Uncovering statistical differences in unlabeled structural brain networks with variational graph auto-encoders

Andrew R. Sedler
Georgia Institute of Technology
North Ave NW, Atlanta, GA 30332
asedler3@gatech.edu

December 4, 2020

Abstract

The human brain is a highly complex organ whose anatomical characteristics vary on an individual basis and can explain variability in social, emotional, and cognitive processing among healthy individuals as well as differences in disease states and susceptibility to disease. Individualized structural brain networks can be extracted via noninvasive imaging techniques like diffusion tensor imaging (DTI) and high-angular resolution diffusion imaging (HARDI). One challenge in finding statistical differences between structural brain networks of different individuals is the complexity of the graphs themselves, and another is measurement noise. Additionally, we do not have labeled graphs for the different characteristics of healthy individuals, as most research is devoted to understanding disease states. Variational graph auto-encoders (VGAEs) provide a simple and flexible framework for unsupervised learning on graph structured data and an excellent starting point for uncovering differences between structural brain networks of healthy individuals. In this work, we extend the original VGAE by incorporating edge weights into the encoder and optimizing the network for link weight prediction rather than simple link prediction and aggregate node features to find an embedding each graph. We find that the resulting model, which we call the weight VGAE (wVGAE), is able to embed whole brain graphs into a latent space where distance agrees with several other measures of graph similarity.

Introduction

The human brain is one of the most intricate and complex networks found in nature, consisting of hundreds of billions of neurons and hundreds of trillions of synapses. It is characterized by highly organized structures and communication pathways at many different scales. This inherently relational structure makes network science a particularly useful approach for studying the brain. Many neuropsychiatric diseases are known to affect its connectivity and network-level functioning [1]. Network science provides the tools necessary to precisely characterize brain networks in health and disease, allowing us to better understand and identify abnormalities.

A common approach to generating graphical representations of structural brain connectivity is to first divide the brain into discrete regions of interest (ROIs), whether by sensor placement

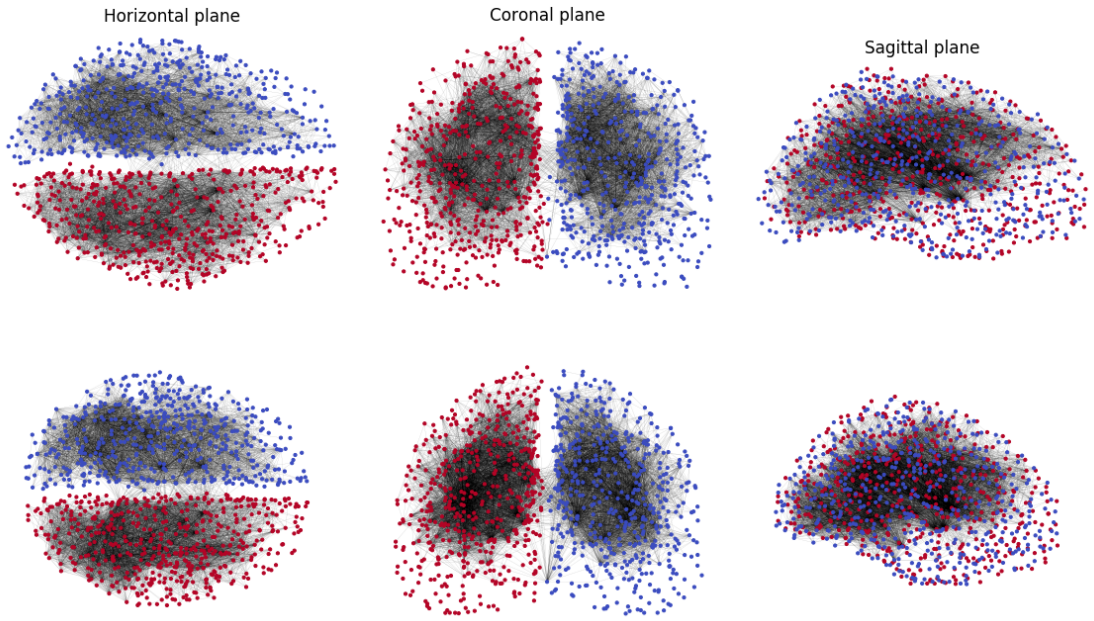


Figure 1: Point clouds representing the physical locations of and connections between ROIs for two randomly selected structural brain graphs in three different anatomical planes. Nodes from left hemispheres are indicated in blue and right hemispheres are red. Edges are represented by thin black lines between the nodes.

