometry itself can be mimicked by finding the correct k,t pair. In that vein, we create FD mode for the HGCN in which the model's normal parameters are frozen and the Fermi-Dirac features are adjusted by gradient descent. The training protocol alternated so that the model is put in FD mode on every third epoch, jointly learning the embeddings and the optimal distance function. This feature is used for exploratory purposes only to prevent instability in the analyzed embeddings.

Chapter 4

Results

In this section, we will describe results of link embeddings through several lenses. First, we look at the quality of embeddings in terms of link prediction in 4.1.1. Then we will analyze the patterns of brain regions and rsns and differences across brain waves in 4.2. Finally, we will examine differences between disease groups from statistical method and classification in 4.3 and 4.4.

4.1 Embedding Results

4.1.1 Link Prediction

Although our models are trained using the fermi-dirac decoder for probability prediction, we measure the embedding quality using a standard metric, mean average precision (mAP) so that comparisons can be made on a standard scale. We use the mAP to assess both the optimal embedding dimension and to show the benefits of using the hyperbolic geometry for our embeddings in Figure 4-1. We find that the hyperbolic embeddings consistently outperform the euclidean embeddings, especially at the lowest dimensions. Based on the diminishing improvements after adding more than 3 dimensions, we use 3 dimensions for the remainder of our analysis.

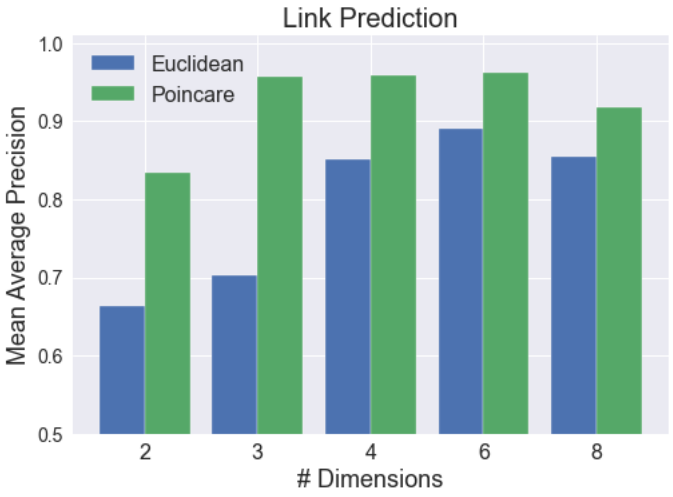
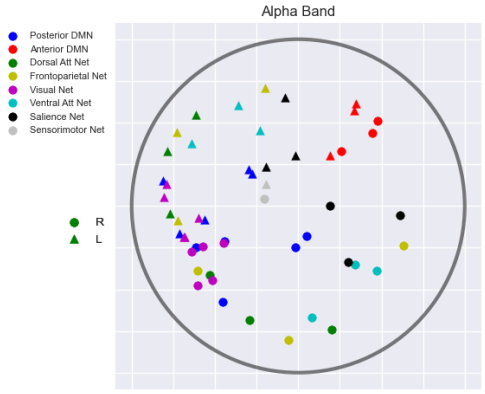


Figure 4-1: Mean Average Precision for Hyperbolic and Euclidean embeddings of the alpha band with different numbers of dimensions.

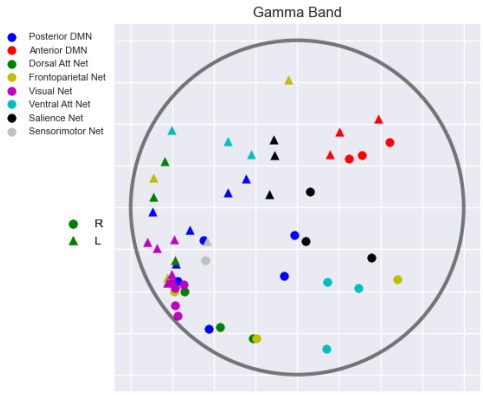
4.2 Brain Analysis

In this section, we analyze the embeddings themselves for patterns and hierarchies. On visual inspection of the embeddings in Figure 4-2, we see that the ROIs in the Salience Network and Posterior Default Mode Network seem to be the most central to hyperbolic hierarchy. Additionally, we see a strong divide between regions on the left and right hemispheres. RSNs generally are divided symmetrically across the axis of the hemisphere. Certain RSNs (Visual, Anterior DMN) are clustered tightly along that hemisphere so that all the nodes connect, where as other RSNs (Ventral Attention, Frontoparietal) are symmetrical at a distance. One may predict that the ROIs assigned to an RSN are the most important to brain function and would all be centrally located. Figure 4-3 show that many unassigned nodes are located more centrally than nodes in RSNs.