or by image segmentation. Then, specialized imaging techniques like DTI or HARDI followed by processing techniques can be used to detect the fibers that connect various ROI's [2]. These graphs can then be characterized using techniques from graph theory and network science to make predictions about disease.

Unfortunately, the complexity of these networks makes them difficult to compare, so discovering any statistically significant natural variability is no easy task. More recently, graph neural networks (GNNs) have been applied to these brain graphs, allowing more effective propagation of features along the graph's edges [7]. When labels are not available, unsupervised methods like VGAEs can be applied to model the latent variables in the data [5]. In this work, we apply a modification of VGAEs with the added reconstruction of edge weights and show that the resulting whole-graph embeddings reflect real differences in the associated structural brain networks.

Methods

Data Preprocessing

The data used in this experiment was originally collected as part of the Human Connectome Project (HCP), which regularly releases functional and structural MRI datasets for hundreds of subjects with the goal of mapping the structural connections of the healthy adult human brain [9]. HARDI structural MRI datasets from the HCP were then passed through a computational pipeline that identified and characterized neural fiber connections between 1015 regions of interest (ROIs) that were shared across subjects [4]. The resulting dataset consisted of 1064 undirected brain graphs,

where each ROI is represented by a node and edges connect nodes that have neural fiber connections between them.

The dataset as described above was downloaded as 1064 `.graphml` files for further processing. We first removed the isolated nodes because they do not contribute in the message passing layers of the VGAE, and would only add noise to our whole-graph embeddings. We added ROI identities as a one-hot feature vector for each node so the network could leverage ROI identities across different graphs. Finally, we added the number of fibers as edge weights between nodes so this information could be incorporated in the generative model. We combined data from all graphs into a custom `Dataset` object in PyTorch Geometric, which made it easy to perform batching and shuffling on multi-graph datasets.

Model Architecture

As in the original VGAE, our model consists of both encoder and decoder modules that were implemented in PyTorch Geometric. The encoder is a two-layer graph convolutional network (GCN) [6] with a ReLU nonlinearity, where the second layer has two separate branches that parameterize μ and σ for each node. In our experiments, the hidden dimension is 64, the latent dimension is 32, and we add improved self-loops to the weighted adjacency matrix ($\hat{\mathbf{A}} = \mathbf{A} + 2\mathbf{I}$) to the adjacency matrix used by the GCN. We also impose a KL penalty on the divergence of $q(\mathbf{Z}|\mathbf{A}, \mathbf{X})$ from a multivariate normal prior $p(\mathbf{Z})$ and weight its contribution to the loss by a hyperparameter λ.

The decoder module is similar to the inner product decoder used in the original VGAE, except we remove the sigmoid and instead compute MSE between inner products and edge weights. For computational efficiency, we use negative sampling (neg) to encourage the inner products to be zero for unconnected nodes. Thus, the complete loss is

$$\mathcal{L} = \sum_{i,j \in E} (\mathbf{z_i^T z_j} - \mathbf{A_{ij}})^2 + \sum_{i,j \in neg(E)} (\mathbf{z_i^T z_j})^2 + \lambda KL[q(\mathbf{Z}|\mathbf{A}, \mathbf{X})||p(\mathbf{Z})]$$

Finally, we pooled all of the node embeddings for each graph using mean, median, and max to generate a 96-dimensional embedding for each graph, as is commonly done in graph classification architectures [3].

Traditional Metrics

To investigate whether graph embeddings were informative about real similarities and differences within the graphs, we selected the graph pairs that were most similar and most different as measured by Euclidean distance in the graph embedding space. We then quantified the similarity of these graphs by more established network science metrics like Jaccard similarity on nodes and edges, degree distributions, correlations between edge attributes, clustering coefficients, and centrality. Importantly, we also plotted the networks according using node position information to subjectively compare the networks.

Other Architectures

Throughout the evolution of this model, we also experimented with several other architectures that are worth mentioning. However, after each improvement we didn't see any obvious structure in the graph embeddings after applying t-SNE [8] for visualization so we continued to increase

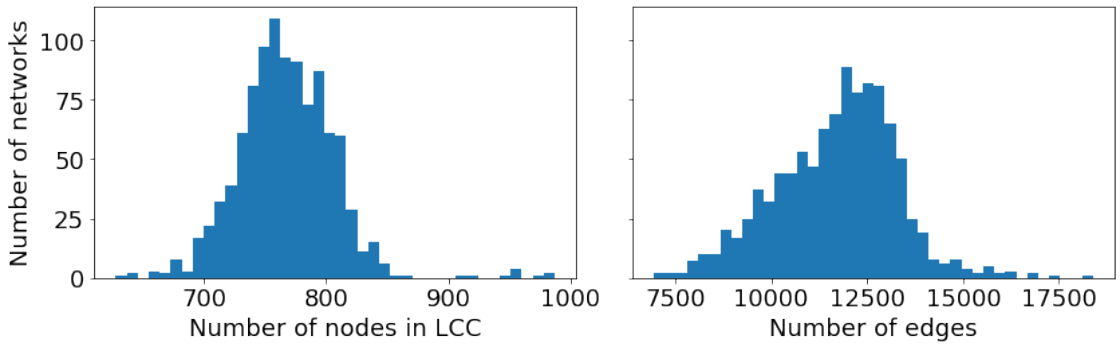


Figure 2: Histograms showing the total number of nodes in the largest connected component and the total number of edges. These plots indicate a significant amount of diversity among the brain graphs.

the representational power of the model. We first implemented a GAE for link prediction, then converted it to a VGAE, then added added the edge weight prediction.

Results and Discussion

Network exploration with traditional metrics is difficult

We started our analysis by characterizing randomly sampled brain networks by a few established metrics. As shown in **Fig. 1**, the node connectivity patterns display a distinct two-lobe structure that is consistent with the two hemispheres of the brain. Generally, the networks show the highest connectivity in the parietal and occipital lobes, with more isolated nodes in the frontal and temporal regions. We found that these networks typically have about 75% of their nodes in the largest connected component (LCC, 767.2 ± 40.1 nodes), and many of the remaining nodes are disconnected. However, we do see pretty significant variation in the number of nodes in the LCC, with some networks connecting nearly all nodes and some connecting only about half (**Fig. 2**). Additionally, the total number of edges varies from nearly 20,000 to as few as 7,000 (**Fig. 2**). Degree distributions follow a power law with the exceptions for very high and very low degree nodes (**Fig. 3**), consistent with the phenomena of low-degree saturation and structural limitations for high-degree nodes. The networks tend to be slightly dissortative on average ($r = -0.012 \pm 0.025$), but degree correlations range from as low as -0.086 to as high as 0.144. Nodes in these networks also tend to have much higher clustering coefficients than a similar $G(n, p)$ network, and a negative correlation between average clustering coefficient and node degree (plots not shown). Additionally, betweenness centrality (weighted by number of fibers) reveals that a very small number of nodes mediate the major connections in the network (plots not shown).

Variational graph autoencoders reveal structured latent space

We trained a VGAE to perform link prediction on these networks and a wVGAE to perform link weight prediction. The VGAE and wVGAE (with a link-prediction decoder) both performed fairly well on the link prediction task, achieving 0.977 ± 0.003 and 0.878 ± 0.008 AUC, respectively.