A visual comparision of the gamma embeddings shows that many nodes are pushed closer to the edges than in the alpha embeddings, particularly one cluster that involved the dominated by the Visual Network. In general, the alpha band embeddings seem to be spread more evenly across the disk than gamma band embeddings.

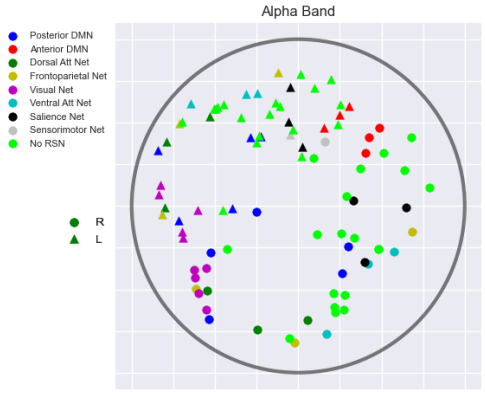


(a) Alpha band embeddings show clear distinctions across side of the brain and symetrical clusters of RSNs

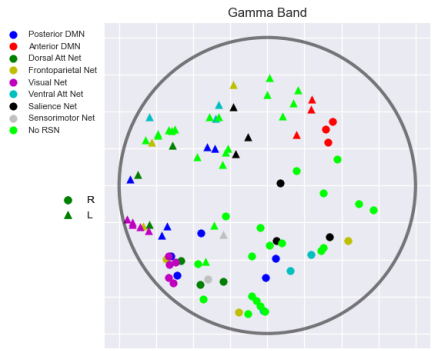


(b) Embedding of the Gamma band shows similar patterns, but with certain clusters pushed toward the disk boundary.

Figure 4-2: 2D Hyperbolic Embeddings of Patient 50 (plotting only ROIs that are assigned to an RSN)



(a) Alpha Band embeddings show nodes spread evenly across the disk.



(b) Gamma Band embeddings show a skewed distribution of brain regions.

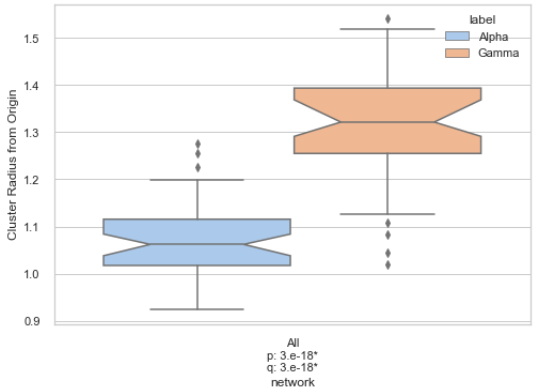
Figure 4-3: 2D Hyperbolic Embeddings of Patient 90 (plotting all nodes)

4.2.1 Hyperbolic Feature Analysis

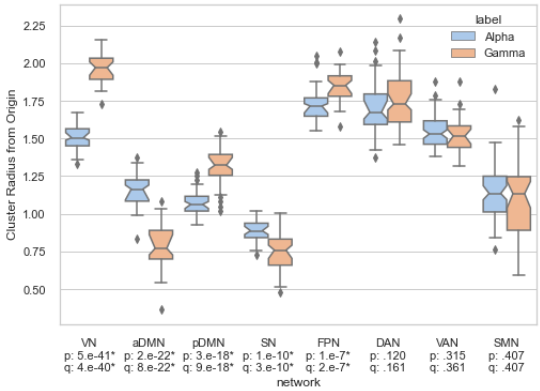
Our analysis in the following passages shows several plots similar to 4-4. The metric "Cluster Radius from Origin" describes the average hyperbolic distance of every node in any particular cluster. A low radius corresponds to a more central cluster. The metric "Cluster Cohesion" describes the average pairwise distance between all nodes in the same RSN. A low cohesion corresponds to a tightly embedded cluster. The cluster in question can be found directly below the plot. The two cluster types we look at are "All" and "RSN", where "All" is every node in a scan and "RSN" is every node assigned to that RSN. The plots show the group-wise probability distributions for a metric for some split in the data (ie. Alpha/Gamma bands, Healthy/SCD patients). P-value statistics indicate the probality that the group distributions are statistically significant. A star denotes significant p-value and q-values, where the q-value is an FDR correction. Each node radius is averaged over 10 random initialized HGCN training sessions.