We then examined the node and graph embeddings produced by these networks (**Figs. 4** and **5**, respectively). The node embeddings produced by the VGAE model were distributed more

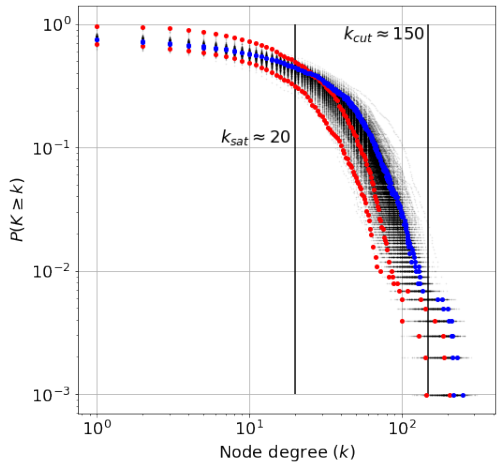


Figure 3: The complement of the cumulative power-law degree distributions for all 1064 structural brain networks. Each network contributes one set of black points. The degree distributions for graphs that the wVGAE embedded similarly in the latent space are shown in blue and the those that the wVGAE embedded differently are shown in red.

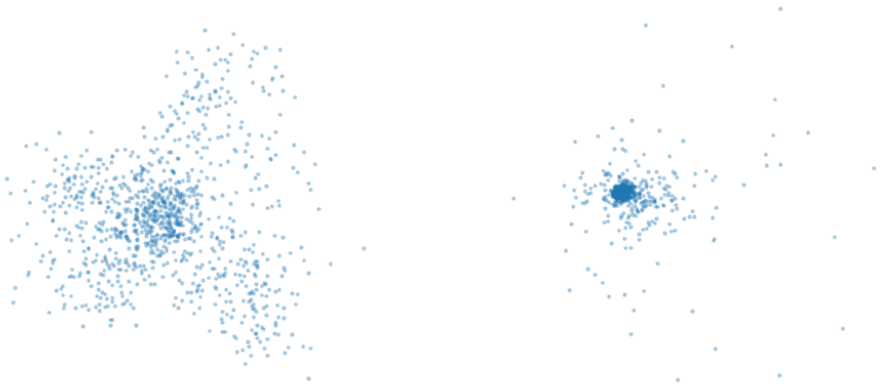
	Nearest Graphs		**Furthest Graphs**	
Nodes in LCC - Number	759	764	980	696
Nodes in LCC - Jaccard	0.883		0.678	
Edges - Number	13159	13206	12598	7619
Edges - Jaccard	0.402		0.182	
Correlation - Fiber count	0.628		0.1258	
Correlation - Betweenness	0.558		0.299	

Table 1: Quantitative comparison between graphs that are nearest and furthest apart in the graph embedding space from the wVGAE. Counts and similarities were quanitifed for the nodes in the largest connected component and all edges in the network. Pearson's r is also shown between fiber counts on the shared edges and betweenness centrality on the shared nodes.

evenly throughout the space, while node embeddings produced by the wVGAE were much more heterogenous. This is likely because for link weight prediction the model must account for significant differences in scale among various links. Graph embeddings reduced to 2D using t-SNE appear to follow a fairly homogenous distribution, with only hints of distinct subpopulations. Overall, this indicates that there do not seem to be distinctly different groups of brain types, but rather a continuum of properties across the population. Of course, this could be partly due to the Gaussian prior used to regularize the VGAEs.

To validate that the models were discovering real differences between the networks, we identified the pairs that were nearest and furthest apart in the graph embedding space to quantitatively and qualitatively compare them. The differences are summarized in **Table 1**. The nearest graphs are more similar by every metric we tested, including shared edges, shared nodes in the LCC, and correlation between edge fiber counts betweenness centralities.

Finally, we plotted the graphs anatomically to see whether distance in the embedding space was an indication of whether the brain structural networks were subjectively different (**Figs. 6** and



(a) **(b)**

Figure 4: First two PCs of node embeddings from (a) a model trained to predict links and (b) a model trained to predict edge weights.

(a) **(b)**

Figure 5: Embeddings of all 1064 graphs from the link prediction and weight prediction models, dimensionality reduced by t-SNE. (a) Model trained to predict links and (b) model trained to predict edge weights.