4.2.2 Alpha-Gamma Band Analysis

Following that format, we analyze differences in the Alpha and Gamma band embeddings. In Figure 4-4a, we broadly analyze the embeddings by overall hyperbolic radius. We find that the average node is closer to the origin in the Alpha than the Gamma embeddings. This may indicate a more connected network. Next, we analyze the RSNs in 4-4b and find that the changes in hyperbolic radius are not uniform across functional subnetworks. Rather some RSNs tend closer to the center in Gamma bands (aDMN, SN) while others tend rather away (VN, FPN). Overall, RSNs radii in the Alpha band are more evenly distributed, compared to the Gamma band in which certain RSNs are very central and others are much more distal.



(a) Alpha Embeddings result in nodes that are closer to the origin on average.



(b) Closer analysis show the changes of in hierarchy of RSNs.

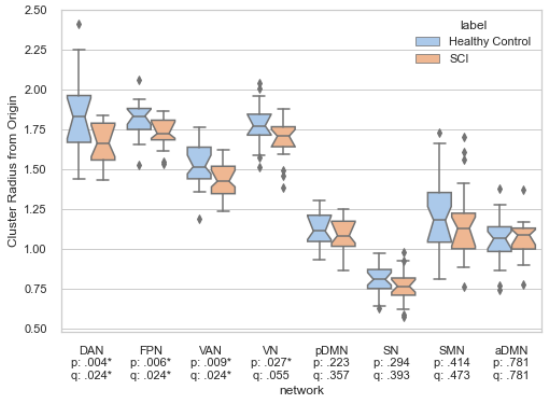
Figure 4-4: Statistical Difference in embedding radius (q value $<$ .05 considered significant, FDR corrected for p-values)

4.3 Disease Analysis

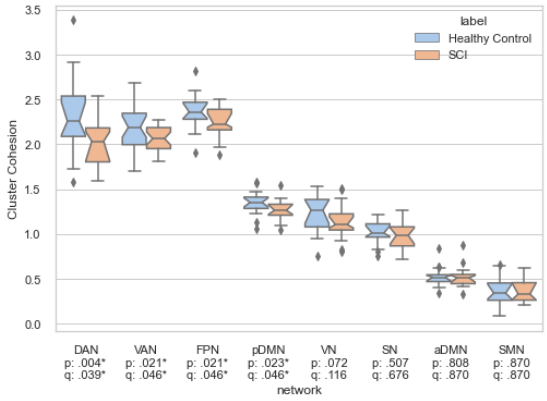
In this section, we perform a similar analysis on embedding differences in the SCD and HC groups with separate comparisons made for the Alpha and Gamma bands. Figures 4-5a and 4-5c show a similar trend among RSNS: the most dorsal regions increase in centrality. This corresponds with existing literature that describes an hypersynchronicity of neural activity in early stages of the AD cascade. Figures 4-5b and 4-5d show a related trend: a tightening of clusters. Most of RSNs with significant cluster tightening also significantly increased their hierarchy. This may mean that the nodes in that region became more active overall, and not just within their cluster. The exception to this is the Posterior Default Mode Network. While the pDMN experiences slight, but non-significant, decrease in radius, the difference in cluster cohesion is quite significant in both sets of embeddings, possibly signifying an insulation of the network.

4.4 Disease Classification

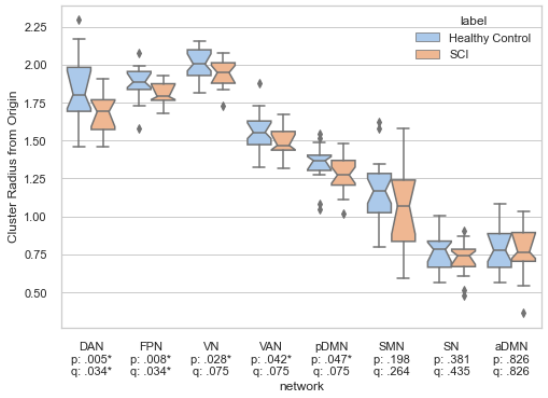
Finally, we show the usefulness of our method feature extraction for downstream tasks by performing classification tasks with and without the hyperbolic radius features.



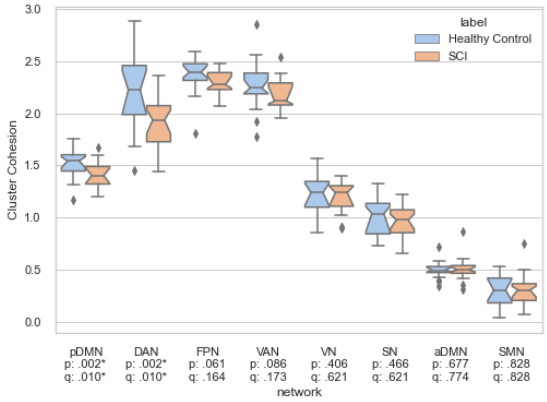
(a) RSN Cluster Radius (Alpha)



(b) RSN Cluster Cohesion (Alpha)



(c) RSN Cluster Radius (Gamma)



(d) RSN Cluster Cohesion (Gamma)

Figure 4-5: Between group changes in RSN embedding statistics between HC and SCD groups

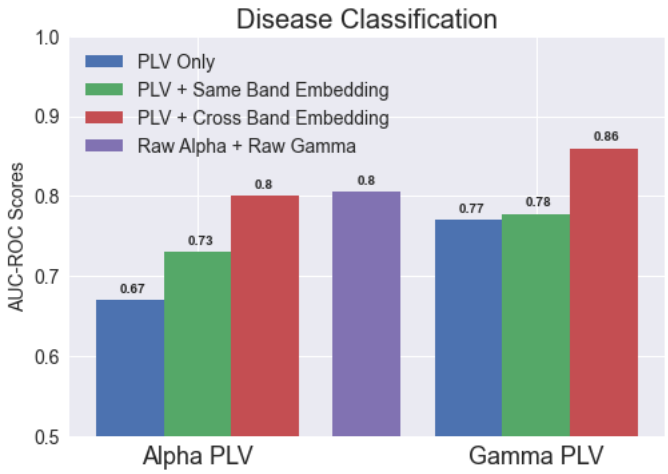


Figure 4-6: Disease Classification AUC-ROC Scores.

We discuss results as seen in Figure 4-6. Using an SVC classifier, we first attempt to classify our patients as HC or SCD with the raw PLV adjacency matrices. This method provides a .67 and .77 area under the ROC curve (AUC-ROC) for Alpha and Gamma bands respectively. Next we augment the PLV matrix of each band with the RSN Features from that band (Same Band Embedding), which increases AUC-ROC scores to .73 and .78.

Given the lack of complete correlation show in 3.3.1, we anticipate an increase in classification score from combining the information of the two bands. By combining the two PLV matrices (Raw Alpha + Raw Gamma), we increase ROC-AUC to .8. Finally, we augment each PLV value with the RSN Features from the other band (PLV + Cross Band Embedding), and see our highest scores of .8 and .86.

These results demonstrate two ways that the embeddings are useful. i. In the case of same band embeddings, the embeddings have extracted information that was not useful in the PLV form, ii. in the case of cross band embeddings, the embeddings' fewer dimensions creates a dense packaging of information compared to concatenating two already large matrices.

Chapter 5

Discussion and Limitations

Our study supports analysis done by [45] suggesting a hypersynchronicity in subjects with SCD. However, while their analysis finds specific increases in the posterior Default Mode Network, our method suggests an alternate framework in which the pDMN remains stationary at the center of a shrinking hierarchy.