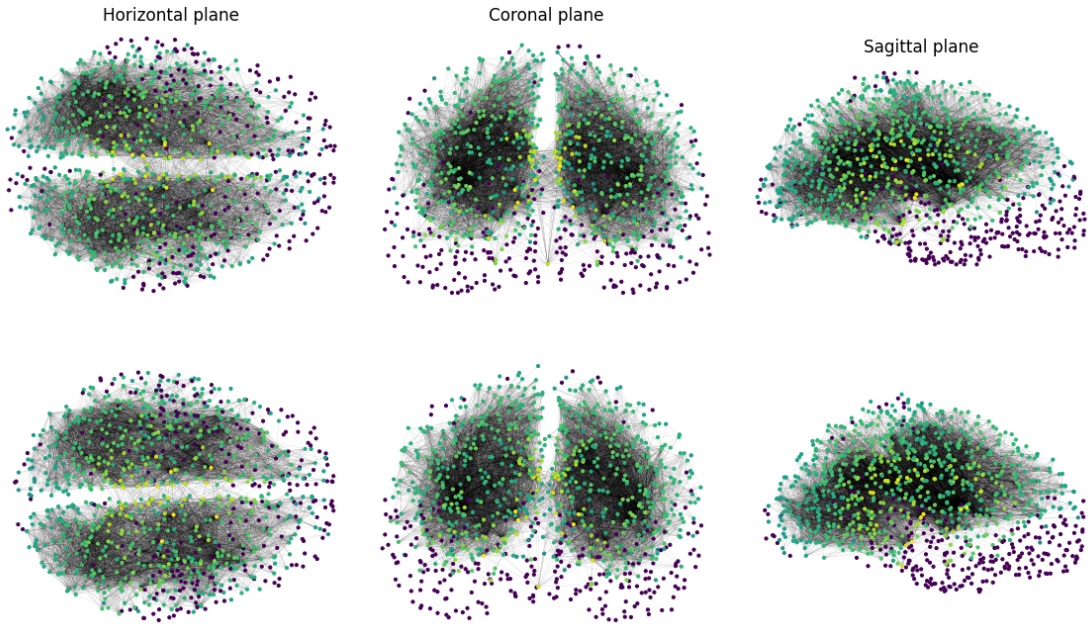


Figure 6: Point clouds representing the physical locations of and connections between ROIs for the two graphs with the smallest euclidean distance in the graph embedding space. Conventions are the same as in **Fig. 1**, except nodes are colored by closeness centrality determined using the mean fiber length.

7). The nearest networks appear very similar in all three anatomical planes, with the connections spanning roughly the same regions with roughly the same densities. In contrast, the networks which were embedded some distance away in the latent space were very qualitatively different from each other. One network was highly connected throughout, including nearly all nodes in the LCC. It was even highly connected in the frontal and temporal regions, which is atypical for this dataset. The second network was even more lacking in frontal and temporal connections than any previous network. The densities of the edges are visually much sparser than a typical network in nearly all regions. These qualitative differences support our assertion that the wVGAE is discovering the latent structure underlying the dataset.

Conclusion and Future Directions

We have shown that the wVGAE is able to find a structured latent space from unlabeled, healthy structural brain networks of adult humans. This approach, when combined with other participant data, could help scientists identify a landscape of previously undiscovered types of neurodiversity.

Future work could incorporate reconstruction of more features into the framework (including node positions and anatomical distances), giving the wVGAE the information necessary to detect more subtle types of variation. It may also be of interest to run this analysis on labeled structural and functional brain networks and seeing if the structure of the embedding space contains information about the class labels. Finally, I think it may be interesting to use node pooling layers like in Graph U-Nets to reduce the influence of less relevant nodes on the final graph embeddings [3].