These implications are further supported by hypersyncronicity in the gamma frequency band. Further research into the similarities and differences between the bands could examine the role of structural hierarchies of the resulting functional networks. Based on the visual skew of the embeddings and the lack of information gain from Same-Band embedding augmentation, the gamma band initially seems to be a less natural fit in the hyperbolic space. However, this may be an artifact of an unrelated circumstance and should be investigated further.

Additionally, this study did not explore the effects of cognitive training on the subjects. In keeping with the spirit of early training and understanding of the AD cascade, an in-depth analysis of the cognitive training on the changes in hierarchy could provide a more causal understanding of resulting changes in hierarchy.

Appendix A

Resting State Network Assignments

Dorsal Attention Network (DAN)	
Names	Nicknames
['Left Precentral gyrus']	['lPreCG']
['Right Precentral gyrus']	['rPreCG']
['Left Inferior Parietal gyrus']	['lIPG']
['Right Inferior Parietal gyrus']	['rIPG']
['Left Superior Parietal gyrus']	['lSPG']
['Right Superior Parietal gyrus']	['rSPG']

Figure A-1: DAN ROI Assignment

Fronto-Pariental Network (FPN)	
Names	Nicknames
['Left Inferior Frontal gyrus, Triangular']	['lIFGt']
['Right Inferior Frontal gyrus, Triangular']	['rIFGt']
['Left Superior Occipital lobe']	['lSOccL']
['Right Superior Occipital lobe']	['rSOccL']
['Left Inferior Parietal gyrus']	['lIPG']
['Right Inferior Parietal gyrus']	['rIPG']
['Left Supramarginal gyrus']	['lSMG']
['Right Supramarginal gyrus']	['rSMG']
['Left Angular gyrus']	['lAng']
['Right Angular gyrus']	['rAng']

Figure A-2: FPN ROI Assignment

Posterior Default Mode Network (pDMN)	
Names	Nicknames
['Left Cingulate gyrus, Posterior part']	['lPCC']
['Right Cingulate gyrus, Posterior part']	['rPCC']
['Left Hippocampus']	['lHip']
['Right Hippocampus']	['rHip']
['Left Parahippocampus']	['lParahip']
['Right Parahippocampus']	['rParahip']
['Left Angular gyrus']	['lAng']
['Right Angular gyrus']	['rAng']
['Left Precuneus']	['lPrecu']
['Right Precuneus']	['rPrecu']

Figure A-3: pDMN ROI Assignment

Sensori-Motor Network (SMN)	
Names	Nicknames
['Left Supplementary Motor area']	['lMotor']
['Right Supplementary Motor area']	['rMotor']

Figure A-4: SMN ROI Assignment

Ventral Attention Network (VAN)	
Names	Nicknames
['Left Precentral gyrus']	['lPreCG']
['Right Precentral gyrus']	['rPreCG']
['Left Inferior Frontal gyrus, Opercular']	['lIFGo']
['Right Inferior Frontal gyrus, Opercular']	['rIFGor']
['Left Inferior Frontal gyrus, Triangular']	['lIFGt']
['Right Inferior Frontal gyrus, Triangular']	['rIFGt']
['Left Insula']	['lInsula']
['Right Insula']	['rInsula']
['Left Supramarginal gyrus']	['lSMG']
['Right Supramarginal gyrus']	['rSMG']
['Left Angular gyrus']	['lAng']
['Right Angular gyrus']	['rAng']
['Left Middle temporal gyrus']	['lMTG']
['Right Middle temporal gyrus']	['rMTG']

Figure A-5: VAN ROI Assignment

Visual Network (VN)	
Names	Nicknames
['Left Calcarine fissure']	['lCalc']
['Right Calcarine fissure']	['rCalc']
['Left Cuneus']	['lCu']
['Right Cuneus']	['rCu']
['Left Lingual gyrus']	['lLingual']
['Right Lingual gyrus']	['rLingual']
['Left Middle Occipital lobe']	['lMOccL']
['Right Middle Occipital lobe']	['rMOccL']
['Left Inferior Occipital lobe']	['lIOccL']
['Right Inferior Occipital lobe']	['rIOccl']

Figure A-6: VAN ROI Assignment

Anterior Default Mode Network (aDMN)	
Names	Nicknames
['Left Superior Frontal gyrus, Medial']	['lSFGm']
['Right Superior Frontal gyrus, Medial']	['rSFGm']
['Left Superior Frontal gyrus, Medial Orbital']	['lSFGmo']
['Right Superior Frontal gyrus, Medial Orbital']	['rSFGmo']
['Left Cingulate gyrus, Anterior part']	['lACC']
['Right Cingulate gyrus, Anterior part']	['rACC']

Figure A-7: aDMN ROI Assignment

Bibliography

[1] Priyanka Abhang. *Introduction to EEG- and Speech-Based Emotion Recognition*, chapter 2, pages 19–150. Addison-Wesley, 2016.

[2] Sylvain Baillet. Magnetoencephalography for brain electrophysiology and imaging. *Nature Neuroscience*, 2017.

[3] R. Bajo. Functional connectivity in mild cognitive impairment during a memory task: implications for the disconnection hypothesis. *J Alzheimers Dis*, 2010.

[4] Kaj Blennow. Biomarkers for alzheimer's disease: current status and prospects for the future. *Journal of Internal Medicine*, 2018.

[5] Pierre Boveroux. Breakdown of within- and between-network resting state functional magnetic resonance imaging connectivity during propofol-induced loss of consciousness. *Anesthesiology*, 2010.

[6] Pierre Boveroux. Insula–dorsal anterior cingulate cortex coupling is associated with enhanced brain reactivity to smoking cues. *Neuropsychopharmacology*, 2010.

[7] R. Brookmeyer. Forecasting the global burden of alzheimer's disease. *Alzheimers Dement*, 2007.

[8] Ricardo Bruña. Phase locking value revisited: teaching new tricks to an old dog. *J. Neural Eng*, 2018.

[9] Konnerth Busche. Impairments of neural circuit function in alzheimer's diseasel. *Philos Trans R Soc Lond B Biol Sci*, 2016.

[10] L. Canuet. Network disruption and cerebrospinal fluid amyloid-beta and phospho-tau levels in mild cognitive impairment. *Journals of Neuroscience*, 2015.

[11] Ines Chami. Hyperbolic graph convolutional neural networks. *NeurIPS 2019*, 2021.

[12] Hejie Cui. Brainnnexplainer: An interpretable graph neural network framework for brain network based disease analysis. *Eorkshop onInterpretable ML in Healthcare at International Converence on Machine Learning (ICML)*, 2021.

[13] Hejie Cui. On positional and structural node features for graph neural networks on non-attributed graphs. *DLG-KDD'21, Virtual Event*, 2021.

[14] J. Cummings. Alzeimer's disease drug-development pipeline: few candidates, freqent failures. *Alzheimers Res Ther*, 2014.

[15] Christos Davatzikos. Prediction of mci to ad conversion, via mri, csf biomarkers, and pattern classification. *Neurobiology of aging*, 2011.

[16] Octavian-Eugen Ganea and Gary Bécigneul. Hyperbolic neural networks. *NIPS '18*, 2018.

[17] Siyuan Gao. Poincaré embedding reveals edge-based functional networks of the brain. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2020.

[18] P Garces. Quantifying the test-retest reliability of magnetoencephalography resting-state functional connectivity. *Brain Connect*, 2016.

[19] Anna Gillespie. Apolipoprotein e4 causes age-dependent disruption of slow gamma oscillations during hippocampal sharp-wave ripples. *Neuron*, 2016.

[20] Martin Grunwald. Theta-power differences in patients with mild cognitive impairment under rest condition and during haptic tasks. *Alzheimer Dis Assoc Disord*, 2002.

[21] Biyu He. Breakdown of functional connectivity in frontoparietal networks underlies behavioral deficits in spatial neglect. *Neuron*, 2007.

[22] S Hoops. Validity of the moca and mmse in the detection of mci and dementia in parkinson disease. *Neurology*, 2009.

[23] Matti Hämäläinen. Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics*, 1993.

[24] Alzheimer's Disease International. Alzheimer's disease international: World alzheimer report 2015: The global impact of dementia: An analysis of prevalence, incidence, cost. *London Alzheimer's Dis. Int*, 2015.