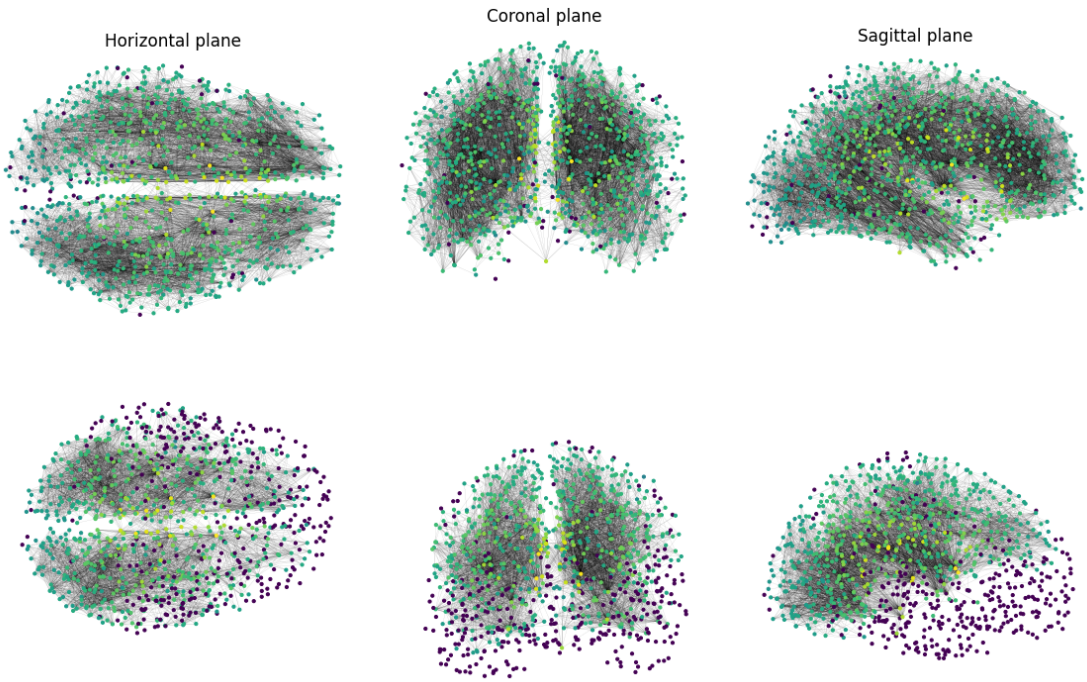


Figure 7: Point clouds representing the physical locations of and connections between ROIs for the two graphs with the largest euclidean distance in the graph embedding space. Conventions are the same as in **Fig. 6**.

References

[1] Danielle S Bassett and Edward T Bullmore. "Human brain networks in health and disease". In: *Current opinion in neurology* 22.4 (2009), p. 340.

[2] Danielle S Bassett, Ankit N Khambhati, and Scott T Grafton. "Emerging frontiers of neuroengineering: a network science of brain connectivity". In: *Annual review of biomedical engineering* 19 (2017), pp. 327–352.

[3] Hongyang Gao and Shuiwang Ji. *Graph U-Nets*. 2019. arXiv: `1905.05178 [cs.LG]`.

[4] Csaba Kerepesi et al. "The braingraph. org database of high resolution structural connectomes and the brain graph tools". In: *Cognitive neurodynamics* 11.5 (2017), pp. 483–486.

[5] Thomas N Kipf and Max Welling. "Variational graph auto-encoders". In: *arXiv preprint arXiv:1611.07308* (2016).

[6] Thomas N. Kipf and Max Welling. *Semi-Supervised Classification with Graph Convolutional Networks*. 2017. arXiv: `1609.02907 [cs.LG]`.

[7] Xiaoxiao Li and James Duncan. "BrainGNN: Interpretable Brain Graph Neural Network for fMRI Analysis". In: *bioRxiv* (2020).

[8] Laurens van der Maaten and Geoffrey Hinton. "Visualizing Data using t-SNE". In: *Journal of Machine Learning Research* 9.86 (2008), pp. 2579–2605. URL: `http://jmlr.org/papers/v9/vandermaaten08a.html`.

[9] Jennifer A McNab et al. "The Human Connectome Project and beyond: initial applications of 300 mT/m gradients". In: *Neuroimage* 80 (2013), pp. 234–245.