[25] Frank Jessen. A conceptual framework for research on subjective cognitive decline in preclinical alzheimer's disease. *Alzheimer's I& Dementia*, 2014.

[26] Sheraz Khan. Using the magnetoencephalogram to noninvasively measure magnetite in the living human brain. *Human Brain Mapping*, 2019.

[27] Dmitri Krioukov. Hyperbolic geometry of complex networks. *Physical Review E)*, 2010.

[28] Xiaoxaio Li. Braingnn: Interpretable brain graph neural network for fmri analysis. *Med Image Anal*, 2021.

[29] Qi Liu. Hyperbolic graph neural networks. *NeurIPS 2019*, 2019.

[30] Aaron Lou. Differentiating through the frechet mean. *Proceedings of the 37th International Conference on Machine Learning*, 2020.

[31] Fernando Maestú. A multicenter study of the early detection of synaptic dysfunction in mild cognitive impairment using magnetoencephalography-derived functional connectivity. *NeuroImage: Clinical*, 2015.

[32] Douwe Kiela Maximilian Nickel. Poincaré embeddings for learning hierarchical representations. *31st Conference on Neural Information Processing Systems*, 2017.

[33] P Carrasco Montejo. Programa de memoria. método umam. *Ediciones Díaz de Santos*, 2001.

[34] Davide Moretti. Theta and alpha eeg frequency interplay in subjects with mild cognitive impairment: evidence from eeg, mri, and spect brain modifications. *Frontiers in Aging Neuroscience*, 2015.

[35] Dinavahi Murty. Stimulus-induced gamma rhythms are weaker in human elderly with mild cognitive impairment and alzheimer's disease. *ELife*, 2021.

[36] N Tzourio-Mazoyer N. Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage*, 2001.

[37] G Nolte. The magnetic lead field theorem in the quasi-static approximation and its use for magnetoenchephalography forward calculation in realistic volume conductors. *Phys Med Biol*, 2003.

[38] Christos Papadelis. Pediatric magnetoencephalography in clinical practice and research. *Neuroimaging Clinics*, 2020.

[39] Sandra Pusil. Hypersynchronization in mild cognitive impairment: the 'x' model. *Brain*, 2019.

[40] Laura A. Rabin. Subjective cognitive decline in preclinical alzheimer's disease. *Annu. Rev. Clin. Psychol.*, 2017.

[41] Barry Reisberg. The pre-mild cognitive impairment, subjective cognitive impairment stage of alzheimer's disease. *Alzheimer's I& Dementia*, 2008.

[42] Ryohei Shimizu. Hyperbolic neural networks ++. *ICLR 2021*, 2021.

[43] W.R. Shirer. Decoding subject-driven cognitive states with whole-brain connectivity patterns. *Cerebral Cortex*, 2011.

[44] Milan Stoiljkovic. Altered cortical and hippocampal excitability in tgf344-ad rats modeling alzheimer's disease pathology. *Cerebral Cortex*, 2018.

[45] Isabel Suarez-Mendez. Cognitive training modulates brain hypersynchrony in a population at risk for alzheimer's disease. 2021.

[46] Damian Szklarczyk. The string database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic acids research*, 2016.

[47] S Taulu. Removal of magnetoencephalographic artifacts with temporal signal space separation: demonstration with single-trial auditory-evoked responses. *Hum Brain Mapp*, 2009.

[48] Van Veen. Localization of brain electrical activity via linearly constrained minimum variance spectral filtering. *IEEE Trans Biomed Eng*, 1997.

[49] Wonseok Whi. Hyperbolic disc embedding of functional human brain connectomes using resting state fmri. *bioRxiv*, 2021.

[50] Mengjia Xu. A graph gaussian embedding method for predicting alzheimer's disease progression with meg brain networks. *preprint*, 2020.

[51] Mengjia Xu. A new graph gaussian embedding method for analyzing the effects of cognitive training. *PLOS Computational Biology*, 2020.

[52] Jing Zhang. Classifying post-traumatic stress disorder using the magnetoencephalographic connectome and machine learning. *Scientific Reports*, 2020.